Durham University

Durham E-Theses

Evaluating teachers' use of research evidence to improve students' literacy outcomes

CHEN, WENQING

How to cite:

CHEN, WENQING (2025) *Evaluating teachers' use of research evidence to improve students' literacy outcomes*, Durham theses, Durham University. Available at Durham E-Theses Online: http://etheses.dur.ac.uk/16282/

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full Durham E-Theses policy for further details.

Academic Support Office, The Palatine Centre, Durham University, Stockton Road, Durham, DH1 3LE
e-mail: e-theses.admin@durham.ac.uk Tel: +44 0191 334 6107
http://etheses.dur.ac.uk

Evaluating teachers' use of research evidence to improve students' literacy outcomes

Wenqing Chen

PhD thesis

School of Education

Durham University

2025

Abstract

Policymakers, practitioners, and researchers from various sectors have long been curious about bridging the gap between research and practice, seeking to incorporate research evidence into real-life applications and policymaking across diverse domains. In education, it appears both logical and desirable to apply the best available research evidence to enhance teaching quality and effectiveness.

However, using research evidence in daily teaching in schools presents considerable challenges. There is growing but still incomplete knowledge of 'what works' in classroom interaction, and there is even less secure evidence on the most promising approaches to encourage and facilitate teachers' use of research evidence. Moreover, even the feasibility of asking teachers to engage with and utilise research evidence is uncertain, let alone determining how to maintain high-quality implementation over time. Based on these research gaps, this study is a further exploration to understand the use of research evidence in teaching practice.

The study starts by examining teachers' access to research information, their views of the usefulness of research findings to improve teaching and the strongest predictors of these views. These are investigated with TALIS data and the NFER Survey of Teachers 2010. The secondary data analysis of TALIS revealed that reading professional literature was a common way for teachers worldwide to access educational research. Over the past decade, there has been little change in teachers' engagement with professional literature, research conferences, or conducting research, suggesting that analysing the Survey of Teachers 2010 remained meaningful. Findings from the Survey of Teachers suggested that using research findings was not a popular approach to improve teaching practice. Confidence in understanding and using research findings emerged as the strongest predictor of evidence-use. The data also highlighted the role of supportive school environments and teacher peer networks in facilitating the transition from accessing research to integrating it into teaching.

Next, the study summarises the best approach to disseminating and engaging teachers with research evidence via a structured review. 2,257 existing studies were screened, and 13 studies were found that evaluated the impact of teachers' evidence use on students' literacy outcomes. Despite the limited research on this topic, the structured review identified that having researchers engage in transferring and disseminating modified research evidence was the most tested and promising approach. Research evidence that is simplified to easy-understanding versions and activities, such as workshops and regular follow-ups, has the potential to encourage teachers' use of research evidence

and promote learning outcomes. Although the observed impact was modest, most high-quality studies highlight the cost-effectiveness of these approaches.

With the foundation of existing evidence on teachers' use of research evidence, the current study then estimates the impact of a use-of-research-evidence intervention in the form of engaged dissemination of modified research evidence. The fieldwork was completed in China to fill key gaps in the exploration of using evidence in educational research so far in China. A pilot randomised controlled trial (RCT) was conducted in May 2023, involving 2,022 students and 46 teachers across two primary schools, to test the feasibility of the intervention materials. Lessons from the pilot highlighted the need for more detailed instructions, improved communication, and example lessons to support teachers in incorporating research evidence. Additionally, insights gained during the pilot led to better data collection methods to maintain accuracy and efficiency.

The main RCT expanded to 6,904 students and 154 teachers across four primary schools in the same region, with an additional school participating via a quasi-experimental approach. For the impact analysis, the study primarily focused on data from the four RCT schools, as the quasi-experiment school lacked an appropriate comparison group and its data were incomplete. An impact analysis including the quasi-experiment school can be found in the Appendix. Classroom observations from the quasi-experiment school were included in the descriptive analysis of the investigation process.

The impact analysis for the four RCT schools revealed a small positive effect size of +0.07 on students' Chinese literacy scores. The use-of-research-evidence intervention also potentially fostered students' positive attitudes towards reading. These positive effects were particularly pronounced among students with rural Hukou and those from migrant backgrounds. A slight reduction in homework time for Chinese literacy was also observed (-0.06). The use-of-research-evidence intervention influenced teachers' decision-making and research engagement. It enhanced their awareness of the usefulness of research evidence (+0.26) and improved their ability to interpret research findings (+0.15). Although the impact on students' Chinese literacy outcomes was marginal, the study underscores the feasibility of incorporating high-quality research evidence into teaching practice. It demonstrates that such interventions can enhance teachers' awareness and skills, forming a foundation for sustainable, long-term implementation.

The process evaluation explored the process of using research evidence and teachers' implementation quality based on the Complier Average Causal Effect (CACE) analysis, four headteacher interviews, five

school research leads interviews, informal teacher conversations, and forty-four classroom observations. The role of teacher compliance in realising the intervention's potential impact is reflected a CACE of +0.22 on students' Chinese test scores among those whose teachers adhered to the implementation protocol. Factors associated with compliance included school-level planning, teachers' confidence and skills, and collaboration among teachers. Classroom observations suggested that the adaptation of research evidence evolved iteratively, with initial attempts often being procedural and inaccurate. Teachers required time and effort to refine their approaches, incorporating classroom feedback and tailoring research-based materials to better meet their own students' needs.

An implementation quality cycle was discussed at the school, individual teacher, and teacher interactive levels. At the school level, effective planning, monitoring, and clear communication with teachers were key to ensuring exposure to research findings. At the teacher's individual level, skills and confidence in using research evidence were relevant to promote active engagement with research evidence. At the teacher interactive level, teacher-peer collaboration may help them identify and adjust ineffective practices based on research findings.

This study offers valuable insights into the impact of teachers' use of research evidence on students' literacy outcomes. It highlights the feasibility and sustainability of using research evidence in teaching practice through a use-of-research-evidence intervention. It also identifies essential factors for the high-quality implementation of evidence-informed teaching strategies, offering advice to further promote the effective use of research evidence in educational practice. Further research should focus on generating high-quality evidence on 'what works' in getting evidence into use with more open-access materials to facilitate educators' use. Additional studies are needed to explore the impact of using research evidence across other subjects, with more diverse samples, broader age groups, and varied evidence-use approaches.

Table of contents

Abstract i
List of tables xii
List of figures xv
List of abbreviations xvi
Declaration xvii
Statement of Copyright xviii
Acknowledgements xix
Chapter 1 Introduction 1
1.1 Use of research evidence in practice 1
1.1.1 Research evidence use across fields 1
1.1.2 Development of evidence-use in teaching 2
1.2 Urgent need for the use of research evidence 4
1.3 Aims of the study 6
1.4 Research questions 6
1.5 Structure of the thesis 8
Chapter 2 What is research evidence? 11
2.1 Description of research evidence 11
2.1.1 What are high-quality primary studies? 12
2.1.2 Uncertainty of existing primary studies on 'what works' 13
2.1.3 Are synthesis studies trustworthy? 14
2.2 Use of research evidence in this study 15
2.2.1 Transformation of research evidence 16
2.2.2 Two dimensions of using evidence 17
2.3 Why does the use of research evidence matter? 20
2.3.1 Role of research evidence in school improvement 20
2.3.2 Potential benefits of evidence-use in teaching practice 21
2.3.3 Struggles for research evidence 23
2.4 Conclusion 24
Chapter 3 Barriers to and facilitators of using research evidence 25
3.1 How to use research evidence? 25
3.1.1 Models and frameworks 25
3.1.2 What is the best approach to using research evidence? 27
3.2 Use research evidence from users' views 28
3.2.1 Are teachers aware of the value of research evidence? 28

3.2.2 Access, time, and skills 30
3.2.3 Support use environment 31
3.3 Quality of implementation 32
3.4 Specifying education contexts in Mainland China 35
3.4.1 Overview of the Chinese primary education system 35
3.4.2 New curriculum standard of Chinese literacy in primary schools 36
3.4.3 Investigation of research evidence-use 37
3.5 What can existing studies not answer? 40
3.5.1 Limited knowledge about barriers to evidence users 40
3.5.2 Uncertainty of the most effective approach for evidence-use 41
3.6 Conclusion 41
Chapter 4 Research design and methods: Secondary data analysis 44
4.1 TALIS 44
4.1.1 Research questions 44
4.1.2 Design 44
4.1.3 Sample description 45
4.1.4 Items description 45
4.2 Analysis for TALIS 46
4.3 Survey of Teachers 2010 46
4.3.1 Research questions 47
4.3.2 Design 48
4.3.3 Sample description 48
4.3.4 Items description 49
4.4 Analysis for Survey of teachers 49
4.4.1 Data cleaning and missing data 50
4.4.2 Descriptive analysis 50
4.4.3 Logistic regression modelling 50
4.4.4 Path analysis 53
4.5 Conclusion 54
Chapter 5 Research design and methods: Structured review 55
5.1 Research questions 55
5.2 Search strategy 55
5.2.1 Search terms 55
5.2.2 Databases 56
5.3 Inclusion and exclusion criteria 56
5.4 Screening 57
5.5 Data extraction 57

5.6 Quality assessment....58
5.7 Conclusion....59
Chapter 6 Research design and methods: evaluating teachers' use of research evidence....60
6.1 Research questions....60
6.1.1 Research questions for impact evaluation....60
6.1.2 Research questions for process evaluation....61
6.2 Design....61
6.2.1 Sample....62
6.2.2 Randomisation....62
6.3 Development of the use-of-research-evidence intervention....63
6.3.1 Booklet of evidence-informed teaching strategies....64
6.3.2 Teacher training presentation....70
6.3.3 Teacher workshops and follow-ups....72
6.4 Outcome measurement....72
6.4.1 Primary outcome....72
6.4.2 Secondary outcome....73
6.4.3 Non-cognitive outcome....76
6.5 Data analysis for impact evaluation....77
6.5.1 Intention-to-treat analysis....78
6.5.2 Imbalance at baseline....78
6.5.3 Missing data....79
6.5.4 Primary outcome analysis....80
6.5.5 Secondary outcome analysis....81
6.5.6 Non-cognitive outcome analysis....81
6.5.7 Effect size....81
6.5.8 Regression-to-the-mean analysis....82
6.6 Data analysis for process evaluation....83
6.6.1 Complier average causal effect analysis....84
6.6.2 Qualitative comparative analysis....85
6.6.3 Classroom observation and interviews....87
6.7 Ethics....88
6.8 Conclusion....88
Chapter 7 Situation of teachers' use of research evidence: results from secondary data analysis....90
7.1 Teachers' access to educational information around the world....90
7.2 Trend of teachers' access to educational research information....92
7.3 Do teachers find research evidence useful for improving teaching?....93
7.4 What is the strongest predictor of teachers' perceptions of using research findings?....94

7.4.1 Variables in the model....94
7.4.2 Results of the logistic regression modelling....96
7.5 What is the role of support from schools and colleagues?....97
7.6 Conclusion....99
Chapter 8 The most promising approach to encourage teachers' use of research evidence: results of structured review....101
8.1 Descriptive analysis....101
8.2 Approaches for research evidence use....103
8.2.1 Passive transfer of plain evidence....104
8.2.2 Passive transfer of modified evidence....105
8.2.3 Passive transfer of engineered evidence....106
8.2.4 Engagement in transfer of modified evidence....107
8.2.5 Engagement in transfer of engineered evidence....110
8.3 Summary: A potential approach....112
Chapter 9 Lessons from the pilot study....115
9.1 Teachers' adaptation of research findings into teaching practice....116
9.1.1 Peer tutoring....116
9.1.2 Dialogic teaching (Philosophy for Children)....118
9.2 Do teachers care about research evidence?....119
9.3 What do teachers need to apply evidence-led teaching strategies?....121
9.3.1 Peer tutoring....121
9.3.2 Dialogic teaching....122
9.3.3 Writing about Values....123
9.4 Adjustment for design and methods....124
9.4.1 Randomisation....124
9.4.2 Contamination....124
9.4.3 Survey revision....125
9.4.4 About the process evaluation....126
9.4.5 Generalisation....127
9.5 Adjustment of intervention....127
9.5.1 Use-of-research-evidence intervention....127
9.5.2 Materials for evidence-informed teaching strategies....128
9.6 Conclusion....128
Chapter 10 Impact evaluation: students' Chinese literacy scores....129
10.1 Description of participant students....129
10.1.1 Student background....129
10.1.2 Student baseline Chinese scores....130

10.1.3 Missing data 132
10.1.4 Flowchart for the main analysis 133
10.2 Impact on Chinese literacy gain scores 135
10.3 Impact analysis for different strategies 137
10.4 Regression-to-the-mean effect 138
10.5 Subgroup analysis 140
10.5.1 Impact on students of different Chinese proficiency levels 140
10.5.2 Impact for different sex 140
10.5.3 Impact on ethnicity minority students 141
10.5.4 Impact for rural Hukou students 142
10.5.5 Impact for migrant students 142
10.5.6 Impact on low-income students 143
10.6 Additional analysis for post-test only 143
10.7 Conclusion 144
Chapter 11 Impact evaluation: students' enjoyment of reading 146
11.1 Students in the student survey 146
11.2 Impact on reading material preference 147
11.3 Impact on reading media and purposes 149
11.4 Attitudes towards reading 150
11.5 Self-assessment of Chinese reading ability 151
11.6 Time for Chinese literacy homework 153
11.7 Other changes in students perceived by teachers 154
11.8 Conclusion 156
Chapter 12 Impact evaluation: teachers' perceptions and skills in evidence-use 157
12.1 Teacher characteristics 157
12.2 Missing cases in teacher survey 160
12.3 Does the intervention change teachers' views on research evidence? 160
12.3.1 Resources for decision making 160
12.3.2 How do teachers find research evidence useful for improving teaching? 162
12.3.3 What is required to improve evidence-use? 163
12.4 To what extent do teachers feel confident in their ability to use research evidence? 165
12.5 Does the intervention increase teachers' skills in using research evidence? 168
12.5.1 How do teachers describe research evidence? 168
12.5.2 Access to research evidence 169
12.5.3 Impact on teachers' skills in understanding research findings 172
12.5.4 Impact on teachers' skills in judging research quality 174
12.5.5 Impact on teachers' awareness of evidence-led decision-making 175

12.6 Change in teachers' professional engagement 178
12.6.1 Views of Headteachers 178
12.6.2 View of school research leads 179
12.7 Conclusion 180
Chapter 13 Process evaluation: compliance analysis 181
13.1 Description of school-level planning 181
13.2 To what extent does teachers' compliance make a difference to the intervention effectiveness? 183
13.3 Variables relevant to teacher compliance 185
13.3.1 Leadership participation and research opportunity 185
13.3.2 Confidence and skills 186
13.3.3 Support from peers 187
13.4 What are the key conditions for successful use of research evidence in teaching practice? 187
13.4.1 Success combinations for embedding formative assessment 189
13.4.2 Success combination for peer tutoring 190
13.4.3 Success combination for Growth mindset 195
13.4.4 Success combination for Writing about Values 195
13.5 Conclusion 196
Chapter 14 Process evaluation: what is the role of leadership in the use of research evidence? 198
14.1 Value of research evidence from school leaders' perspectives 198
14.1.1 Promotion of school teaching practice 198
14.1.2 School decision-making and management 199
14.1.3 Adaptation for societal changes and policy shifts 200
14.2 Barriers to the experiment management 201
14.3 Efforts to organise and coordinate the experiment 202
14.3.1 Workload management 202
14.3.2 Leadership engagement and monitoring 204
14.3.3 Effective communication 205
14.4 What may happen after the experiment? 207
14.5 Conclusion 208
Chapter 15 Process evaluation: how is research evidence adapted into teaching? 210
15.1 Embedding formative assessment (EFA) 210
15.1.1 Initial implementation: translating Wiliam's ideas into class-level goals 211
15.1.2 Challenge identification: difference in teachers' interpretation of EFA 211
15.1.3 Solution proposal and practice refinement: emphasising feedback 213
15.1.4 Reflection: better integrating formative assessment into teaching 216
15.1.5 Conclusion 217

15.2 Peer tutoring 217
15.2.1 First cycle: students' ability and rules of peer tutoring 218
15.2.2 Second cycle: looking for the most suitable peer tutoring activity 222
15.2.3 Third cycle: embedded peer tutoring 226
15.2.4 Conclusion 227
15.3 Growth mindset 228
15.3.1 Initial implementation: direct use of materials 228
15.3.2 Challenge identification: overload of content 229
15.3.3 Solution proposal: selective use of materials 229
15.3.4 Practice refinement: refined growth mindset workshop 230
15.3.5 Reflection: skills in adjusting materials from research 230
15.3.6 Conclusion 232
15.4 Writing about Values 232
15.5 Conclusion 234
Chapter 16 Limitation, implication, and discussion 235
16.1 Limitations 235
16.1.1 Secondary data analysis 235
16.1.2 Structured review 236
16.1.3 Randomised controlled trial 237
16.2 Findings summary 238
16.3 Use of research evidence: a feasible movement for teaching 239
16.3.1 Teachers' views on the value of research evidence 239
16.3.2 Empowering teachers' use of research evidence through training 241
16.3.3 Can teachers adapt research findings into teaching practice? 242
16.3.4 Cost-effectiveness and minimal disruption 243
16.4 Implementation quality 244
16.4.1 School-level factors: planning, monitoring, and research tradition 246
16.4.2 Teacher individual level factor: perceptions and skills 248
16.4.3 Teacher interactive level: collaboration and adjustment 249
16.4.4 Cycle of adapting research evidence 250
16.5 Sustainability of using research evidence in teaching practice 251
16.5.1 Potential long-term impact 251
16.5.2 A positive loop 252
16.6 Implications for researchers and policymakers 254
16.6.1 Recommendations for further research 254
16.6.2 Recommendations for policymakers 256
16.7 Conclusion 257

References 258
Appendices 274
Appendix A. Search terms for the structured review 274
Appendix B. Use-of-research-evidence intervention booklet 276
Appendix C. Use-of-research-evidence training slides 309
Appendix D. Implementation guide for evidence-informed strategies 330
Appendix E. Teacher Use-of-Research-Evidence Survey 394
Appendix F. Student Enjoyment of Reading Survey 406
Appendix G. Classroom observation form 410
Appendix H. Interview questions 411
Appendix I. Ethical approval 412
Appendix J. Trends of teachers' access to educational research information by countries 421
Appendix K. Refinement of the logistic regression model 424
Appendix L. Structured review data extraction 427
Appendix M. Impact analysis for the pilot 432
Appendix N. Impact analysis including the quasi-experiment school 439
Appendix O. Truth table 445
Appendix P. Example classroom observation record 448

List of tables

Table 2.1 Two dimensions of evidence-use 17
Table 4.1 Description of TALIS 2008-2018 45
Table 4.2 Items in TALIS 45
Table 4.3 Teacher background characteristics 48
Table 4.4 Items in the Survey of Teachers 2010 49
Table 4.5 Predictors in each block in the regression model 51
Table 5.1 The Gorard 'SIEVE' for quality assessment 59
Table 6.1 Evidence-led strategies for the main study 66
Table 6.2 Evidence on embedding formative assessment 67
Table 6.3 Evidence on peer tutoring 67
Table 6.4 Evidence on dialogic teaching 68
Table 6.5 Evidence on Writing about Values 69
Table 6.6 Evidence on Growth mindset 69
Table 6.7 CACE estimation formula 84
Table 6.8 Compliance Criteria 85
Table 6.9 Calibration of conditions and potential outcome for success 86
Table 7.1 Access to educational information relevant to research, TALIS 2018 91
Table 7.2 Trends of access education information relevant to research 93
Table 7.3 Usefulness of ways to improve teaching (percentage) 94
Table 7.4 Odds ratios of finding research evidence useful 95
Table 7.5 Percentage correct 96
Table 7.6 Coefficient for each predictor 97
Table 7.7 Model explanatory power 97
Table 7.8 Path 'effect' and ratio 99
Table 8.1 Searching results in each database 101
Table 8.2 Summary of impact and research quality 103
Table 8.3 Summary: passive transfer, plain evidence 105
Table 8.4 Summary: passive transfer, modified evidence 106
Table 8.5 Summary: passive transfer, engineered evidence 107
Table 8.6 Summary: engagement in transfer, modified evidence 110
Table 8.7 Summary: engagement in transfer, engineered evidence 112
Table 10.1 Student allocation and background characteristics 130
Table 10.2 Comparison of students' Chinese scores in the pre-test, all students 130

Table 10.3 Comparison of the pre-test, by trial settings 131
Table 10.4 Comparison of the pre-test, by grades 131
Table 10.5 Comparison of the pre-test scores of students missing post-test 132
Table 10.6 Comparison of standardised pre-, post-tests, and gain scores, all students 135
Table 10.7 Comparison of standardised post-test and gain scores, by trial settings 136
Table 10.8 Comparison of standardised post-test and gain scores, by grades 137
Table 10.9 Comparison of standardised post-test and gain scores, by strategies 137
Table 10.10 Comparison of the lowest and the highest 50% students' gain scores (standardised) 139
Table 10.11 Estimation of RTM effect 139
Table 10.12 Effect sizes of students at different reading proficiency levels 140
Table 10.13 Comparison of standard pre, post-tests, and gain scores, by sex 141
Table 10.14 Comparison of standard pre, post-tests, and gain scores, ethnicity minority 141
Table 10.15 Comparison of standard pre, post-tests, and gain scores, rural Hukou students 142
Table 10.16 Comparison of standard pre, post-tests, and gain scores, migrant students 142
Table 10.17 Comparison of standard pre, post-tests, and gain scores, low-income students 143
Table 10.18 Comparison of the post-test scores of students missing pre-test 144
Table 10.19 Comparison of post-test only, all students 144
Table 11.1 Number of students in the survey, by trial settings 147
Table 11.2 Comparison of means in reading material preference, all students 148
Table 11.3 Effect sizes of reading material preference, by student characteristics 149
Table 11.4 Comparison of means in reading media and purposes, all students 149
Table 11.5 Effect sizes of reading media and purposes, by student characteristics 150
Table 11.6 Comparison of means in reading attitudes, all students 151
Table 11.7 Effect sizes of reading attitudes, by student characteristics 151
Table 11.8 Comparison of means in perceived Chinese reading ability, all students 152
Table 11.9 Effect sizes of perceived Chinese reading ability, by student characteristics 152
Table 11.10 Comparison of time for Chinese literacy homework 153
Table 12.1 Teacher background information and allocation 158
Table 12.2 Teacher major subjects for their highest degree qualification 159
Table 12.3 Comparison of resources for decision-making, all teachers 161
Table 12.4 Effect sizes of resources for decision-making, by trial settings 162
Table 12.5 Comparison of perceived usefulness of research evidence, all teachers 162
Table 12.6 Effect sizes of perceived usefulness of research evidence, by trial settings 163
Table 12.7 Comparison of external support for evidence-use, all teachers 164

Table 12.8 Effect sizes of required external support, by trial settings 165
Table 12.9 Comparison of confidence in research evidence-use, all teachers 167
Table 12.10 Effect sizes of confidence in using research evidence, by trial settings 168
Table 12.11 Means of the best description of research evidence, all teachers 169
Table 12.12 Comparison of teachers correctly describing research evidence, all teachers 169
Table 12.13 Effect sizes of correct description of research evidence, by trial settings 169
Table 12.14 Comparison of frequency of accessing research evidence, all teachers 171
Table 12.15 Effect sizes of approaches for research evidence, by trial settings 172
Table 12.16 Means of understanding research findings, all teachers 173
Table 12.17 Comparison of ability to understand research findings, all teachers 173
Table 12.18 Effect sizes of skills in understanding research findings, by trial settings 173
Table 12.19 Means of rating research quality, all teachers 174
Table 12.20 Comparison of research quality identification, all teachers 174
Table 12.21 Comparison of perceived research quality difference, all teachers 175
Table 12.22 Effect sizes of skills in judging research quality, by trial settings 175
Table 12.23 Means of rating actions for the conflict teaching situation, all teachers 176
Table 12.24 Comparison of evidence-led decision-making, all teachers 177
Table 12.25 Adjusted effect size of evidence-led decision-making 177
Table 12.26 Effect sizes of skills in evidence-led decision-making, by trial settings 177
Table 13.1 Average frequency of using strategies 184
Table 13.2 Complier average causal effect, full compliance 184
Table 13.3 Complier average causal effect, intervention activity 185
Table 13.4 Complier average causal effect, implementation frequency baseline 185
Table 13.5 Teacher compliance, leadership, and school research 186
Table 13.6 Comparison of teachers' confidence and compliance 186
Table 13.7 Comparison of teachers' skills and compliance 187
Table 13.8 Comparison of demonstration lesson opportunity and compliance 187
Table 13.9 Necessary condition table 188
Table 13.10 Truth table, Embedding formative assessment 189
Table 13.11 Success combinations, embedding formative assessment 190
Table 13.12 Truth table, peer tutoring 192
Table 13.13 Success combinations, peer tutoring 194
Table 13.14 Truth table, Writing about Values 196
Table 13.15 Success combinations, Writing about Values 196

List of figures

Figure 3.1 What Works Centres' areas of activities (Cabinet Office & HM Treasury, 2018, pp. 5–6) ..27
Figure 4.1 path analysis of using research evidence 53
Figure 7.1 Coefficients for the path analysis 98
Figure 8.1 PRISMA flow diagram 102
Figure 9.1 Flowchart of the pilot study 115
Figure 10.1 Flowchart of RCT schools.....134
Figure 12.1 Years of teaching 159
Figure 15.1 Discussion of quality by using two pieces of anonymous work.....215
Figure 15.2 Example of computer-assisted all-student response system 216
Figure 15.3 Amended rules for peer tutoring, School A.....219
Figure 15.4 Group role assignment card, School C.....221
Figure 15.5 Story ladder 223
Figure 16.1 Diagram of implementation at school and teacher levels.....245

List of abbreviations

AERA	American Educational Research Association
BEI	British Education Index
BERA	British Educational Research Association
BME	Black and minority ethnic
CACE	Complier Average Causal Effect
DfE	Department for Education (UK)
EEF	Education Endowment Foundation
EFA	Embedding formative assessment
ERIC	Educational Resources Information Centre
ES	Effect Size
GTCE	General Teaching Council for England
NFER	National Foundation for Educational Research
OECD	Organisation for economic Co-operation and Development
P4C	Philosophy for Children
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-analyses
QCA	Qualitative comparative analysis
RCT	Randomised Controlled Trial
RLC	Research Learning Communities
ROC	Receiver Operating Characteristic
RTM	Regression to the mean
SDGs	Sustainable Development Goals
TALIS	Teaching and Learning International Survey
UNESCO	United Nations Educational, Scientific and Cultural Organisation
WaV	Writing about Values
WWC	What Works Clearinghouse

Declaration

I declare that this thesis is my own work and has not previously been submitted elsewhere for any other qualification or degree.

Part of this thesis, the structured review and secondary data have been published as book chapters and papers. References below:

Chen, W. (2025). The Usefulness of Research Evidence for Improving Teaching: Insights from a Secondary Analysis of a Survey of Teachers. *International Journal of Education and Research. 13*(7).

Gorard, S., & Chen, W. (2025). What is the way forward for research-informed education?. In D. Wyse, V. Baumfield, N. Mockler, & M. Reardon (Eds.), The BERA-Sage Handbook of Research-Informed Education Practice and Policy. BERA-Sage

Chen, W. (2024). Improved classroom interaction matters to the reading proficiency of low-income students. In Stephen Gorard, Nadia Siddiqui (Eds.), *An International Approach to Developing Early Career Researchers* (pp. 131-143). Routledge.

Statement of Copyright

The copyright of this thesis (including any appendices or supplementary materials to this thesis) rests with the author, unless otherwise stated.

This copy has been provided under licence to the University to share in accordance with the University's Open Access Policy, and is done so under the following terms:

- this copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

- any quotation from the work (e.g. for the purpose of citation, criticism or review) should be insubstantial, should not harm the rights owner's interests, and must be accompanied with an acknowledgement of the author, the work and the awarding institution.

- the content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

Acknowledgements

I would like to express my sincere gratitude to my supervisors, Professor Stephen Gorard and Professor Nadia Siddiqui for their invaluable guidance and support throughout this research. Special thanks go to Professor Beng Huat See for her kindness and thoughtfulness. I am also grateful to my colleagues, friends, and family for their encouragement and patience. Finally, my thanks go to the teachers and students who participated in this study, without whom this research would not have been possible.

Chapter 1 Introduction

For perhaps 30 years, an international movement advocating the use of research evidence in practice has gained momentum. There have been efforts to improve the use of research evidence in educational decision-making. While this approach has attracted considerable attention and investment across various fields due to its potential to enhance effectiveness and strengthen the scientific basis of practice and decision-making, it also faces a number of challenges.

This chapter serves as the introduction to a new study on this topic, providing a comprehensive overview of the purpose and structure of the whole thesis. It begins with a description of the research background and the broader issues that motivate the investigation. The research aims, questions, and designs of the current study are summarised.

1.1 Use of research evidence in practice

'Evidence' can be information, facts, or data that can be trusted to alter our beliefs about how the world appears or is working. Evidence in education can include various forms such as internal evaluation, standardised tests, and statutory inspection that could all be used for school improvement and policy development. For example, teachers who use students' test scores to adjust their teaching could be said to be based on evidence. Findings that have been systematically collected and analysed as part of research studies are an essential sort of 'evidence'. This study focuses on the use of evidence from systematic research, particularly evidence that has been tested robustly.

1.1.1 Research evidence use across fields

The 'research to practice gap' has been a longstanding concern across various fields. The movement of integrating research evidence into practice first emerged most strongly from the healthcare field in the 1990s. Within the medical profession, evidence-based practice is characterised as a decision-making approach that draws from three main sources: (1) the best available research evidence, (2) clinical expertise, and (3) client values (Sackett et al., 1996). Rycroft-Malone et al. (2004) discussed four sources of evidence for patient-centred, evidence-based practice, including research, professional knowledge and clinical experience, local data and information, and patient experience and preferences.

This synthetic form of evidence-based practice has been embedded in the training of medical practitioners. Ilic and Maloney (2014) conducted a systematic study of approaches to teaching evidence-based practice to incoming medical trainees. The review covered a variety of transformative methods from nine randomised controlled trials, including self-directed learning, small group work, and online lectures. While these approaches were relevant to an increase in knowledge, there was no longer-term data on the impact of their practices.

This movement of research evidence use has spread to other fields, such as environmental science, psychology, and social work. Toomey (2016) investigated the use of scientific knowledge in environmental practice and revealed the gap between environmental scientists and local people. In psychology, the American Psychological Association introduced the scientist-practitioner model in 1949 as a foundation for training psychologists (Drabick & Goldfried, 2000). Drake et al. (2009) conducted a meta-analysis of all research evidence that could be located to determine what works to reduce crime and estimated the potential economic benefits of evidence-based options for the Washington (US) legislature.

Beyond supporting practice, attempts have been made to incorporate research evidence into policymaking. Innvaer et al. (2002) conducted a systematic review of the perceptions of health policymakers and emphasised the role of personal contact in using research evidence. This review, comprising 24 studies and 2,041 interviews with health policymakers, shed light on the subject by categorising evidence use into direct, selective, and enlightening forms, considering different types of decision-makers, policy questions, and issues.

The Chinese Clinical Epidemiology Network has been actively promoting evidence-based policymaking in China, making it a key component of the government's strategy for policy development (Wang & Jin, 2011). Bowen and Zwi (2005) proposed an 'evidence-informed policy and practice pathway' to 'adopt, adapt, and act' to support policymaking, based on existing literature from health, public policy, and the social sciences. However, these policy areas offer few signs that, as yet, the use of evidence has markedly increased.

1.1.2 Development of evidence-use in teaching

In a lecture about 'Teaching as a research-based profession', Hargreaves (1996) compared educators with medical practitioners and proposed the idea that teachers could use evidence more in teaching practice and decision-making. Davies (1999) argued that evidence-use in education is not a panacea

but rather a framework of principles and practices aimed at improving educational policy and practice. High-quality evidence can apparently make a difference to educational experiences and lead to important gains for children, the public, and society (Palmer, 1999). Slavin (2002) argued the necessity of large-scale experimental evaluations as high-quality evidence of effectiveness to achieve the potential of a scientific revolution in education to transform education practice.

However, the use of research evidence has been met with critiques and resistance. One concern was that evidence-based practice restricts teachers' professional discretion and authority, posing a broader threat to the democratic purpose of schooling (Holloway & Larsen Hedegaard, 2023). Such concerns often reflect an overemphasis on the role of research evidence in decision-making, portraying teachers as being constrained or even displaced by external sources of authority.

Beyond concerns about professional autonomy, some scholars have also questioned the applicability of evidence-based approaches in education by drawing comparisons with other disciplines. Although medicine and education share certain similarities, their differences make it challenging to assume that what is suitable in one field will automatically be appropriate in the other, particularly considering external validity. Allen (2013) questioned the applicability of findings from randomised controlled trials, arguing that while such studies may indicate what works for the specific schools involved, they offer limited insights into what might be effective in different social contexts with varying teachers and student populations.

From the perspective of users, education practitioners have little incentive to engage with research, which is often seen as alienating, especially when it involves statistical analysis (Hemsley-Brown & Sharp, 2003). Those who do read research are generally doing so as part of academic study rather than to inform their teaching. Teachers require more sustained opportunities to connect research findings with their teaching practice and tend to prefer sources that are both accessible and directly relevant to their needs.

Despite such concerns, in England, the Department for Education (DfE) funded the Education Endowment Foundation (EEF) to support schools in finding and implementing high-quality research evidence in education (EEF, 2021). The What Works Clearinghouse (WWC) was established by the Department of Education's Institute of Education Sciences (n.d.) in the US. Other countries have produced their own versions of many of these initiatives, such as the Swedish Institute for Educational Research, Norwegian Knowledge Centre for Education, and Evidence for Learning in Australia. Social

media has also become a key platform for disseminating educational research, providing teachers with free access to resources that can potentially inspire and inform their practice.

Educational institutions and organisations have started to provide training and support for research users. The American Educational Research Association (AERA) provides advocacy training, including guidance on using education research when engaging with political representatives. EdReports.org offers recommendations to help educators evaluate classroom materials. In the UK, initiatives like ResearchED, and the Chartered College of Teaching, have emerged claiming to support teachers in adopting evidence-based practices. Several websites that summarise findings of educational programme evaluations have also appeared, including Social Programs That Work (www.evidencebasedprograms.org) and the Promising Practices Network (www.promisingpractices.net).

Many schools have established roles such as Research Leads, Research Champions, or Researchers-in-Residence to facilitate discussions on the optimal use of evidence (e.g., Sing et al., 2018). Burn et al. (2021) examined the impact of Research Champions through the Oxford Education Deanery project, which fosters partnerships between the University of Oxford and local schools to support continuous professional development and initial teacher education. Konstantinou (2022) explored the process of establishing a Research and Evidence Centre at a secondary school in south England. At a more structured level, the Research Schools Network has developed support systems through a network of schools aimed at using evidence to enhance teaching practices (DfE, 2018).

To capture this evolution, White (2019) proposed a four-wave model outlining efforts by researchers to promote evidence use in decision-making. The four waves include: (1) the results agenda introduced as part of New Public Management in the 1990s; (2) the growing use of impact evaluations, particularly RCTs, since the early 2000s; (3) the increased production of systematic reviews over the past decade; and (4) efforts to institutionalise evidence-use through knowledge brokering agencies, exemplified by the What Works movement in the US and the UK. These diverse efforts reflect a growing emphasis on promoting evidence use in education.

1.2 Urgent need for the use of research evidence

The quality of teaching is widely regarded as one of the most important factors at the school level that affects student achievement. In September 2015, the United Nations General Assembly adopted the Sustainable Development Goals (SDGs), setting out a global agenda for transformation. The SDG Goal

4, ‘Quality Education’, has been defined as ‘ensure inclusive and equitable quality education and promote lifelong learning opportunities for all’ (UNESCO, 2017).

To improve teaching quality, an increasing number of studies have examined the impact of various programmes or approaches on student learning outcomes. Research evidence now plays an essential role in decision-making and school improvement. A recent document from UNESCO suggested that policy-makers should ‘use solid and reliable data and evidence to ensure that countries progress towards their education targets and understand policies and programmes that could improve their educational outcomes’ (Steiner-Khamsi et al., 2024).

Schools are encouraged to apply research findings in the classroom to develop strategies that enhance student learning. Mincu (2015) indicated that school improvement relies on establishing the capacity for change, fostering high teacher and student morale, and providing positive incentives for teachers to participate in the improvement process.

While there are increasing resources for teachers to get access to research evidence, the quality of these resources is not always reliable, with inaccurate or misleading information (e.g., Test et al., 2015). Some resources use persuasive language to enhance their appeal and perceived credibility, yet they rarely focus on the evidence of their effectiveness (Carrier, 2017). Teachers face the risk of adopting low-quality research or false information, which may negatively affect their teaching. This not only undermines educational outcomes but also represents a lost opportunity for each generation of students and their one-chance at formal education (Gorard & Chen, 2025).

Furthermore, having access to research evidence does not automatically translate into its use in teaching practice. Findings from impact studies are typically presented and interpreted at the average level. Teachers must adapt them to fit the specific needs of their students. This process requires a deep understanding of their students and sufficient resources to facilitate teachers’ interpretation and application of research findings. The time and effort involved in this adaptation can add to teachers’ already heavy workload, potentially becoming an additional burden or leading to research being ignored.

Considering the Chinese teaching context, the requirement of integrating educational research into teaching plans, policy decisions, institutional frameworks, and public discourse is increasing (Ministry of Education of the People’s Republic of China., 2019). This may force the whole system to be involved

in the movement of using research evidence to implement effective teaching practices. However, the investigation of research evidence use is still at the beginning stage. Specifically, no study has been found that evaluates the impact of teachers' use of research evidence on primary school students' first-language learning outcomes in China.

This new study is therefore unique and fills a gap in current research in this area. The findings of this study will enhance current understanding of the feasibility and impact of integrating research evidence into teaching. Notably, this study conducts fieldwork in China, where the movement towards evidence-informed education is still in its early stages, providing a timely and important exploration of this emerging area.

1.3 Aims of the study

The aim of this study is to gain a deeper understanding of how teachers perceive and use research evidence, identify effective approaches for disseminating research evidence into teaching practice, and rigorously estimate the feasibility and impact of teachers' use of research evidence. It gathers accessible data on teachers' views and access to research evidence, followed by a review of promising strategies to promote evidence-use, particularly in reading literacy instruction.

Drawing on existing evidence, this study evaluates a use-of-research-evidence intervention that consists of a menu of evidence-informed teaching strategies, training, and monthly follow-ups. The intervention was carried out with primary school teachers in mainland China to examine the feasibility of teachers' use of evidence in the Chinese teaching context and to generate primary evidence on its effectiveness for improving students' Chinese literacy outcomes. In doing so, the study addresses a critical gap in understanding the effectiveness of teachers' use of research evidence in mainland China.

Importantly, rather than enforcing specific evidence-informed teaching strategies, the study focuses on teachers' own adaptation of research evidence into their practice. Teachers participating in the intervention did not receive prescriptive training from a particular programme. Instead, their implementation was guided by their own understanding of research evidence. The strategy menu provides options and scaffolds teachers' ability to critically select, adapt, and apply relevant research findings to their teaching.

1.4 Research questions

There are five main research questions and sub-questions proposed as follows:

1. What are the current situation and emerging trends in school teachers' access to research information on education worldwide?

- What is the situation of teachers' access to research information on education around the world?
- What is the trend of teachers' access to research information on education around the world?

2. How do teachers perceive and engage with research evidence to improve teaching practice?

- How do teachers value the usefulness of research findings?
- What is the strongest predictor of teacher-perceived usefulness of research evidence for improvement of teaching?
- To what extent can school encouragement and colleague discussion support teachers' confidence in using research evidence?

3. What is the best approach to facilitate teachers' use of research evidence, particularly in teaching literacy?

- What is the impact of teachers' use of research evidence on students' literacy outcomes?
- What is the most promising approach to encourage teachers' use of research evidence, in order to improve students' literacy outcomes?

4. What are the impacts of teachers' use of research evidence on students' literacy outcomes, teachers' perceptions and skills in using research evidence, and classroom teaching practice in China?

- To what extent does teachers' use of research evidence improve students' Chinese literacy outcomes?
- Is there an impact of teachers' use of research evidence on student's enjoyment of reading?
- Does the impact of teachers' use of research evidence on students' literacy outcomes differ according to students' demographic characteristics (sex, age, ethnicity, and hukou)?
- What impact does the use-of-research-evidence intervention have on teachers' perceptions of research evidence?
- To what extent does the use-of-research-evidence intervention make a difference to teachers' skills of using research evidence?

5. To what extent is the intervention delivered as intended?

- Is it feasible for teachers to incorporate research evidence into the Chinese primary school context?
- To what extent does teacher compliance make a difference to the impact of teachers' use of research evidence on students' Chinese literacy outcomes?
- How can schools support and facilitate teachers' use of research evidence?
- What conditions play a role in the successful use of research evidence in classroom practice?
- How is the implementation of evidence-informed strategies different among teachers?

This study adopts different research designs to answer these research questions, combining cross-sectional secondary data analysis, a structured review, and a large-scale randomised controlled trial with integrated process evaluation.

The first two main questions are largely answered based on existing secondary datasets, TALIS (2008, 2013, and 2018) and the Survey of Teachers 2010, with a cross-sectional design. While there is bias in the sampling of TALIS and the Survey of Teachers only included about four thousand teachers in England, these remain the only available open-access sources that can provide valuable insights into teachers' access to research information and their views on the use of research findings.

The third research question is largely explored through a structured review of over two thousand studies to summarise the most effective approaches for supporting teachers' use of research evidence and enhancing students' literacy outcomes. The review follows the PRISMA guidelines, and the quality of included studies is evaluated using the Gorard SIEVE (Gorard, 2024).

The fourth and fifth questions look for new evidence on the effectiveness of teachers' evidence-use in the Chinese teaching context. To investigate the feasibility and impact of teachers' use of research evidence, a large-scale randomised controlled trial is conducted with primary school teachers in China. Prior to the main experiment, a pilot experiment is carried out to refine intervention materials and training procedures. Students' Chinese literacy test results are collected from school records. Students' enjoyment of reading and teachers' perceptions and skills towards research evidence are captured through surveys. In addition, classroom observations and interviews are conducted as part of the process evaluation to gain further insights into implementation outcomes.

1.5 Structure of the thesis

This thesis involves sixteen chapters.

Chapter 1, Chapter 2, and Chapter 3 establish the background and rationale for this study. Chapter 1 outlines the origins and development of evidence-use and the urgent need for a deeper understanding of evidence-use in teaching practice. It also presents the research aims, questions, and an overview of the thesis structure. Following the first chapter, Chapter 2 clarifies the concepts of research evidence and the use of research evidence. It also explains the importance of using research evidence in education. Chapter 3 critically reviews the literature on barriers and facilitators of using research evidence. The chapter also introduces the Chinese education system and current exploration of evidence-use.

Chapters 4 to 6 explain the research questions and corresponding designs in detail. Chapter 4 presents the design, datasets, and data analysis methods used in the secondary data analysis. Chapter 5 introduces the design and methods used in the systematic review, which includes search strategies, selection and screening, data extraction and quality appraisal. Chapter 6 consists of the methods and procedure for the randomised controlled trial, involving participant recruitment, intervention design, outcome measures, missing data, and ethical issues. This chapter also explains the procedure undertaken for the process evaluation.

The findings of this study are reported in Chapters 7 to 15. Chapter 7 describes teachers' access to research information and identifies the strongest predictors of their perceived usefulness of research evidence for teaching. Chapter 8 summarises the impact of using research evidence on students' literacy outcomes and the best approach to engage teachers with research evidence use. Chapter 9 presents lessons from the pilot experiment that may support better delivery of the use-of-research-evidence in the main experiment. Chapter 10, Chapter 11, and Chapter 12 include impact analyses in terms of students' Chinese literacy test results, enjoyment of reading, and teachers' perceptions and skills in using research evidence.

In addition to the impact analysis, this study focuses on feasibility, implementation quality, and teachers' exploration process. Chapter 13 investigates teacher compliance with the protocol and identifies pathways through which teachers successfully integrate high-quality research evidence to improve student outcomes. Chapter 14 describes views from school headteachers and research leads regarding their efforts and challenges in supporting evidence-use within schools. Chapter 15 describes the exploration process of each selected evidence-informed strategy and the difficulties teachers encountered during practice, based on classroom observation and informal interviews.

Chapter 16 is the final chapter of this thesis, which reflects on the study's limitations, summarises key findings, and discusses implications for practice. The implication part consists of feasibility, implementation quality, and sustainability of teachers' use of research evidence.

Chapter 2 What is research evidence?

Following the introductory chapter, this chapter defines 'research evidence' and the 'use of research evidence' within the context of this study. Conceptual issues such as high-quality research evidence, synthesis studies, waves of evidence-use, and types of evidence-use are discussed, based on the existing literature. This chapter also explains the potential significance of using research evidence in educational decision-making and practice.

2.1 Description of research evidence

The first issue is to have an operational definition of research evidence. Various factors shape decisions, and research findings frequently play a relatively minor role in education (Trinder & Reynolds, 2000). For schools and school administrators, other types of evidence can be helpful besides research. The Social Care Institute for Excellence produced a typology in 2003 based on Pawson et al. (2003) that listed five knowledge sources that were pertinent to social care:

- Organisational knowledge
- Practitioner knowledge
- User/client knowledge
- Research knowledge
- Policy community knowledge

They argued that all these sources contribute to an evidence base, with no inherent hierarchy among them. However, they emphasised the importance of recognising that some types of knowledge are more relevant to specific purposes than others. Users must also consider the quality and reliability of the knowledge. Additionally, since these resources form the evidence basis, users may aggregate all forms of evidence.

When creating and refining instructional strategies, educators may use two distinct sorts of research evidence. The first type comes from teacher practitioner inquiry, such as when teachers participate in action research projects and/or lesson studies and use data at the student or school level to help make decisions (e.g., BERA-RSA, 2014, p.11). The second is obtaining and using external research evidence, which includes primary studies and evidence reviews that are frequently created by the research community. To avoid confusion, this study will specifically focus on research evidence from external research communities.

Research Excellence Framework (n.d.) defined research as 'a process of investigation leading to new insights effectively shared'. Scott and McNeish (2013) highlighted that research can serve multiple purposes, including identifying needs and issues, providing evidence on what works, and assessing progress towards school improvement. The second purpose, determining what works, is most often referred to in discussions of the use of research evidence. In this study, research evidence, as well as research findings, are defined as outcomes of a systematic investigative process designed to expand or refine knowledge, particularly regarding programmes and strategies that have been repeatedly or robustly shown to improve students' learning outcomes.

It is widely acknowledged that users should rely on the best available evidence, often referred to as high-quality or robust evidence (Cooper et al., 2009; Hollands et al., 2019; Sackett et al., 1996; Slavin, 2002). However, not all studies meet the standards of sufficient quality and reliability. Building on the description of research evidence, many scholars have debated what constitutes high-quality research, the criteria for assessing research quality, and the current state of primary research. These discussions also highlight the risks of using research without critically evaluating its quality, emphasising the potential consequences of relying on flawed or unreliable findings. This following section considers three key challenges related to evidence: defining high-quality research, the lack of strong primary studies, and the limitations of systematic reviews.

2.1.1 What are high-quality primary studies?

Various frameworks have been proposed to assess the trustworthiness of research findings across fields. For primary research, some existing frameworks for judging research quality consider research designs with randomised controlled trials (RCTs) listed as the 'gold standard'. For example, the Maryland Scientific Methods Scale has five grades of quality ranging from comparisons between naturally occurring groups with the lowest quality to full randomisation with the highest quality (Farrington et al., 2003). Similarly, Guyatt et al. (2008) proposed GRADE for clinicians that classifies evidence quality into four categories: high, moderate, low, and very low. In this system, randomised trials typically start as high-quality evidence, while observational studies begin as low-quality evidence.

However, the determination of what constitutes the best research depends largely on the specific questions being addressed. Not all research questions require RCTs to generate valid and meaningful answers. For instance, a longitudinal design with observational data could be sufficient to understand the process of implementing a certain intervention and assess the fidelity of its implementation. Evans

(2003) introduced a framework for ranking evidence evaluating healthcare interventions with consideration of three major research purposes, involving effectiveness, appropriateness, and feasibility. According to Scott and McNeish (2013), this framework could be used in education studies as well. For estimating the impact of an intervention at the primary study level, Evans rated RCTs and observational studies as good, while descriptive studies, case studies, expert opinions, and studies of poor methodological qualities were categorised as of poor quality. Evans acknowledged interpretive studies and descriptive studies for testing appropriateness and feasibility.

Hierarchies based solely on study design can overlook other critical factors, such as sample size, missing data, and the quality of measurements. A study with small sample size, significant participant attrition, or poorly standardised and non-independent outcome measures may produce results that are not trustworthy, even if the research design is an RCT. Gorard (2024) developed a research quality judgement tool, a 'SIEVE', with consideration of design, sample size, attrition rate, and measurement quality. The Gorard 'SIEVE' does not directly appraise the randomised controlled trial as a 'gold standard' but considers whether the design is appropriate for the research question. It provides more detailed criteria for other dimensions of research quality, offering a more balanced and comprehensive evaluation of individual studies. Where appropriate the SIEVE is used in this thesis.

2.1.2 Uncertainty of existing primary studies on 'what works'

Research quality forms the foundation for the use of research evidence to guide policy and practice decisions based on 'what the evidence says'. While there is some debate about 'what good evidence is', there is also a lack of trustworthy and conclusive evidence regarding what strategies are truly effective for school improvement and effectiveness. Many of these approaches have not been robustly or independently evaluated (See & Gorard, 2020). It appears that much of the existing research in education does not meet high-quality standards. Twenty years ago, key UK stakeholders such as policymakers, practitioner bodies and senior researchers agreed reasonably well that education research is not good enough for real-life impact (Taylor, 2002). The same concern still persists today.

Researchers conducting reviews often find that the numbers of high-quality studies are limited. For instance, See et al. (2021) reviewed 56 studies on the impact of computer-assisted formative assessment on students' learning outcomes, but only 22 were rated with a trustworthiness of 2 padlocks (out of 4) using the Gorard 'SIEVE', and the rest were worse, indicating that most studies were of very low quality. Similarly, in a review of metacognitive interventions, only three out of 29

studies received the highest rating of two padlocks (Wongdaeng & Higgins, 2023). Even for widely discussed topics such as computer-assisted formative assessment and metacognitive interventions, the number of high-quality studies on their effectiveness remains limited.

Interestingly, most interventions have more apparent observable effects on students in weaker studies, which may be more misleading to evidence users due to their 'promising' effect sizes. Slavin (2008) found that for the same intervention in mathematics, the median effect size for studies with smaller sample sizes (fewer than 250 students) was +0.21, while studies with larger samples yielded a smaller effect size of +0.11.

Identifying and recognising high-quality research is essential for advancing knowledge and ensuring that the evidence used in practice is valid. The use of low-quality evidence can lead to wasted time, effort, and investment, which raises ethical issues. Individual teachers may invest in ineffective practices, ultimately damaging their students' learning outcomes. At the school or regional level, the consequences extend beyond the time and energy of users. Poor decisions based on unreliable evidence can lead to misallocation of resources, wasted investments, and far-reaching negative effects on broader educational outcomes.

2.1.3 Are synthesis studies trustworthy?

While some primary studies with high quality can provide reliable research findings, they can hardly provide definitive and conclusive answers to real-world educational questions on their own. Thomas (2021) argues that the disappointing results frequently observed in evaluations are not solely due to the ineffectiveness of the interventions but to the powerful contextual variables inherent in educational settings. There may be several pieces of similarly good research that contradict a single high-quality study (Gorard et al., 2020).

Since it is uncommon to directly replicate experiments in education, it might be more appropriate to summarise all of the relevant evidence on any issue. Consequently, systematic reviews are widely valued for their ability to synthesise evidence and offer more conclusive insights (Evans, 2003; Higgins et al., 2016). These syntheses may be conducted by academics, organisations and government departments. Traditionally, the emphases have been on scrupulous searching to try and collate all relevant evidence, and on transparency so that the findings can be checked or even replicated. These are ideals. In education and social policy, more generally, there is an increasing number of syntheses

of existing evidence intended to summarise what is known on any topic in a way that helps to promote the use of evidence or identify evidence gaps for future research.

Though this approach offers a more comprehensive view of the existing evidence base, it also draws attention to a more sobering reality that many interventions, even those widely promoted or theoretically promising, show limited or no measurable impact. For example, the majority of educational interventions synthesised by the WWC and the EEF Toolkit have been reported as ineffective when summarising the overall effect sizes. This reflects the limitations of individual interventions and raises questions about the reliability of the reviews themselves at the same time.

The inconsistent judgement of study quality within syntheses might be misleading. It is necessary to make reviews as valid, unbiased, and meaningful as possible for their intended purpose (Slavin, 2008). Wadhwa et al. (2024) examined 12 educational clearinghouses to evaluate their effectiveness criteria and assess the consistency of their programme ratings. Among the 1,359 programmes analysed, only about 30% received similar ratings across the clearinghouses.

Some syntheses have simply aggregated the effect sizes together. Poor quality studies tend to produce more extreme 'effect' sizes than better-designed ones and so will dominate the results when all are aggregated. There has been considerable criticism of meta-analyses like that of Hattie (1992) or the EEF Toolkit in England, indicating that these analysts have engaged in unsuitable aggregation (Bergeron & Rivard, 2017). It has been demonstrated that giving equal weight in synthesis to studies of different quality yields invalid and misleading results (Gorard & Chen, 2025).

2.2 Use of research evidence in this study

There has also been an increasing emphasis on applying evidence in recent years. Various terms and concepts describe the process of evidence-use such as evidence-based, evidence-informed, and evidence-led. Some researchers interpret use of research evidence differently, offering distinct definitions for evidence-based, research-based, and evidence-informed practice, while others use these terms interchangeably (Nelson & Campbell, 2017).

This study predominantly employs terms like 'use', 'adapt', and 'incorporate' research evidence into (teaching) practice. In this context, teachers are both 'encouraged to' use and actively 'engaged with' research evidence. The intervention in this study is designed to 'facilitate' teachers' use of research

evidence. These terms emphasise teachers' autonomy in using evidence, allowing them to interpret research findings themselves and adjust their teaching practices to suit specific teaching contexts.

Teachers applying strategies based on promising research evidence is referred to as evidence-informed teaching practice, with the strategies themselves termed evidence-informed teaching strategies. According to England's Department for Education (DfE), evidence-informed practice is defined as practice that is influenced by robust research evidence (Coldwell et al., 2017). In Nelson and Campbell (2017), evidence-informed practice may emphasise that the research evidence is just one factor that influences educational decisions with educators needing to apply professional judgement.

In this study, 'dissemination' refers to the process of delivering research evidence to teachers through various forms and activities. There is an ongoing debate about whether dissemination involves merely communicating research evidence in a one-way manner to improve accessibility, or engaging users collaboratively with the evidence. For instance, while schools are increasingly using research evidence due to a stronger evidence base for effective approaches, this does not necessarily mean they are adopting these practices primarily because they are evidence-based. More research is needed to understand how research can be effectively channelled to encourage the adoption of evidence-based practices for their proven benefits. To clarify these issues, this section explores different types of evidence-use.

2.2.1 Transformation of research evidence

Research evidence is often transformed into various formats to enhance its accessibility and usability. White (2020) proposed an evidence architecture based on acknowledge translation pyramid that comprises nine stages to explain how plain evidence can be translated into useable and accessible formats. From the base of the pyramid upwards, these include data, primary studies, systematic reviews, databases, evidence maps, evidence platforms, evidence portals, guidelines, and checklists. White suggested that checklists, guidelines, and evidence portals can support decision-making without requiring users to read research papers. However, it is hard for users to assess the quality of the underlying evidence with these refined summaries.

Beyond how evidence is presented, researchers have also examined the different ways in which it is used in practice. In some literature, the use of research evidence in practice is generally defined as either (a) the implementation of programmes that have a solid research evidence basis in schools,

such as adopting a 'proven' curriculum, or (b) the integration of broadly interpreted evidence into decision-making processes (Hood, 2003). Langer et al. (2016) categorised research evidence use into instrumental use and conceptual use. The instrumental research evidence use refers to the direct use of research evidence (e.g., Doabler et al., 2014; Van Staden, 2011), while the conceptual use highlights its 'enlightenment function', which emphasises users' awareness and understanding of research evidence (e.g., Cain, 2015b; Rose et al., 2017).

2.2.2 Two dimensions of using evidence

While previous classifications about evidence transformation provide valuable insights into the various forms of evidence and its use in practice, they do not fully integrate the format of evidence with the level of engagement between researchers and users. Gorard et al. (2020) suggested two dimensions of evidence-use, which categorised evidence-use by dissemination approach and evidence formats (Table 2.1). This framework will also be used later in the structured review to categorise and describe the included studies.

Table 2.1 Two dimensions of evidence-use

	Passive transfer	Engagement in transfer	(inter) Active transfer
Plain evidence	e.g., open access to journals	e.g., journal clubs	e.g., practitioner inquiry
Modified evidence	e.g., EEF toolkit to practitioners	e.g., think tanks	e.g., internship, research schools
Engineered evidence	e.g., lesson plans	e.g., hotlines, helpdesks	e.g., legislation for population measures

The framework categorises studies based on the format of presenting research evidence and user engagement in the process of evidence-use into nine snapshot cells. It can cover various studies that promote using research evidence in a broad sense involving studies that focus on policy decisions, like government legislation or local funding and practical education strategies, such as classroom organisation and language instruction.

The format of evidence is separated into three categories: plain evidence, modified evidence, and engineered evidence. Evidence such as research papers of primary studies or systematic reviews from academic journals is categorised into plain evidence without revision. For instance, in the study of See et al. (2016), teachers were provided with a systematic review of the power of feedback to improve teaching and learning of Hattie and Timperley (2007).

Despite search engines such as Google Scholar, in recent years some databases have focused on gathering education systematic reviews. For example, the International Database of Education Systematic Reviews, developed by Oxford University, is a freely accessible electronic resource that provides published systematic reviews in the field of education (https://idesr.org). It also serves as an online platform where planned and ongoing systematic reviews can be registered, searched, and viewed.

While plain and raw evidence can accurately present the research findings and provide enough information to judge its trustworthiness, it requires skilful users to search, read and summarise the evidence, and then implement the changes. Transferring research evidence into modified or engineered form can be easier to read but the simplification of the raw evidence may provide inaccurate information. Evidence that is revised but retains noticeable research elements is modified evidence such as EEF Toolkit and WWC. Erkan (2023) provided an evidence kit identifying a list of promising approaches to improve literacy in primary schools in the form of handouts to teachers.

Some researchers develop more practical resources, with engineered evidence that could hardly be recognised as research such as lesson plans and curricula developed with research evidence. Kim et al. (2017) delivered a package of READS summer literacy programme which had been evaluated for its effectiveness in prior research for teachers, including the use of teacher voice and lesson activities. Similarly, Doabler et al. (2014) delivered an evidence-led mathematics curriculum to kindergarten teachers, where it is assumed that teachers can follow the curriculum without being aware of the audit trail leading back to the evidence for its use.

Another dimension in Table 2.1 is the transfer approach according to the role played by researchers and developers, involving passive transfer, engagement in transfer, and (inter)active transfer. When the research evidence is directly delivered to teachers without activating effort, the transfer is passive. As an example, Lord et al. (2017a) provided schools with research summaries and evidence-based resources to improve teaching. However, the passive transfer requires resources and training to be sufficiently detailed. Otherwise, teachers may be left unclear about what actions to take. Moreover, passive reception of evidence may affect the sustainability of evidence-use, as teachers might not actively incorporate other research evidence once the experiment concludes.

Disseminating research evidence with some activities such as workshops to help teachers better understand and apply evidence is an approach where researchers engage in transfer. This kind of

transformation of research evidence focuses on the communication between researchers and research users. Teachers are usually aware of the evidence that they are using. Van Staden (2011) delivered a series of evidence-based instructional strategies, where instructors were trained in a three-hour event to apply these strategies, to support students who learnt English as a second language. Cain (2015b) investigated how teachers transformed research-generated knowledge into pedagogical knowledge by presenting research about teaching gifted and talented students. Teachers were also invited to express their understanding of the papers and supported adapting findings into practitioner research projects of their own devising for 12 months. Rose et al. (2017) estimated the impact of participating in Research Learning Communities, in which experts discussed research evidence with evidence champions from schools.

Some studies emphasise how research may solve real-life teaching issues. Evidence for the Frontline provided an interactive approach for educators' evidence-use for science teaching with an online broker (Lord et al., 2017c). Teachers and school leaders formulate their own questions related to enhancing teaching and learning and share them on an online platform. An online broker then connects these questions with academic researchers for responses or directs the teachers to relevant published evidence and existing answers provided by the service. Tortorelli and Bruner (2022) developed a research-practice partnership to address teacher-generated questions related to how children learn to spell words.

When the roles of research provider and user overlap, the transfer of evidence becomes interactive. Research schools exemplify this approach by enabling researchers and teachers to collaborate on research directly relevant to classroom practice (Hinton & Fischer, 2008). The research process involves formulating questions, testing new techniques through fieldwork, entering results into a national database, using classroom outcomes to modify local practices and research directions, and distributing findings. This cyclical approach integrates theory and practice through collaboration between researchers and practitioners. In England, the Research School Network, funded by EEF, demonstrates this model by supporting other schools in translating research evidence into practice through training, exemplification, and school-to-school collaboration (https://researchschool.org.uk).

In summary, evidence-use encompasses various forms, considering both the transformation process and the format in which research evidence is delivered. This diversity allows for the broad utilisation of research evidence but also complicates its integration into practice, adding to the uncertainty of its effectiveness. Chapter 8 will classify and describe the impact of teachers' use of research evidence on

students' literacy outcomes based on different types of evidence-use, to identify the most promising approaches for facilitating teachers' use of research evidence.

2.3 Why does the use of research evidence matter?

Research evidence plays a role in the educational system at the levels of governmental policy, school, and individuals with the potential to support effective decision-making and practice and inspire innovation in teaching practice. While studies about the benefits of using research evidence are limited, the use of research evidence may still make a difference to teaching practice and influence school performance.

2.3.1 Role of research evidence in school improvement

Research matters for school improvement across national, school, and individual levels. At the policy level, research can reflect on the policy and reveal its potential limitations and problems, or it may provide direct support for policymaking, focusing on solving practical, real-world issues (Lingard, 2013). It has the potential to highlight critical issues for policymakers and suggest solutions, prompting adjustments and resource allocation at the national level. Mayer (2021) employed teacher education policy as an example and summarised that research evidence intervened in policy primarily by shaping the discourse, justifying reforms, and influencing regulatory frameworks. Researchers have also adopted various strategies to enhance the use of research evidence in policymaking at both national and local levels (e.g., Bartholo & Koslinski, 2020).

At the whole-school level, research provides insights into what works broadly and for specific student groups, such as those eligible for free school meals or those identified with special education needs. Research evidence can serve as a holistic approach to educational decision-making that critically appraises different forms of evidence before key improvement decisions are made (Brown et al., 2017). It highlights the factors with the greatest impact on student outcomes, helping to guide priorities, focus, and investment. These insights are particularly valuable in guiding resource allocation, such as through initiatives like the Pupil Premium funding in England.

In England, the Pupil Premium funding gives schools the flexibility to allocate resources. According to Sutton Trust (2019), 74% of senior leaders considered research evidence, with 70% of secondary leaders using the EEF Toolkit, up from 63% last year. Teachers using research as reference were more likely to report effective Pupil Premium spending (46% vs. 32%). These findings suggest that research

evidence plays a role in decision-making at the school level, potentially leading to more informed and effective use of funding.

The EEF, supported by the UK government, was established in 2010 particularly to improve the achievements of disadvantaged students (See, 2020). Its reports often estimate the effect sizes of educational programmes, particularly for sub-groups such as free-school-meal eligible students. By estimating the effect sizes of educational programmes for specific sub-groups, the EEF provides schools with evidence to tailor their funding decisions and support strategies to meet diverse student needs, thereby enhancing overall school performance.

Research evidence may also provide inspiration for educational innovation for school improvement. With the increasing pressures of school improvement and accountability, many school leaders have sought ways to enhance their schools' performance. In this context, using research evidence was seen as a way to make their schools stand out (Graves & Moore, 2018). This approach helped them set an agenda that aligned with national-level policies and met the expectations of inspectors.

For teachers, research evidence serves as a valuable tool to help them make informed instructional decisions, by offering a sound understanding of what are promising approaches that support student learning. It provides insights into which teaching approaches are most effective for specific groups of students (Gersten, 2001; Scott & McNeish, 2013). Teachers may adapt and integrate teaching approaches or programmes with strong evidence of effectiveness into their practice to fit their unique teaching contexts and students' characteristics.

Though many educational interventions lack evidence or have been proven ineffective (see https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects), recognising this ineffectiveness is also crucial. It informs teachers about what practices to avoid. This process of engaging with evidence fosters critical reflection on their teaching practice, allowing teachers to improve their teaching qualities. A recent survey-based study involving 534 teachers across Catalonia (Spain), Poland, and England indicates that engaging with research empowers teachers to question traditional beliefs and adopt innovative teaching strategies (Ion et al., 2024).

2.3.2 Potential benefits of evidence-use in teaching practice

Evidence-use is sometimes required at the policy level. For instance, No-Child-Left-Behind in the US. specified the use of scientifically based research to inform instructional and school reform decisions,

requiring districts to seek and apply evidence from rigorous studies on the effectiveness of programmes and practices (US Department of Education, 2002). In England, the Department for Education required schools to justify their use of Pupil Premium funding with research evidence. Consequently, most schools maintain websites detailing the interventions they have implemented and the supporting evidence base.

Many researchers argue that evidence-informed schools and colleges are an essential feature of effective education systems. Using research evidence might be beneficial for teachers to enhance the effectiveness of learning processes, innovate new approaches to pedagogical issues, and improve school quality (Diery et al., 2021; Mayer, 2021). Teaching and instructional quality is one of the most important school-based factors for student learning and attainment and has the potential to close the attainment gap (Blömeke et al., 2016; Slater et al., 2012; Wiliam, 2013). If the use of research evidence can be shown to have a positive impact on teaching, it would contribute directly to school improvement.

Teachers as users of research may receive direct benefits of research evidence use such as updating their knowledge, evaluating their approaches to teaching and introducing new resources and schemes of work. Beyer et al. (2025) provided pre-service teachers with a review article, a Clearing House-style format, and a chat-based analysis tool, finding that all three resources supported participants' research-informed reasoning and practice. Similarly, Cain (2015a, 2015b), in a 12-month collaboration with secondary school teachers, found that teachers demonstrated greater willingness to experiment, enhanced critical thinking skills, a deeper understanding of evidence, and heightened ethical awareness. Flynn (2019) indicated that the online platforms and action research potentially engaged teachers with research, increased teachers' knowledge, and changed teaching practices.

Teachers who engage with research evidence may adapt their teaching practices informed by reliable, scientific findings that may be relevant to students' performance. A study based on 73 English primary schools indicated a positive correlation between teachers' perceptions of research use climate and student performance with a coefficient β of 0.13 (Groß Ophoff et al., 2023). Clarke et al. (2011) stated that a 120-lesson evidence-based kindergarten curriculum improved at-risk students' mathematics attainment and effectively narrowed the attainment gap with their not-at-risk peers. Wiggins et al. (2019) claimed a small positive impact on students' English (+0.05) and mathematics (+0.09).

2.3.3 Struggles for research evidence

Although research evidence is not the sole resource supporting decisions in education policy and practice, it undoubtedly offers valuable insights and guidance. However, currently, policymaking often lacks a strong foundation in research evidence, and practitioners' implementation is not always supported by it either, which may risk a waste of financial investment and resources. This places research evidence in a somewhat precarious position.

For instance, a number of British policy initiatives have firmly embedded phonics in the curriculum for early reading development, while some uncertainty about the most effective approaches to teaching reading remains (Torgerson et al., 2019). In a school district's central office in the US, research is often used symbolically during textbook adoption, with decisions primarily driven by experience and financial considerations (Farley-Ripple, 2012). References to research are vague or made after decisions had already been made, while decisions directly informed by research rarely happen.

Similarly, while recent years have seen numerous documents advocating educational innovation, media coverage often overlooks the importance of research evidence. Even when the language in these documents appears promising and reliable, they are likely to prioritise the appeal and perceived credibility of the innovation over its research-based effectiveness. As Carrier (2017) found in an analysis of 63 documents, research citations relating to the effectiveness of innovations appeared in only eight, and none provided an in-depth exploration of studies testing the innovations' effectiveness.

At the practice level, many teaching practices and school-wide interventions are based on experiential judgment and prevailing trends rather than being grounded in research evidence. When selecting interventions, school leaders tend to favour programmes that are popular or align with past or current practices. However, these choices often lack robust evidence to demonstrate a causal impact on improving learner outcomes (Brown & Greany, 2018).

This issue is further illustrated in a study by Pegram et al. (2022) of 138 SEN interventions implemented across ten schools in Wales. Only 30% of these interventions had some evidence of a positive impact on student outcomes, while 67% had no published evidence, and 3% had causal evidence indicating they were ineffective. Even within this 30%, most evidence was weak, with 11% rated as moderate to high quality, 19% as low quality, and just two interventions (FRIENDS and Accelerated Reader) were backed by high-quality replicated evidence. Furthermore, a follow-up conducted a year after the

report was shared with schools revealed that not all schools used this information to make informed decisions about their interventions.

2.4 Conclusion

In brief, this chapter has clarified the definitions of research evidence and its application, helping to prevent conceptual misunderstandings. The study focuses on the use of high-quality research evidence to inform teaching practice and emphasises its critical role in education. Furthermore, it highlights that currently education policymaking and practice rarely have a ground research evidence basis. It would be necessary to explore the barriers and facilitators of evidence-use, understand users' needs, and identify effective strategies for sustaining high-quality implementation, which will be discussed in Chapter 3.

Chapter 3 Barriers to and facilitators of using research evidence

Using research evidence has become an important basis for informing teaching decisions. However, there are still many challenges to its effective use in practice. If the use of research evidence is to be truly feasible and impactful, it requires concerted efforts from researchers, practitioners, and the broader evidence-use environment. This chapter explores the potential barriers to and facilitators of using research evidence in education, drawing on insights from existing literature. It also introduces the educational context in China, where the fieldwork for this study was conducted. Finally, it critically reflects on limitations of previous studies on teachers' use of research evidence.

3.1 How to use research evidence?

While many researchers argue that evidence-informed schools and colleges are an essential feature of effective education systems, there is still uncertainty about the best ways to encourage and implement evidence-use. Researchers keep proposing frameworks for the process of using research evidence. However, existing reviews suggest that our understanding of how to optimise evidence-use remains limited and insufficiently tested.

3.1.1 Models and frameworks

The research-practice gap involves assumptions and perspectives, including the usefulness of research products, nature and quality of research, problems research addresses, structure-process-incentives, and specific relationships between communities (Farley-Ripple et al., 2018). Frameworks have been proposed to guide the process of using research evidence. For instance, Spencer et al. (2012) introduced a framework to guide practitioners from formulating practical questions to selecting, adapting, and implementing treatments to achieve positive outcomes. This framework emphasises using the best available evidence, considering client values and contexts, and applying professional judgment in practice.

Both researchers and research users are responsible for the use of research. Starting with the selection of trustworthy research evidence, researchers and research users should think critically about well-designed research to enhance the overall trustworthiness and utility of their findings (Owen et al., 2022). Higgins (2020) outlined three key responsibilities for research and research providers: ensuring that evidence is accurate, accessible, and actionable, while also being appropriate, applicable, and acceptable for policy and practice users. Higgins emphasised concerns about actionability and

appropriacy, particularly its relevance to classroom application and alignment with specific curriculum contexts.

The process of using research evidence extends beyond researchers and users to include mediators who facilitate the connection between research production and practice. Additional involvement in intermediary stages may also be necessary, for teachers to use research evidence sustainably over the long term. Mediators are also important in bridging the gap between research producers and users (Campbell & Levin, 2012). These intermediaries can take various forms, including research centres, think tanks, lobby groups, charities, professional organisations, providers of professional development, parent groups, and publishers of media in both print and digital formats.

Gough et al. (2022) introduced the concept of an 'evidence ecosystem', which includes decision making, research production and some engagement between such decision making and research production. The system involves players and interests such as actors, perspectives, issues, questions, and power. This ecosystem emphasises the alignment of decision-makers' needs with research questions and their application to specific contexts, facilitated through active engagement between decision-making and research production.

Bringing individual, organisational, and system levels factors together, Rickinson et al. (2021, 2022) synthesised a Quality Use of Research Evidence Framework with 112 studies. Facilitators of quality use of research evidence involved individual enabler such as skillsets, mindsets, and relationships and organisational enablers such as leadership, culture, infrastructure. The framework also emphasised system-level enablers, referring to the complex interactions and interdependencies across the education sector that support the thoughtful engagement with and implementation of appropriate research evidence.

While researchers propose frameworks that outline how research evidence can be used in practice, some centres have taken an active role in facilitating evidence-use. For example, the What Works Network combines ten independent What Works Centres in the UK, proposed six key activities for 'generating evidence, translating that evidence into relevant and actionable guidance, and helping decision-makers act on that guidance' (Cabinet Office & HM Treasury, 2018, pp. 5–6) as shown in Figure 3.1.

In addition to producing primary evidence and synthesising existing research, the What Works Centres actively translate evidence into accessible and easy-to-understand formats as an intermediary organisation. They also provide users with adequate support during the process of receiving and implementing research evidence. Furthermore, the centres encourage practitioners to evaluate their own practices, contributing new insights and evidence to supplement primary studies.

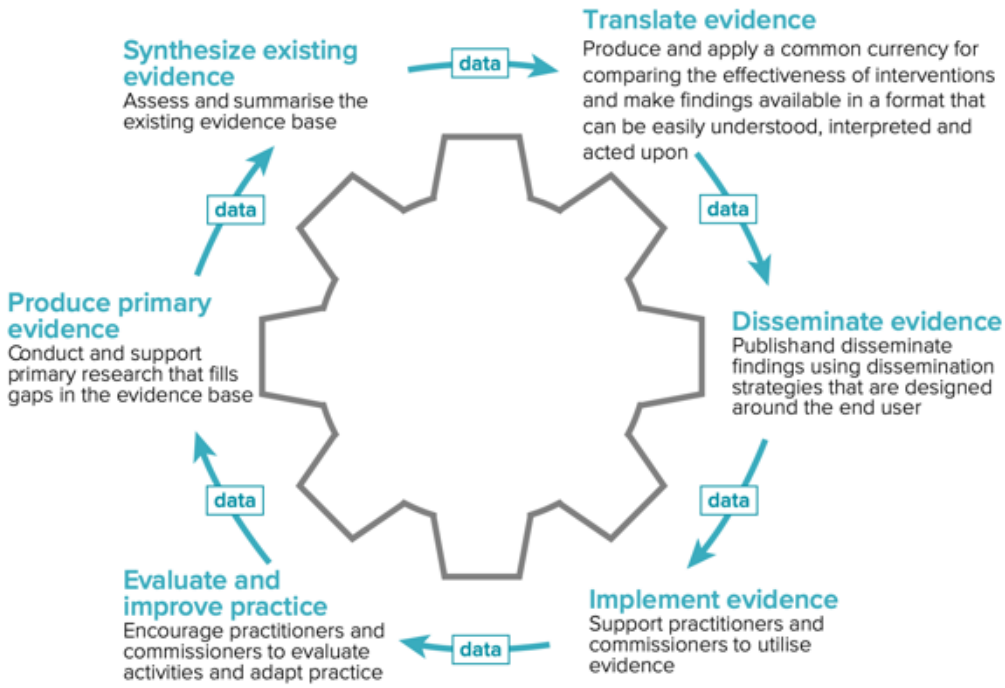


Figure 3.1 What Works Centres' areas of activities (Cabinet Office & HM Treasury, 2018, pp. 5–6)

3.1.2 What is the best approach to using research evidence?

Systematic reviews have synthesised existing studies that examine the effectiveness of various approaches to promoting the use of research evidence. Overall, passive transmission of research is largely ineffective unless interventions are heavily engineered, often obscuring their evidence base, whereas active engagement can facilitate the use of research evidence (e.g, Bero et al., 1998; Gorard et al., 2020). Notably, across all areas of public policy, the most promising approaches seem to be those that embed evidence use into everyday practice, such as through curriculum requirements or legal mandates.

In education, Erkan (2021) investigated the best ways to disseminate research evidence to teachers based on 25 studies, only two of which had high quality. The review suggested that passive dissemination, such as emailing evidence-based materials to teachers was ineffective but encouraged active multi-component approaches. However, knowledge users rarely participate in the

dissemination and application stages of research (Sengalrayan & Harvey, 2023). Researchers might be following a linear research process, with dissemination engagement occurring after publication, thus leading to underreporting.

In summary, while frameworks and reviews provide valuable insights to guide the use of evidence, challenges remain in translating evidence into effective practice. The diversity of approaches and their varying levels of effectiveness underline the need for a more systematic and rigorous approach to investigate the conditions, methods, and processes that can maximise the benefits of using research evidence in education. On the other hand, beyond the impact of using research evidence, there are barriers faced by users who seek to apply it in educational settings.

3.2 Use research evidence from users' views

The effective use of research evidence in practice is further constrained by various barriers faced by evidence users. In the clinical field, Abu-Odah et al. (2022) revealed that the primary individual-level barriers were insufficient knowledge and skills to conduct, organise, utilise, and evaluate research literature from ten existing reviews. Similarly, across all areas of public policy, users' skills and attitudes were emphasised as key barriers (Gorard et al., 2020). In school improvement, Scott and McNeish (2013) summarised seven facilitators and barriers, involving access, relevance, credibility, engagement, usability, a combination of time, skills, and confidence, and organisational support.

3.2.1 Are teachers aware of the value of research evidence?

For teachers, knowledge from rigorous research can be used to enrich, explain, and justify professional actions and judgments, as well as to stimulate thoughtful deliberation on one's practice. Some teachers regard research as a resource to enhance their understanding of educational theories, providing valuable insights to inform their teaching practices (e.g., Ion et al., 2024; Sato & Loewen, 2019).

There is some evidence suggesting that teachers seem to hold positive views on the value of research evidence. Kostoulas et al. (2019) surveyed 257 secondary and tertiary teachers in Austria, finding that most educators demonstrated a Growth mindset and regularly engaged with research findings, primarily through professional journals and online resources. Booher et al. (2020) collected survey data from 452 STEM teachers in K-12 in the US who voluntarily enrolled in a week-long integrated STEM professional development. 71% of the participant expressed that using research would not be a barrier to their teaching creativity, which acknowledged the value of research evidence for practice.

Teachers believe that research findings have a practical impact on their teaching methods. A study surveying 250 high school teachers in Turkey showed positive views on research for classroom practices and also wanted to participate in research processes (Beycioglu et al., 2010). Similarly, Brown and Greany (2018) surveyed 696 primary school practitioners working in 79 schools in England, with 76% of teachers choosing to strongly agree or agree that research plays an important role in informing their practice. 86% indicated that information from research was useful in helping them apply new classroom approaches.

However, there is some resistance and reluctance to use evidence. Education is a complex field with many 'grey' areas, making it hard to distinguish myths from facts. Critics of educational myths face the risk of being perceived as unempathetic or pedantic, yet their role in promoting evidence-informed practices is vital to prevent the proliferation of harmful educational myths (Siegel, 2024). Practitioners may resist using research evidence because it often challenges their long-held beliefs and practices, leading to defence mechanisms like rejection or misinterpretation.

Informal knowledge, gained from life experience, is a primary influence and likely to be prioritised over theory and research in teachers' pedagogical decision-making in early childhood education (Hedges, 2012). Teachers' professional expertise is mainly gained from their daily teaching practice rather than research findings (Weimer, 2008). Teachers may think of research as a 'threat' which challenges their expertise and adds extra workload. They are sceptical and reject using educational research because they think that either the research evidence is not contextualised enough for specific teaching or that educational research is not scientific (Sato & Loewen, 2022; Vanderlinde & van Braak, 2010). Some teachers believe that research does not directly translate into their classroom, particularly if their school or the students that they work with differ significantly from the research context (Willingham & Daniel, 2021).

More importantly, although studies highlight the importance of teachers' positive views on research evidence are predictive of their willingness and ability to use it in daily teaching (e.g., Lysenko et al., 2014), there remains uncertainty about the actual adoption of evidence-based practices. A consistent gap between how much teachers value the use of research and how much they use research in their daily practices occurs (Procter, 2015). In England, a survey by Walker et al. (2019) of 1,670 teachers revealed that, despite generally positive attitudes towards research evidence, its actual use in practice

was limited. Teachers had limited interest and confidence in using knowledge from research findings to support teaching (van Schaik et al., 2018).

3.2.2 Access, time, and skills

Research evidence is just one of many factors that influence teachers' decision-making. The absence of research evidence in teaching practices is not necessarily due to teachers' rejection of it. Instead, teachers may face challenges during implementation that their own skills or resources cannot resolve.

Engaging teachers with research evidence is always challenging with insufficient preparation, lack of reinforcement, competing demands, the absence of evidence-based practice culture, challenges in maintenance and generalisation, and the breadth of expertise required (Scheeler et al., 2016). At teachers' individual level, they often encounter obstacles such as limited and inaccessible research resources, confidence in skills and time needed to identify research evidence, and the ability to understand and adapt research knowledge into practice (Ion et al., 2024; Sato & Loewen, 2022; van Schaik et al., 2018).

Access to research evidence is fundamental in the initial stage of engagement. It is necessary to provide teachers with effective structures, systems and resources that facilitate research use and the sharing of the best practices (Brown & Zhang, 2017; Williams & Coles, 2007). Teachers' willingness to engage with research evidence may be hindered without physical access, especially if teachers are not sure about where they can get adequate research findings.

It is necessary for teachers to make sense of research evidence before they judge its quality and decide how to apply it in their specific teaching contexts. Thomm et al. (2021) conducted path analysis among 2,549 secondary mathematics teachers, indicating teachers who are confident in their interpretation skills are more likely to consult research findings for their teaching practice and attach more value to the usefulness of research evidence.

However, understanding research findings may not be as easy as using general information for teachers. Available research is often published in academic language and journals (Procter, 2013). Teachers' limited evidence literacy may undermine their confidence in defining information needs, locating relevant sources, and effectively organising and synthesising information. Teachers often struggle to accurately interpret inferential statistics and effect sizes, underscoring the need to

enhance their research literacy for a better understanding of research findings and evidence-based decision-making (Schmidt et al., 2023).

Identifying 'high-quality' research evidence poses an additional challenge. Use of evidence in education may have varied interpretations of research quality among users, such as emphasising scientific rigour, cumulative theory, or underdeveloped concepts of research (Coburn & Talbert, 2006). School-based decision-makers often prioritise evidence from organisations or contexts that resemble their own in terms of demographics, location, or performance, regardless of the study design (Farley-Ripple et al., 2018; Vanderlinde & van Braak, 2010). Such tendencies risk the adoption of low-quality or non-promising evidence, which may fail to support meaningful improvements in teaching practices or student outcomes.

Additionally, all these processes of searching for, reading, interpreting, and judging research evidence require considerable time, which is often not accounted for within teachers' official workload. Consequently, they may perceive this as an extra burden.

3.2.3 Support use environment

A teacher's potential to use research evidence might be limited. Senior leadership and support are crucial for the successful integration of research evidence into practice (Burn et al., 2021; Konstantinou, 2022). Schools with a culture of evidence-use and research-informed learning environments can foster support for evidence-based change and enhance teachers' evidence-informed practices (Brown & Zhang, 2017).

An active, evidence-informed school environment and culture may stimulate teachers to 'buy into' research evidence for improving teaching. Teachers' beliefs and attitudes about the usefulness of educational research findings for teaching practice are highly related to the degree to which they feel supported in their efforts to read and use research and their voice in decision-making in the classroom (Joram et al., 2020). However, as with most studies on this topic, the research was weak.

Still, the importance of school encouragement and colleague collaboration is often ignored. About half of 131 teachers in England reported that they had received good professional development in engaging with educational research (Mills et al., 2021). Booher et al. (2020) found that 95% of 456 participating teachers reported that school leaders do not share education research with them.

3.3 Quality of implementation

Implementation is one of five crucial stages in the wide-scale diffusion of innovations (Rogers et al., 2014): (1) dissemination (conveying information about the existence of innovation to potentially interested parties), (2) adoption (an explicit decision by a local unit or organisation to try the innovation), (3) implementation (executing the innovation effectively when it is put in place), (4) evaluation (assessing how well the innovation achieved its intended goals), and (5) institutionalisation (the unit incorporates the innovation into its continuing practices).

The process of translating evidence into classroom practices is often fraught with challenges, and the quality of implementation plays a pivotal role in determining whether the intended outcomes are achieved. This reinforces the point that disseminating evidence alone does not ensure its impact on practice. James (2013) warned against simply disseminating and communicating research findings, emphasising that the real challenge often lies not in a lack of knowledge but in ineffective implementation.

Poor implementation often undermines the success of even well-supported programmes, highlighting the importance of implementation science in promoting the effective adoption of evidence-based interventions (Moir, 2018). Rojas-Andrade and Bahamondes (2019) systematically reviewed 31 studies and found that the relationship between implementation fidelity components and outcome variables was measured 273 times, with a positive association identified in 40% of cases.

Teachers may access new teaching programmes and strategies through various channels or even receive training, yet they often struggle to implement them effectively. For instance, a survey involved 1,495 language teachers in Poland reported that while teachers were well-informed about the cooperative learning principles and agreed their effectiveness, they often lacked knowledge of implementation methods and rarely applied them in practice (Abramczyk & Jurkowski, 2020).

Similar challenges emerge when engaging teachers in evidence-informed practice. The gap between research-based practices and classroom realities can limit their effectiveness (Woodard & Kline, 2015). Jacob (2017) investigated the impact of Evidence-Based Literacy Instruction (EBLI) on primary school students' reading, which found no difference in the attainment of students in grades 2 to 5 at seven primary schools in the US. One key issue was the difficulty teachers faced in incorporating EBLI techniques into their reading instruction throughout the school year. Andreassen and Braten (2011) noted that while teachers had appropriate implementation of relevant background knowledge and

reading comprehension strategies, they struggled with organising reading groups and motivating students.

This raises the question of whether any intervention or policy can be effective if it is not feasible for teachers to implement or if teachers struggle to implement it with fidelity. Therefore, understanding the factors that influence implementation quality and developing strategies to support teachers and schools are critical to addressing the broader challenges of evidence-use in education.

The implementation of evidence-informed teaching practices is relevant to diverse factors. The Implementation Components Framework identifies two key areas for successful implementation: organisational and competency drivers (Fixsen et al., 2009). The organisation drivers are decision support data systems, facilitative administrative support, and systems interventions. The competency drivers involve staff selection, pre-service training, consultation and coaching, and staff performance evaluation. Besides individual competence and organisational support, the effective use of research evidence by teachers may also require resource allocation at a more macro level. The macro level includes systemic influences such as policies, community capacity, and the availability of qualified professionals and trainers (Domitrovich et al., 2008).

A high-quality adaptation of research evidence involves complex stages. Fixsen et al. (2009) covers the entire process from exploring the current situation to implementation and maintenance. The framework describes six additive stages toward full implementation of programmes, including current situation exploration, consideration of change, initial implementation phase, full implementation, innovation, and maintenance of procedures to ensure sustainability.

While Fixsen's (2009) framework presents a comprehensive implementation pathway, its limited alignment between stages and core components constrains step-specific guidance, and it offers little support for refining or sustaining implementation, which is vital for schools. This limitation is addressed by the Quality Implementation Framework developed by Meyers et al. (2012), synthesised from 25 existing models. It kept initial considerations of the implementation setting with strategies such as self-assessment tools and capacity building. Then, it emphasises on creating a structure for implementation (e.g., implementation plans), ongoing structure once implementation begins (e.g., technical assistance/ coaching/ supervision/ process evaluation and supportive feedback mechanisms), and improving future applications (e.g., learning from experience).

Although teachers put practices in place that provide effective and beneficial support to students, it is necessary to remove ineffective practices to make room for evidence-informed practices. Wilcox et al. (2023) provided nine phases for implementation and de-implementation. The nine phases involved identifying ineffective practices, estimating the prevalence, examining sustaining factors, facilitating behaviour change, ensuring usability, conducting de-implementation trials, evaluating student outcomes, analysing cost savings, and identifying the next target for de-implementation.

Despite frameworks, empirical research has emphasised the importance of specific enabling conditions for effective implementation. High-quality implementation requires aligning policy with research on curriculum, teacher development, leadership, and assessment to support educators' professional discretion and improve literacy instruction (Woodard & Kline, 2015). From the views of intervention developers, successful implementation and sustainability of evidence-based interventions in school settings is largely relevant to administrative and teacher support, financial resources, training, alignment with school goals, visible outcomes, and strategies for managing staff turnover (Forman et al., 2009).

At the school and district level, implementation quality is linked to the quality of school planning and arrangement. Strunk et al. (2016) investigated plan quality and its relationship with implementation by rating 206 school plans submitted to the district in the first 3 years of the initiative and collected implementation plans via survey data. The findings indicated positive relationships between plan quality and reported implementation outcomes.

At the teacher's individual level, coaching might be a useful tool to facilitate teachers' practice and increase the accuracy of implementation. In-service coaching helps foster ongoing professional growth and enhance teachers' precision in applying evidence-informed practices (Kretlow & Bartholomew, 2010; Kretlow et al., 2012). Key components across coaching interventions included small-group initial training, multiple observations, feedback, and modelling. The extent of coaching teachers received was strongly and positively linked to their performance in implementing proactive classroom management techniques (Reinke et al., 2014).

Observing example practices may be another approach to enhance implementation quality. Kennedy et al. (2017) evaluated the impact of a multimedia-based intervention that provides example videos (Content Acquisition Podcasts for Teachers with Embedded Modelling Video) on teachers' implementation of evidence-based classroom management practices. Their small-scale randomised

controlled trial suggested that teachers who accessed the video-embedded condition used more evidence-based practices than those in the control condition, leading to greater student engagement.

Taken together, these studies illustrate that implementation quality is relevant to multiple levels, including systemic, organisational, and individual. Existing research underscores that the effective use of research evidence in schools depends not simply on providing high-quality research evidence but on ensuring the conditions necessary for its successful implementation are in place. However, most studies on the implementation quality were theoretical frameworks, lacked comparison groups, or were based on a small scale, which requires further investigation for grounded evidence.

3.4 Specifying education contexts in Mainland China

Mainland China is currently at the beginning stage of integrating research evidence into teaching practice. This section starts with an overview of the Chinese primary education system and curriculum standards, providing the contextual foundation for understanding the role of evidence-use in this setting. It then describes current efforts on evidence-use in education.

3.4.1 Overview of the Chinese primary education system

The Compulsory Education Law of the People's Republic of China mandates nine years of free and compulsory education for all school-age children and adolescents. Children typically begin compulsory education at age six or seven, starting with six years of primary education followed by three years of junior secondary education. According to the Ministry of Education's statistics, in 2023, China had 143,472 primary schools, 5,818,159 full-time primary school teachers, and 108,360,253 enrolled primary school students. The current standard class size for primary schools is 45 students.

The education system follows a subject-based curriculum. In primary schools, students take classes in Chinese literacy, mathematics, English, Science, morality and rule of law, sports, music and art, comprehensive practice, information technology, and labour skills. China has long maintained a highly centralised curriculum administration system (Cui et al., 2018). The Ministry of Education sets the curriculum standards for each subject and approves textbooks published by designated publishers. In the new curriculum plan, the subject categories, total weekly class hours, and hour distribution according to the national curriculum standards are clearly outlined. A specific number of weekly class hours are allocated for curricula that local governments and schools can develop or choose.

Based on the standard national curriculum, schools have some flexibility to develop innovative school-based curricula and establish distinctive teaching practices tailored to their unique characteristics. For instance, Han et al. (2024) introduced a school-based curriculum at a primary school that focuses on students' mental health through the use of drama. Wang (2021) adapted dialogic teaching into daily practice by posing guiding questions to initiate conversations between teachers and students, creating a student-centred learning environment that fosters student interaction.

3.4.2 New curriculum standard of Chinese literacy in primary schools

The current curriculum standard for Chinese literacy is published by the Ministry of Education (2022). The standard provides detailed guidelines on the nature of the curriculum, concepts, objectives, content, academic quality monitoring, and curriculum implementation.

Overall, the curriculum standard requires students to 'learn to read in various ways and develop the ability for independent reading' and to 'appreciate the beauty of language, understand the ideological content and artistic value of works, and, by drawing on their own experiences, comprehend, appreciate, and make initial evaluations of literary works.' The standard established three layers of Chinese literacy learning task groups involving basic task groups, developed task groups, and exploratory task groups. The six grades were separated into three-year groups with different task requirements.

The basic task group emphasised students' accumulation and organisation of Chinese language and characters. For grades 1 and 2, the standard requires students to recognise and write basic Chinese characters, learn pronunciation with Pinyin, and discover writing patterns of Chinese characters. For grades 3 and 4, students are expected to independently recognise and write characters and to organise the relationships between the pronunciation, form, and meaning of commonly used characters. For grades 5 and 6, students should use various methods to independently recognise and write characters, enrich their vocabulary, and pay attention to the emotional connotations of words.

The developed task group encompasses practical reading and communication, literary reading and creative expression, and critical reading and expression. Practical reading and communication focus on helping students gather and integrate valuable information through listening, reading, and observation during practical activities, and effectively convey it to meet communication needs. Literary reading and creative expression guide students to appreciate the unique charm of literary language and imagery and to understand the basic characteristics of literary works. Critical reading

and expression require students to analyse the relationship between ideas, facts, and evidence through reading, comparison, inference, questioning, and discussion, fostering logical thinking and a rational mindset. As students' progress from lower to higher grades, reading content expands from everyday life topics to science and society, and reading forms evolve from short texts, poetry, and picture books to literary works, scientific articles, argumentative essays, and brief commentaries.

The exploratory task group involves whole book reading and interdisciplinary learning. Whole-book reading requires students to use various reading methods to engage with entire books. The curriculum standard recommends different types of books for three-year groups, ranging from picture books and nursery rhymes to children's literary classics and outstanding works of literature, popular science, and science fiction. Interdisciplinary learning connects classroom and extracurricular activities, as well as school and community environments, broadening the scope of students' language learning and application.

Teachers are expected to organise the content and timing of learning activities thoughtfully, based on students' learning conditions and task requirements, to encourage active participation and enhance learning effectiveness (Mou, 2024). The curriculum standard also indicates the need for teachers to adopt a rigorous approach to improving their teaching skills, enabling them to better adapt to new challenges and emerging issues.

The standard suggests estimating students' Chinese literacy achievement through recognising and writing characters, reading and appreciation, expression and communication, and organising and inquiry (Ministry of Education of People's Republic China., 2022, p.49-51). Despite the learning outcomes, the standard also emphasises students' ability to set up learning goals and enhance self-directed and cooperative learning ability.

3.4.3 Investigation of research evidence-use

Education quality and students' learning achievement are always focused on by the government. The 14th Five-Year Plan for National Economic and Social Development of the People's Republic of China and the Outline of Long-term Objectives for 2035 outlines the orientation and goals for the policy of 'building a high-quality education system'. Promoting education reform through improved assessment mechanisms and developing a high-quality education system are key strategies to drive high-quality development during this period.

Recently, the Double Reduction Policy highlights the value of student-centred and quality-based education (Xue & Li, 2023). At the school level, the policy required schools to provide 'high-quality, well-balanced development of compulsory education needs' and calls for improvements in classroom teaching (Qian et al., 2024). To effectively implement such reforms, there is a growing recognition of the importance of using research evidence to inform teaching practices.

In 2007, the National Assessment of Education Quality was launched as a national monitoring system for the quality of Grades 4 and 8 school education in China (Jiang et al., 2019). The assessment had a stratified and systematic probability proportionate to size sampling, covering 32 provinces. However, the assessment is still in development and validation, with data not being published yet.

In October 2019, the Ministry of Education of China promulgated the 'Opinions of the Ministry of Education on Strengthening Educational Scientific Research in the New Era' (Ministry of Education of the People's Republic of China, 2019). The document calls for integrating educational research into teaching plans, policy decisions, institutional frameworks, and public discourse. It requires innovation for further educational research to 'strengthen empirical research, insist on facts and evidence as the basis, and continuously track major issues, focusing on long-term and systematic research'.

Education research in China has predominantly focused on qualitative, speculative approaches, resulting in a narrow academic paradigm, limited scientific rigour, and a disconnect between theory and practice (Yao, 2021). It is challenging for policymakers, school leads, and teachers to realise 'what works' for improving management and teaching efficiency. Consequently, educational reforms and decisions are often experience-based rather than evidence-driven. Ren and Li (2021) indicated that the limited quantity and low reliability of educational research evidence are key barriers to incorporating research evidence into teaching in China.

A few studies in China investigated the approaches to facilitate teachers' use of research evidence. Song et al. (2023) established a model for using evidence in practice. The model consisted of identifying issues to address, searching for evidence in databases, carefully selecting the best evidence, using evidence to inform teaching, and engaging in observation, reflection, and analysis. Zhou and Jing (2024) proposed the use of research evidence in school-based training. The study suggested 5 As for the mechanism of evidence-based teaching, involving ask, access, appraisal, apply, and assess.

Zhao and Chen (2024) constructed an Evidence-based Practice Advancement in Instructional Teaching framework based on healthcare's evidence-based practice promotion framework to address evidence, context, and facilitation in classroom teaching. However, the impact on classroom effectiveness was estimated by case study without proper comparison, which led to unwarranted results. Specifically, no study was found to evaluate the impact of teachers' use of research evidence on primary school students' Chinese literacy outcomes.

Wang and Ji (2024) explored the relationship between research evidence use and teacher research, surveying 33,332 primary and secondary school teachers in Sichuan Province, China, about their perspectives on teacher research. The findings indicated that while 73% of teachers stated that they understood the process of education research, most teachers lacked a deep understanding of the value of education research. More than half of the participants thought the teaching research was useful for teaching practice. Only 28% of the teachers conducted research to improve teaching. The study also highlighted potential barriers to conducting research by teachers, such as limited research knowledge, access to research and data, time, and motivation.

With the increasing amount of information and growing demands for teaching innovation, teachers are often tempted to follow popular new teaching methods. However, these methods may lack an evidence basis, which increases the risk of ineffective teaching and places an additional burden on teachers.

For instance, a new concept and practice of 'Big Units' (Da Danyuan) has been gaining popularity as an integrated teaching approach. This method aims to enhance the coherence of instruction by linking and integrating teaching content. Schools devote large effort to developing a 'Big Units' teaching approach. Zhang (2024) outlined strategies for implementing holistic teaching in primary Chinese literacy classes based on Big Units, providing detailed examples of how to structure content. The approach follows four steps: interpreting curriculum standards to establish themes for Big Units, analysing textbook content to design layered learning objectives, creating engaging learning scenarios driven by tasks, and developing integrated assessment methods based on student needs within the unit.

However, the effectiveness of the Big Units approach lacks rigorous research backing, raising concerns among scholars and teachers. Many educators feel uncertain and overwhelmed by this new model (Bao, 2024). Wen (2024) cautioned against impractical and hastily implemented teaching method

reforms that overburden teachers and fail to achieve desired outcomes. Wen advocated a collaborative approach, involving experts and practitioners to conduct research, address real issues, and serve as a responsible think-tank for educational reform.

While the use of research evidence in teaching practice is underdeveloped, school-based teaching-research and professional development may support teachers' adaptation of research evidence in China. Compared to Western teaching research institutions and professional bodies for teachers, China's teaching research system has stronger resource allocation and administrative authority (Zhou, 2021). Schools regularly organise teaching workshops and professional development trainings. Ke et al. (2019) indicated that teacher participation in school-based professional development was generally related to the improvement of teacher efficacy and teaching strategies adopted in the classroom, by surveying 1,506 secondary school teachers. Some schools also possess an 'action-research' process, cooperating with universities to enhance in-service teachers' professional development (e.g., Gu & Wang, 2006; Xu, 2009).

3.5 What can existing studies not answer?

The following sections discuss the limitations of existing knowledge on teachers' use of research evidence, emphasising the importance and necessity of this study.

3.5.1 Limited knowledge about barriers to evidence users

Existing evidence on teachers' views on the usefulness of research findings may not be representative. Most studies investigated teachers' views on using research evidence via small-scale surveys or case studies. A limitation of many previous studies on this topic is that they often rely on case studies or surveys with small to medium sample sizes, which may not represent a diverse demographic of teachers. These are generally no causal designs. Studies reviewed by van Schaik et al. (2018) were mainly case studies, and the review did not consider the quality of the underlying research, which makes the findings of the review less trustworthy.

Data collection methods in these studies may also contribute to bias, which could not represent a diverse demographic of teachers. For instance, the survey conducted by Mills et al. (2021) was distributed through specific channels such as the Chartered College of Teaching Twitter account and the UCL Institute of Education newsletter. Teachers who are willing to participate in a survey or interview about using research findings may already have a preference for research evidence.

Most existing studies have focused on the predictive power of teachers' views from a single aspect, either regarding themselves or the school environment. When these two aspects are combined, the strongest predictor of teacher views remains uncertain. The relationship among these predictors, particularly the interaction between teachers' skills and the school research environment, is still unclear.

In addition, these studies primarily focus on the use of research findings to improve teaching and do not compare teachers' perceptions of other approaches for teaching quality improvement. It is unclear if teachers will still value research findings as useful for their teaching when they are facing with other options.

3.5.2 Uncertainty of the most effective approach for evidence-use

There is limited evidence on the most effective ways to support teachers to incorporate research evidence with high quality and on its benefits for improving teachers' skills and perceptions, and students' learning outcomes. While some studies have explored methods to enhance teachers' awareness, skills, and practices in evidence-use (e.g., Beyer et al., 2025; Cain, 2015a, 2015b; Greenwood et al., 2003; Ilhan, 2021), findings from other research have been mixed. For instance, while Research Learning Communities can foster a more positive attitude towards evidence use, they have limited impact on teachers' skills (Rose et al., 2017). Benner et al. (2023) offered online training on effective reading instruction and found no improvement in teachers' knowledge of evidence-based practices, but observed increased self-efficacy and more positive engagement with research use.

Only a small number of studies have examined the direct impact of teachers' research use on student learning outcomes, and even these results remain inconsistent. Although Clarke et al. (2011) and Wiggins et al. (2019) claimed positive impact, experiments of See et al. (2016), Rose et al. (2017), and Lord et al. (2017b) found no difference or mixed results between the treatment and the control groups. Notably, while some frameworks were proposed, there was almost no study that rigorously estimated the impact of teachers' use of research evidence within the Chinese teaching context.

3.6 Conclusion

Public interest and investment in evidence-use have increased, as a potential approach to improve education quality, which urgently requires more evidence on the use of research evidence. However, it is still unclear which approaches to use evidence are effective and what their impacts are. Teachers often remain unconvinced by research evidence. Barriers such as limited access, insufficient skills,

time constraints, and inadequate school support further hinder the integration of research evidence into practice. The existing literature fails to address these challenges, leaving the research-practice gap largely unresolved. By focusing on the unique context of China, this research contributes to the global understanding of teachers' use of research evidence and offers practical insights to bridge the research-practice gap.

Design and Methods

Chapters 4 to 6 are about research designs and methods for different research questions that explore evidence on teachers' use of research evidence. Chapters 4 and 5 introduce the process of using existing datasets and studies to better understand the process of teachers' use of research evidence. Chapter 6 describes the evaluation process of the use-of-research-evidence intervention with relevant outcome measurement methods.

Chapter 4 Research design and methods: Secondary data analysis

This chapter introduces the secondary data analysis process that investigates the situation of teachers' access to educational research information and their perceptions on using research findings for teaching improvement. Data from the OECD Teaching and Learning International Survey (TALIS) and the NFER Survey of Teachers were two high-quality and open-access datasets about teachers and research evidence. The following sections describes the research questions and data analysis methods for each dataset respectively.

4.1 TALIS

TALIS is the first large-scale international survey for in-service teachers and head teachers, exploring teaching as a profession. The survey is carried out every five years, starting from 2008. Participants are invited to report their work situation including instructional practices, school leadership, professional practices, teacher education and initial preparation, teacher feedback and development, school climate, job satisfaction, teacher human resource issues and stakeholder relations, and teacher self-efficacy. In 2018, innovation, and equity and diversity were added as new themes. TALIS data has open access and could be downloaded from the OECD website (https://www.oecd.org/education/talis/).

4.1.1 Research questions

TALIS did not settle specific questions about teachers' use of research evidence but offers insights into potential access to educational research information such as educational conferences, professional literature, and individual or collaborative research. Although lots of the educational conferences and professional literature lack a research evidence basis, they remain common approaches for teachers to reach research information in education. The analysis of TALIS data investigated current opportunities and the tendency for teachers to engage with research information in education.

- What is the situation of teachers' access to research information on education around the world?
- What is the trend of teachers' access to research information on education around the world?

4.1.2 Design

TALIS is a cross-sectional survey that collected teachers' answers to the surveys during a short period of the year. The first research question described access to educational information of teachers from

different countries. This thesis analysed the latest TALIS dataset, TALIS 2018, with a cross-sectional design. The second question explored the change of teachers' access to education information from 2008 to 2018. Different teachers participated in TALIS in different years, therefore, this study employed a repeated cross-sectional design.

4.1.3 Sample description

TALIS mainly focused on teachers and headteachers from lower secondary schools. Participant countries and economies can include teachers from primary to upper secondary schools. Table 4.1 illustrates the numbers of countries or economies, schools, and teachers that participated in TALIS from 2008 to 2018 respectively.

Table 4.1 Description of TALIS 2008-2018

Year	Number of countries/economies	Number of schools	Number of teachers
2008	23	386	72,190
2013	36	1,810	186,346
2018	48	1,714	261,426

From 2008 to 2018, the number of countries and economies that engaged with TALIS increased. TALIS selected schools randomly with stratification. Participant countries and economics were invited to set a certain form of stratification to answer national requirements. For example, in TALIS 2018, the explicit stratification for England contained two types of funding, two school sizes, and four locations.

4.1.4 Items description

The themes of TALIS vary slightly for different years (Table 4.2). Education conferences, seminars, and professional literature were among the ways teachers access educational research findings. The item on access to education conferences or seminars was included in all three TALIS cycles, while access to professional literature was surveyed in TALIS 2008 and TALIS 2018. The item on individual or collaborative research on a professionally relevant topic appeared in TALIS 2008 and TALIS 2013 but was excluded from TALIS 2018. While these surveys provide some insights, they did not fully capture the breadth of teachers' access to research information. Nevertheless, they remained the only available source offering a global perspective on this issue, though with limitations.

Table 4.2 Items in TALIS

Item/Year	2008	2013	2018
Education conferences or seminars (where teachers and/or researchers present their research results and discuss educational problems)	Yes	Yes	Yes

Individual or collaborative research on a topic of interest to you professionally	Yes	Yes	
Reading professional literature (e.g. journals, evidence-based papers, thesis papers)	Yes		Yes

4.2 Analysis for TALIS

This study describes teachers' access to educational conferences or seminars, professional literature, and self-conducted research across TALIS participant countries or economies. To examine the distribution and trends of teachers' access to research-related resources over the past decade using TALIS data, the teachers were recoded into two groups: those who explicitly reported such access and those who did not. Teachers who either did not answer the questions or indicated no experience with the specified educational information were classified into the latter category ('no access or unknown'). The percentage of teachers with access to these resources was calculated by dividing the number of teachers reporting such access by the total number of teachers in each country or economy.

The study started with the latest dataset TALIS 2018 to illustrate the current situation of teachers' access to research-relevant resources including educational conference or seminars and professional literature across 48 countries or economies. This analysis reports the average access rate across all participants and countries and identifies those with higher or lower access rates than the overall average.

This study examined the overall trends in teachers' access to education conferences or seminars, professional literature, and individual or collaborative research from 2008 to 2018. Countries or economies included in the analysis for teachers' access to education conferences should have participated in at least two of TALIS from 2008 to 2018, involving 40 countries and economies. The analysis for access to professional literature was conducted through 20 countries and economies that participated in both TALIS 2008 and 2018. Trends of teachers' opportunity to participate in research was analysed with 16 countries and economies that joint in TALIS in 2008 and 2013.

4.3 Survey of Teachers 2010

TALIS provided international evidence on teachers' access to educational information that was related to research evidence. However, items from TALIS did not illustrate teachers' attitudes towards research evidence and were insufficient to explain predictors of teachers' attitudes and values on research findings. The Survey of Teachers is the only accessible dataset that provides an insight into

teachers' use of research findings with a large sample size. The following session briefly introduce the Survey of Teachers in England and the process of analysis.

The General Teaching Council for England (GTCE) conducted six annual surveys from 2004 to 2010 to obtain first-hand information and perspectives from educators, except in 2008. The information gathered from these self-reported surveys played a crucial role in establishing GTCE policies and making recommendations to the Department for Education. In 2010, the Survey of Teachers aimed to understand how in-service educators improved their teaching. Teachers' opinions on professional development, the utilisation of observation and feedback, the application of research findings, and performance management were collected through a survey.

The report of this survey conducted by Poet et al. (2010) suggested that teachers who had more professional activities in the past 12 months and experienced teachers tended to have a more positive attitude towards research. But this report combined teachers' use of existing research findings and conduction of their own research, which could not accurately illustrate the predictors of perceived usefulness of research findings.

The Survey of Teachers dataset is safeguarded, and should be requested from the UK Data Service by registering and accepting the End User Licence. This study set up a teacher use of research project in the UK Data Service website and applied for the access to data in the Survey of Teachers. The dataset was downloaded with a .sav format that could be analysed in SPSS. While the most recent Survey of Teachers dates back to 2010, it remains the best available large-scale dataset offering insights into teachers' use of research evidence.

4.3.1 Research questions

The study investigated the following questions with Survey of Teachers 2010, from an overview of teachers' attitudes towards the value of research findings to predictors of their attitudes. Potential support from schools and colleagues was also discussed to facilitate teachers' confidence in understanding and using research findings.

- How do teachers value the usefulness of research findings?
- What is the strongest predictor of teacher-perceived usefulness of research evidence for improvement of teaching?

- To what extent may school encouragement and colleague discussion support teachers' confidence in using research evidence?

4.3.2 Design

The research questions explored predictors of teacher-perceived usefulness of research findings. This analysis could illustrate the 'snapshot' of in-service teachers' perceptions of research findings and was based on the Survey of Teachers 2010. The Survey of Teachers is a set of cross-sectional surveys varying topics and items in each year. Teachers participated in the survey voluntarily, so the participants were different every year. The characteristics of the dataset led to a cross-sectional design for the study.

4.3.3 Sample description

The survey contacted 13,500 teachers of which 4,392 teachers responded to the survey. The NFER compared the demographic information of teachers in the survey and the population, indicating that there was only a small difference between the achieved sample and the population in General Teaching Council for England (Poet, Rudd, and Kelly, 2010).

This study focused on the use of research evidence among teachers in compulsory education, therefore, the sample size included 3,982 teachers who had teaching experience from key stage 1 to key stage 4. The other 410 teachers who only worked in the pre-key stage or in key stage 5 were excluded from the dataset. Teachers who did not answer the question 'how useful the research findings in others' research are to improve their teaching' were also excluded from the dataset which finally contains 3,936 teachers. Teachers' background information such as ethnicity, sex, employment status, and work experience were collected by NFER and accessible in the dataset, as shown in Table 4.3. Teachers who answered this survey were mainly white, female, and full-time teachers.

Table 4.3 Teacher background characteristics

		Percentage
Ethnicity	BME	4.4
	White	93.6
	Missing	2.1
Sex	Male	20.7
	Female	79.3
Employment Status	Part-time	27.9
	Full-time	71.2
	Missing	0.8
Worked in KS1		31.0

Worked in KS2	42.0
Worked in KS3	44.8
Worked in KS4	43.6
Number of teachers	3,936

4.3.4 Items description

Teacher-perceived usefulness of research evidence for improving teaching practice is reflected by the question:

Thinking about the different ways that you might improve your teaching, in the past two years, how useful have you found the following? Using findings of other people's research.

This was a four-point Likert scale, from 'not at all useful' to 'very useful'. Teachers could also select 'no experience of it' to indicate that the item was not applicable to their situation.

This self-reported survey provided insights into teachers' perceived skills and opportunities related to research engagement, with items shown in Table 4.4. Teachers responded using a five-point scale ranging from 'strongly agree' to 'strongly disagree', with an additional option of 'do not know'.

Table 4.4 Items in the Survey of Teachers 2010

Confidence and external condition	I know where I can find relevant research that may help to inform my teaching.
	The research I have found I have been able to understand and use.
	My school encourages me to use research findings to improve my practice.
	I have frequent opportunities (at least termly) to discuss relevant research findings with my colleagues.
	I do not have time to use research findings.
Conducting research	In the last 12 months, I have undertaken my own research and enquiry to improve my practice.
	I feel confident in my research skills to conduct my own enquiry.
	Usefulness of doing my own enquiry or research.

4.4 Analysis for Survey of teachers

This study investigated the situation of teacher-perceived usefulness of research findings for their teaching practice with a large-scale dataset covering teachers from England. The analysis started with a descriptive analysis of teachers' views on the usefulness of diverse resources to improve teaching. A multi-variate regression model was constructed to find out the strongest predictor of the willingness to use research evidence. Taking a further step, the study explores the relations among predictors via path analysis.

4.4.1 Data cleaning and missing data

Teachers who did not respond to the question on the usefulness of research evidence were excluded at the first stage (n = 46), as explained in Section 4.3.3. For the remaining cases, missing values in categorical variables (e.g., sex, ethnicity, and employment status) were recoded as 'missing' and the cases retained, while missing values in numeric variables were replaced with the mean of the corresponding items, to retain most of the sample and reduce potential bias (Gorard, 2020).

4.4.2 Descriptive analysis

This study described how teachers valued the approaches for improving teaching, such as self-reflection on practice, school self-evaluation, performance management objectives, conducting enquiry or research, using research findings, and using teacher assessment data. Teachers' answers were re-categorised into 'useful', 'useless', and 'no experience' and the proportion in each category was reported. As some teachers had no experience with certain approaches, the useful rate was calculated as the percentage of teachers who rated the approach as useful among those who had experience with it (with the following formula).

$$Useful\ rate = (useful|experienced) = \frac{percentage(useful)}{percentage(experienced)}$$

4.4.3 Logistic regression modelling

To identify the strongest predictor of teachers' expressed perceptions of the value of research evidence, a multivariate logistic regression model was constructed. The outcome variable was based on teachers' responses regarding the usefulness of research findings for enhancing teaching practice. For this analysis, expressed perceived usefulness was recoded as a binary variable: responses of 'useful', and 'very useful' were coded as 'useful', while responses of 'useless', 'very useless' and no experience were grouped together as 'not (reported as) useful'. A limitation of this approach is that teachers who regarded research evidence as useless and those who reported no experience are conceptually different groups. However, since the primary purpose of this analysis was to identify factors associated with expressing research as useful, rather than to distinguish between different forms of non-use, these categories were combined.

Before the modelling, teachers' demographic characteristics, confidence in using research evidence, and external conditions for using research findings were cross tabulated with the value of research findings. The odds ratio (OR) was employed to create the association between a categorical variable and the outcome variable, which is calculated with the following formula:

$$Odds\ ratio = \frac{n_{exposed\ cases}/n_{unexposed\ cases}}{n_{exposed\ non-cases}/n_{unexposed\ noncases}}$$

The order of input variables was from teachers' background variables, followed by teachers' confidence and external conditions. This order could provide a sense of how much difference the later items make to the usefulness of research evidence that the individual already had with their background. Items about teacher conducting their own enquiries were entered in the last block for a high relevance to use of research findings to avoid absorbing variance that should be attributed to four key predictors, potentially resulting in biased estimates. Predictors in each block were illustrated in the following Table 4.5.

Table 4.5 Predictors in each block in the regression model

Block	Variables
Block 0 – Beginning	--
Block 1 - Background	Ethnicity
	Employment Status
	Sex
	Key Stage 1
	Key Stage 2
	Key Stage 3
	Key Stage 4
Block 2 – Confidence and external condition	I know where I can find relevant research that may help to inform my teaching
	The research I have found I have been able to understand and use
	My school encourages me to use research findings to improve my practice
	I have frequent opportunities (at least termly) to discuss relevant research findings with my colleagues
	I do not have time to use research findings
Block 3 – Conducting research	In the last 12 months, I have undertaken my own research and enquiry to improve my practice
	I feel confident in my research skills to conduct my own enquiry
	Usefulness of doing my own enquiry or research

The model was created in SPSS IBM 29 and employed the forward entry function to determine whether predictors should be included in any one block. Given the presence of missing cases and the non-randomised nature of the data, p-values cannot be calculated mathematically. Even if they could, significance tests merely indicate the probability of observing a given coefficient estimate under the null hypothesis that the coefficient is truly zero, and give no indication of whether the coefficient is zero. Also, such tests do not directly illustrate the magnitude or practical relevance of any observed

effect (Berk & Freedman, 2001). Instead, this study reported the percentage correct of each block to support the model. If the percentage correct increases with adding a block, it suggested that variables in that block have predictive power for teacher-perceived usefulness of research findings.

This logistic regression model was in the form as follows:

$$Logit(Y) = \ln\left(\frac{p}{1-p}\right) = \beta X + \alpha$$

where Y was teacher-perceived usefulness of research findings, X was the matrix of predictors in the three blocks, p was the probability of the outcome of interest, β was the matrix of regression coefficients, and α was the intercept.

This analysis used the column labelled $Exp(\beta)$ to represent the odds by which the outcome of having positive perceptions towards the usefulness of research findings increase or decrease with the increase in each predictor. An $Exp(\beta)$ value greater than 1 suggested that the predictor was associated with increased odds of positive attitudes, while a value less than 1 indicated decreased odds.

The regression model was refined based on the receiver operating characteristic (ROC) curve and adjustment of cut-off point. The ROC provided a more precise and comprehensible picture of whether teachers were categorised into the correct category that had a positive or negative attitude towards research findings. The curve was illustrated by the horizontal axis with false positive rate and the vertical axis with the true positive rate. A model with good predictive power has a high true positive rate and a low false positive rate. The ROC curve of this model exhibits a performance that is close to the (0, 1) coordinates, and the area under the curve is close to 1.

The logistic regression predicted the outcome variable with a probability ranging from 0 to 1 and used the cut-off point to categorise the continuous probability figure to a binary variable. The classification plot was figured for the probability of teachers being predicted to be in one or other of the two outcome categories. With a 0.5 cut value, the horizontal axis was labelled half Y (perceiving usefulness of research evidence) and half N (not perceiving usefulness of research evidence). Each Y or N within the area of the plot is the actual outcome. The plot illustrated the actual outcome to refine the model (Gorard, 2021, p.226).

4.4.4 Path analysis

While discovering the strongest predictor of teachers' views on research findings, this study also explored the interaction between teachers' confidence and external conditions of using research evidence. A path analysis model was constructed to explain how support from schools and colleagues may facilitate teachers' understanding of research findings. The model was illustrated in Figure 4.1.

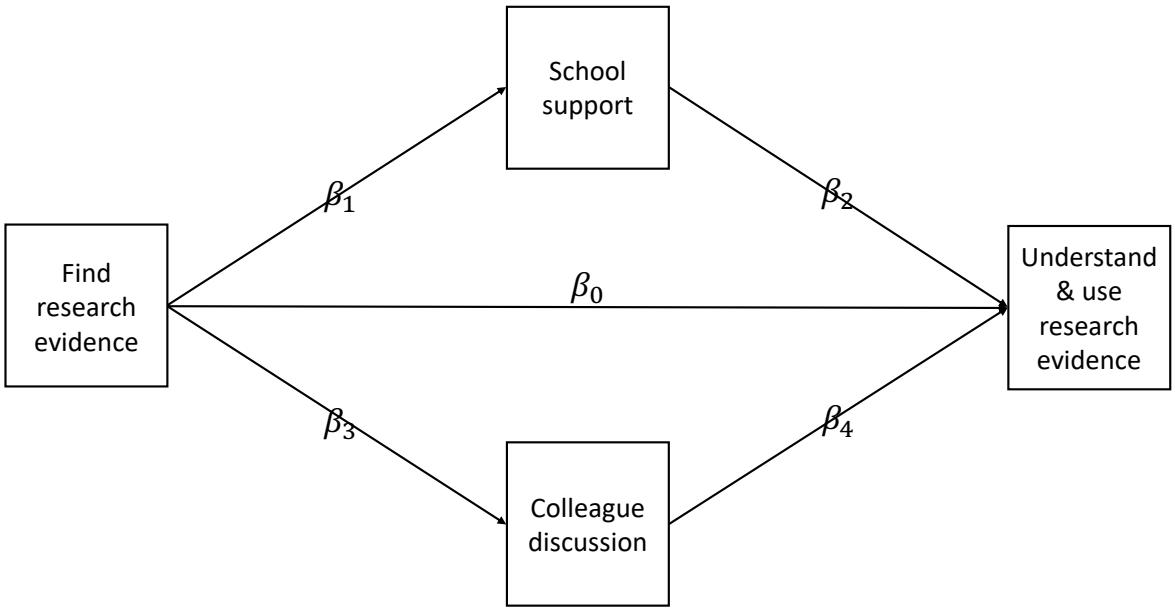


Figure 4.1 path analysis of using research evidence

This study constructed the model in M-plus editor and diagrammer, Version 8 for IOS. The model was based on the relationship between finding research evidence and understanding research evidence. Then variables about school support and opportunity to discuss with colleagues were added. This study employed R^2 as an indicator for the model rationale. If R^2 increases with school support and colleague discussion, it suggested that school support and colleague discussion partially explain the process from finding research evidence to understanding it.

The standardised coefficients on each indirect path were multiplied and added up ($\beta_1 \times \beta_2 + \beta_3 \times \beta_4$) to estimate the explanatory power of support from schools and colleagues in the process of understanding research evidence. The ratio of the total indirect 'effect' over the total 'effect' was calculated with the following formula:

$$Ratio = \frac{\beta_1 \times \beta_2 + \beta_3 \times \beta_4}{\beta_0 + \beta_1 \times \beta_2 + \beta_3 \times \beta_4}$$

The ratio indicated the proportion of the effect of teachers' confidence in finding relevant research findings on understanding research evidence that was mediated by support from schools and colleagues. The ratio could represent the potential difference made by the mediators.

4.5 Conclusion

The chapter presented the design and methods to analyse two existing datasets, TALIS and the Survey of Teachers 2010. This analysis provided evidence on situation of teachers' access to educational information relevant to research and predictors of teachers' attitude towards usefulness of research evidence to improve teaching. The following two chapters will have a further investigation for the promising approaches to encourage teachers' use of research evidence in daily practice.

Chapter 5 Research design and methods: Structured review

Previous reviews about the use of research evidence focused on medicine, public practice, and education but none of them synthesised the existing evidence on teachers' use of research evidence to improve students' literacy outcomes (see Section 3.1.2). The best way to facilitate teachers' use of research evidence and its impact on students' literacy attainment remain unclear.

A systematic review allows an in-depth analysis of existing literature to provide the evidence on the impact of teachers' use of research evidence. The current chapter presents procedures of the structured review which investigates the most promising approach to facilitating teachers' use of research evidence and improve students' literacy attainment.

5.1 Research questions

This structured review summarised findings from existing studies on the impact of using research evidence on students' literacy and the most effective approaches to encourage literacy teachers to engage with research evidence. The review addressed the following questions:

- What is the impact of teachers' use of research evidence on students' literacy outcomes?
- What is the most promising approach to encourage teachers' use of research evidence, in order to improve students' literacy outcomes?

5.2 Search strategy

The review developed keywords highly related to 'teachers use research evidence', 'impact', and 'student literacy' at the first stage. Subsequently, the terms were used and modified in accordance with the idiosyncrasies of different databases.

5.2.1 Search terms

The search terms separated the keywords into five categories linked with 'AND'. The first three categories were around 'teachers use research evidence'. The use of research evidence was developed with phrases such as evidence-based, evidence-informed, research-led, and evidence into practice. The synonyms of the verb 'use' were gathered such as usage, utilise, disseminate, transfer, access, and translate. The search terms restricted the search in the education field and focused on teachers, so terms such as school, classroom, teacher, education, and student were applied.

This structured review explored the impact of teachers' use of research evidence on students' literacy outcomes including both academic attainment and non-cognitive attainment. The fourth category of search terms included skills relevant to literacy such as reading, grammar, spelling, and comprehension, words about academic outcomes such as attainment, assessment, and proficiency, and words about wider outcomes such as attitude, autonomy, and enjoyment.

The review restricted the search to studies that investigate the causal relationship between use of research findings and students' literacy outcomes. Words indicating a causal claim such as evaluation, effect, determinant, and randomised controlled trial were used in the last category of search terms.

5.2.2 Databases

The review was carried out with five electronic databases within the social science field, including Scopus, Web of Science, British Education Index (BEI), Educational Resources Information Centre (Eric), and ProQuest Dissertations and Thesis Global, with Google Scholar as a supplementary source. ProQuest covers Masters' dissertations and PhD theses, so including this database ensures that some high-quality but unpublished work is also considered in this review. Given the differences in how each database indexes publications, search terms were adjusted to capture the most relevant studies. The search terms can be found in Appendix A. In addition, the review also included studies suggested in others' review and Google Scholar.

5.3 Inclusion and exclusion criteria

Before screening, the study set up a list of inclusion and exclusion criteria. The studies were included if they:

- Included teacher level interventions about teachers' use of research evidence
- Focused on in-service, mainstream teachers
- Explored impact on students' literacy
- Were published in English

Studies were excluded if they:

- Evaluated only the impact of an evidence-based intervention without examining how teachers engaged with or used research evidence in their practice

- Were not implemented by in-service, mainstream teachers (i.e., institutional tutors, teaching assistant, or researchers)
- Were in special education
- Focusing on subjects other than literacy
- Did not have measurable outcomes of literacy
- Were not primary research
- Were based on participants' self-report (i.e. subjective opinions or individual experiences)
- Were theoretical or conceptual pieces

5.4 Screening

Because many articles mentioned 'evidence' in their main texts without being directly relevant, the search terms intended to maximise the inclusion of studies related to teachers' use of research evidence. All identified studies were sorted by relevance to search terms. Studies were first screened for relevance by titles and abstracts with the use of inclusion and exclusion criteria. To balance comprehensiveness with feasibility, the screening stopped when there were continuously 100 pieces of work with no relevance to the research topic. Included studies were recorded in EXCEL with their citation and deduplicated to facilitate the in-depth screening. Then the full text of included studies was downloaded and screened. The screening process adopted the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) (Page et al., 2021) to report the total number of screened studies and the number that remained and synthesised.

5.5 Data extraction

This study focused on the impact of using research evidence on students' literacy outcomes and the most promising approach to encourage teachers' use of research evidence. Key information extracted from included studies was about each study's research design, sample size, attrition rate, measurement quality, intervention implementation, and findings. This comprehensive summary of information informs the assessment of the strength of evidence, distinguishing this review from previous ones on this topic. The information was recorded in EXCEL. The table includes the following items:

Study characteristics

- Author(s)
- Research focus: different forms of using research evidence

Intervention

- Which intervention is developed and delivered by the researchers to facilitate teachers' use of research evidence?

Study design

- Does the study have appropriate comparison group?

Sample

- Sample size in each group
- Clarification of nationalities of samples
- Sampling strategy
- Phase of education

Attrition

- How many students left the programme?

Measuring instrument

- How are students' literacy outcomes measured?
- Does the study use independent, standardised, and validated measurements?

Research findings and results

- Literacy performance
- Change in teachers' perceptions, skills, and teaching practice

Trustworthiness of research (explained in Section 5.6)

- Research design
- Scale
- Missing data
- Measurement quality
- Rating outcomes (from 0* to 4*)

5.6 Quality assessment

Only robust research evidence is trustworthy for real-life applications. Low-quality research may lead to misleading results. This structured review assessed the quality of evidence on teachers' use of

research evidence with Gorard 'SIEVE' quality assessment tool including research design, sample size, attrition rate, and measurement quality (Gorard, 2024), as shown in Table 5.1.

Table 5.1 The Gorard 'SIEVE' for quality assessment

Design	Scale	Missing data	Measurement quality	Rating
Strong design for research question	Large number of cases per comparison group	Minimal missing data, no impact on findings	Standardised, independent, accurate	4*
Adequate design for research question	Adequate number of cases per comparison group	Some missing data, possible impact on findings	Standardised, independent, some errors	3*
Weak design for research question	Small number of cases per comparison group	Moderate missing data, likely impact on findings	Not standardised/independent, errors	2*
Very weak design for research question	Very small number of cases per group	High level of missing data, clear impact on findings	Weak measures, high level of error, of many outcomes	1*
No consideration of design	A trivial scale of study	Huge amount of missing data, or not reported	Very weak measures	0*

The Gorard 'SIEVE' assesses the strength of research using five possible levels, ranging from 0 asterisk to 4 asterisks. The evaluation begins in the first column, focusing on research design and measurement quality. Each study is assigned a rating based on its lowest row description across the first four columns. A 4-asterisk rating indicates high-quality research with a high level of trustworthiness, which provides a plausible answer to the most promising approach to engage teachers' use of research evidence. With consideration of research quality, Chapter 8 will report the results of this structured review including the impact of teachers' use of research evidence on students' literacy and the approaches to encourage evidence-use from the most secure studies.

5.7 Conclusion

To sum up, Chapters 4 and 5 were designs and methods for existing evidence on situation and impact of teachers' use of research evidence. Based on existing studies, Chapter 6 is designed to investigate the impact of an intervention about teachers' use of research evidence on students' early literacy outcomes in mainland China.

Chapter 6 Research design and methods: evaluating teachers' use of research evidence

After summarising existing evidence on teachers' use of research evidence, the current study moves forward to estimate the impact of a use-of-research-evidence intervention. This chapter presents the design and methods used to evaluate the effectiveness of teachers' use of research evidence in improving students' first language learning outcomes at primary schools in mainland China. The chapter starts with the statement of research questions, followed by an introduction of the intervention, design, sample, randomisation, outcome measurement, and data analysis.

6.1 Research questions

A use-of-research-evidence intervention is employed in this study, combining the findings from the structured review (see Chapter 8). The study estimated the impact of the intervention on primary students' literacy and teachers' perceptions and skills in using research evidence in China. During the intervention semester, a process evaluation was also conducted to explore teachers' engagement and how they adapted research evidence into their teaching practice.

6.1.1 Research questions for impact evaluation

The impact of teachers' use of research evidence was evaluated based on primary students' Chinese literacy scores and their self-reported enjoyment of reading. The research questions were as follows:

- To what extent does teachers' use of research evidence improve students' Chinese literacy outcomes?
- Is there an impact of teachers' use of research evidence on students' enjoyment of reading?
- Does the impact of teachers' use of research evidence on students' literacy outcomes differ according to students' demographic characteristics (sex, age, ethnicity, and hukou)?

This study also examined the impact of the intervention on teachers' perceptions and skills in using research evidence for improving teaching practice and their ability to use research evidence. Two questions were posed:

- What impact does the use-of-research-evidence intervention have on teachers' perceptions of research evidence?
- To what extent does the use-of-research-evidence intervention make a difference to teachers' skills of using research evidence?

6.1.2 Research questions for process evaluation

In addition to the impact evaluation, this study presented how the intervention was implemented. The process evaluation analysed teacher compliance with the intervention protocol and investigated the 'successful use' conditions with the following five questions:

- Is it feasible for teachers to incorporate research evidence into the Chinese primary school context?
- To what extent does teacher compliance make a difference to the impact of teachers' use of research evidence on students' Chinese literacy outcomes?
- How can schools support and facilitate teachers' use of research evidence?
- What conditions play a role in the successful use of research evidence in classroom practice?
- How is the implementation of evidence-informed strategies different among teachers?

6.2 Design

This study aims to make a causal claim about the impact of teachers' use of research evidence on students' literacy outcomes. A causal claim model should contain four elements: a correlation between cause and effect, a time sequence, changes in the effect produced by varying the strength of the cause, and a plausible explanation of the process (Gorard, 2013, p.59). RCTs have been considered as a strong design for a subset of causal questions about 'what works'.

This study requires teachers to attend teacher training sessions and use certain evidence-led teaching strategies in daily teaching practice. And it is not feasible to re-allocate students into new classes. The appropriate design of the study would be a teacher-level clustered randomised controlled trial to reduce possible contamination. Teachers and their corresponding classes were randomly assigned to receive the intervention about evidence-informed teaching practice.

Treatment group teachers participated in a use-of-research-evidence presentation and got access to intervention materials at the beginning of the experiment semester. School headteachers, research leads, and treatment teachers were told not to share intervention materials with others. The study also involved a waiting-list for teachers in the control group. They were informed that they could participate in the study after the trial. The design promoted an unbiased estimate of the effectiveness of teachers' use of research evidence.

6.2.1 Sample

The participant recruitment started with reaching out to schools. School headteachers were provided with an invitation poster that gave a brief introduction to the project. Headteachers who were interested in the project were invited to a warm-up meeting to gain a deeper understanding. Each school nominated a research lead who assisted other teachers in understanding research evidence and facilitated their evidence-informed classroom practice.

Teacher-level recruitment was processed within schools. Primary teachers who were teaching Chinese literacy were all eligible to participate in this project. Headteachers played a crucial role in presenting the project to eligible teachers and compiling a list of willing participants, which they then gave back to the researcher. The project was voluntary, and teachers could drop out of the project at any time.

Participant schools were in Hefei, the capital city of Anhui province in southeast China. A pilot study was conducted before the main experiment to test the feasibility and readability of intervention materials over six weeks, from May to July 2023. Two primary schools in Anhui, China, participated, with 46 Chinese literacy teachers and 2,033 students from grades 3 to 6. Both schools had a good reputation in the area for high-quality education.

The main study expanded to evaluate the impact of teachers' use of research evidence with a larger sample, comprising 187 teachers and 8,241 students from five schools. One of the original pilot schools continued to participate in the main study to further explore evidence-informed teaching practices. Among the four newly added schools, two schools were situated on the boundary between urban and rural areas and enrolled many rural Hukou and migrant students. Notably, one of the newly added schools implemented a quasi-experimental design and its results were not included in the impact analysis.

6.2.2 Randomisation

Randomisation was conducted at the teacher level within each school, using stratified randomisation based on school, grade, and the number of classes taught. It was carried out with the random number function in EXCEL. Teachers were first categorised by grade and then by the number of classes they taught. Special attention was given to the number of classes per teacher, as assigning multiple-class teachers to the same group could lead to an imbalance in the number of students between groups. Teachers within the same grade and with the same number of classes were randomly allocated to either group.

While the pilot study followed such randomisation process, the main study involved a larger sample size and a more complicated randomisation procedure. Four primary schools in China participated in the randomised controlled trial and one primary school participated in a quasi-experiment setting by nominating participant teachers at their insistence. Before the randomisation, school leaders selected the strategies to be incorporated and provided information about teacher assignment. All schools had sufficient skilful teachers to implement the strategies, and each teacher taught one class only.

Three of the four schools in the main study were randomised by the researcher, while one school randomised teachers themselves through a random draw. For researcher randomised schools, the study gathered all participant teachers to a list and randomly allocated them by year groups. Teachers who had participated in the pilot study and school research leads were excluded from the randomisation and directly allocated to the treatment group. At the self-randomised school, the lot was a small piece of paper with either 'treatment group' or 'control group' written on it. The paper was folded in half before being drawn. Teachers randomly drew lots to determine whether assigned individuals in the treatment group or not.

One additional school participated by nominating 32 teachers for the treatment group. However, this school did not provide data for a control group, as they were unwilling to submit non-experimental data. Their grading system differs from that of the other schools, and a substantial amount of exam scores and questionnaire data from both students and teachers is missing. This school was excluded from the impact analysis. Nevertheless, because its workshops and the process to use research evidence differs from the other schools, the classroom observation results from this school will be reported to offer insights into an alternative approach.

6.3 Development of the use-of-research-evidence intervention

This study provided teachers with training to enhance their knowledge and competence in using research evidence in classroom practice. The intervention encouraged teachers' use of research via multiple approaches, including a booklet featuring five examples of promising evidence-informed teaching strategies, a teacher training presentation, and monthly follow-ups. The intervention was adjusted in light of the results from the structured review.

Notably, the intervention aimed to facilitate teachers' use of research evidence-informed strategies, with an emphasis on the investigation and development rather than just the strategies themselves.

Neither the materials nor the professional development was delivered by external intervention companies commissioned by schools. Instead, the content was carefully extracted from research evidence and relevant books. For example, when a school chose to implement formative assessment, the intervention drew directly from Anders et al. (2022), who specifically validated the effectiveness of Dylan Wiliam's *Embedded Formative Assessment* teacher training. Activities and supporting materials related to formative assessment were selected from this book and summarised into a practical guide.

The strategies that teachers employed also tended to be broader in scope, and less specific in their application. With peer tutoring, for instance, while numerous studies, such as those by Vardy et al. (2022) and Topping et al. (2012), confirmed its effectiveness, they described various ways to organise peer tutoring. This study combined the organisational approaches discussed in these papers with insights from *Using Peer Tutoring to Improve Reading Skills* (Topping et al., 2015) and *Productive Group Work: How to Engage Students, Build Teamwork, and Promote Understanding* (Frey et al., 2009), both of which align closely with the research evidence.

This use-of-research-evidence intervention intended to enhance teachers' skills to select, understand, and adapt evidence-informed strategies along with their related practical activities. The protocol allowed teachers flexibility in how they implemented these strategies in their own classrooms. Therefore, the impact of teachers' use of research evidence was not solely due to the strategies themselves but stems from the interaction between using research evidence and applying those strategies. The following sections describe the components of the use-of-research-evidence intervention.

6.3.1 Booklet of evidence-informed teaching strategies

The intervention booklet contained examples of promising teaching strategies that have been evaluated rigorously in existing literature. The examples of promising teaching strategies were summarised in a table with a brief introduction of each strategy to facilitate school leaders' choice. Once the strategies were chosen from the menu, schools were provided with the detailed user guide extracted from existing research on the chosen strategy to help classroom practice. The following section first introduces the process of selecting strategies. The booklet used in the experiment could be found in Appendix B.

There are diverse teaching strategies that are considered to potentially improve students' literacy skills. However, most of them are not rigorously evaluated or are not suitable for Chinese literacy teaching and learning contexts (e.g., phonics). Strategies selected by this project should satisfy at least two of the four criteria: a high-quality research base, a potential effect size on reading, low cost, and easy implementation.

The selection of strategies to be included in the booklet of evidence-informed teaching strategies began with a review of promising teaching strategies from the EEF toolkit, the US Institute of Education Sciences' What Works Clearinghouse (WWC), and systematic reviews about effective classroom instructions in primary literacy (e.g., See & Gorard, 2020). Ten interventions and strategies that had potential to improve primary literacy were identified. The strategies were *Accelerated Reader*, *Achieve 3000*, *Learning by Questions*, embedding formative assessment, Growth mindset, student tutoring, reciprocal peer tutoring, dialogic teaching, Writing about Values, and text messaging.

A review and synthesis of existing research evidence on the effectiveness of these interventions resulted in some interventions with similar features being combined. For example, *Accelerated Reader*, *Achieve 3000*, and *Learning by Questions* all had online learning with rapid formative feedback as a common feature. These three were therefore grouped under the category of computer-assisted formative assessment, contributing to a final pool of eight promising teaching strategies. These were:

- computer-assisted formative assessment,
- embedding formative assessment,
- Growth mindset,
- student tutoring,
- reciprocal peer tutoring,
- dialogic teaching,
- Writing about Values,
- and text messaging.

From this pool, the researcher identified feasible interventions by considering their suitability for Chinese schools in terms of school funding, facilities, and accessibility of products.

For instance, the pilot study included computer-assisted formative feedback, peer tutoring, dialogic teaching, and Writing about Values. However, no school selected computer-assisted formative

feedback because of insufficient software support. Therefore, in the main study, this option was removed from the booklet and replaced with embedding formative assessment. The strategies used in the main RCT were embedding formative assessment, peer tutoring, dialogic teaching, Writing about Values, and Growth mindset.

The booklet started with a table indicating the suggested grades, estimated effectiveness, and strength of existing research evidence of each strategy (Table 6.1). The suggested grades in the table were determined through a combination of insights from existing studies on the effectiveness of each strategy and the researcher's judgement. While some were directly informed by the grade levels used in prior studies, others were determined by considering the intervention's purpose and the characteristics of the target student group. For example, Writing about Values targets students facing high-stakes exams and has mainly been studied in secondary schools. In the Chinese context, Grade 6 students take a final examination that marks the transition to lower secondary education, making them an appropriate group for this intervention.

The estimated impact was summarised with the effect sizes of the strategy from high quality research. These figures were not exact calculations derived from specific formulas but serve as rough estimates to give teachers a general sense of the magnitude of the effect, which helped them make informed choices when selecting evidence-based strategies. The evidence strength indicated the quality of evidence on the strategy with consideration of the quality of research on the strategy and whether the strategy has been tested widely. Embedding formative assessment, peer tutoring, and dialogic teaching were widely and rigorously evaluated, showing promise and therefore, rated with three asterisks.

Table 6.1 Evidence-led strategies for the main study

Strategies	Suggested grades	Estimated impact	Evidence strength
Embedding formative assessment	1-6	0.20	* * *
Peer tutoring	3-6	0.30	* * *
Dialogic teaching	4-6	0.10	* * *
Writing about Values	6	0.07	* *
Growth mindset	5-6	0.10	* *

The following describe the five high-quality research evidence supporting these strategies.

Evidence for formative assessment. High-quality research evidence supporting these strategies is presented here. A recent structured reviews combining 48 studies suggested that the average effect size of formative assessment on reading achievement was +0.18 (Xuan et al., 2022). Another review suggested a larger effect size +0.33 on reading in the US (Lee et al., 2020). However, both reviews included several small-scale experiments and studies without appropriate comparison groups.

A more robust piece of evidence comes from a recent large-scale randomised controlled trial evaluating Dylan Wiliam's Embedding Formative Assessment programme, which reported an effect size of +0.11 on students' reading outcomes at GCSE level (Anders et al., 2022). This trial was well-designed, involved a large sample, had low attrition, and received a four-asterisk rating (Table 6.2).

Taken together, the evidence for embedding formative assessment is supported by extensive and rigorous evaluations. However, due to the limited number of studies conducted in China, its overall trustworthiness was rated at three stars. Based on both systematic reviews and primary research, the estimated average impact was approximately +0.20.

Table 6.2 Evidence on embedding formative assessment

Report	Year group	Duration	Evidence strength	Estimated impact
Anders et al. (2022)	Year 10 (UK)	2 years	4 *	+0.11

Evidence for peer tutoring. Peer tutoring is also a well-evaluated strategy, with several studies assessing its effectiveness (e.g., Lewin et al., 2024; Vardy et al., 2022). However, the majority of the research has been limited to case studies or lacks balanced comparison groups. The following three studies were randomised controlled trials (Table 6.3). A total of 114 schools, including 4,840 students, participated in the first study that evaluating peer assisted learning strategies (Lewin et al., 2024). More than three thousand students from 125 primary schools in Scotland participated in the second study which implemented both cross-age and same-age peer tutoring over two years (Topping et al., 2012). The third study directly evaluates the effects of peer-assisted learning, randomising 118 students from Grade 4 (Kung & Teo, 2022). These studies all reported promising results in improving reading achievement.

Table 6.3 Evidence on peer tutoring

Report	Year group	Duration	Evidence strength	Estimated impact

Lewin et al. (2024)	Year 5 (UK)	35-minute sessions, 3 times per week, for 20 weeks	3 *	+0.12
Topping et al. (2012)	Year 4 and Year 6 (UK)	30-minute sessions, 1 or 3 times per week, for 15 weeks	3 *	First Year +0.31, Second Year +0.34
Kung and Teo (2022, as cited in Vardy et al., 2022)	Year 4 (Taiwan, China)	40-minute sessions, twice a week, for 15 weeks	2 *	Vocabulary +0.07, Reading Comprehension +0.11

Evidence for dialogic teaching. Dialogic teaching also has strong evidence that suggest positive impact. The implementation guidance selected four randomised controlled trials which evaluated the effectiveness of dialogic teaching (Table 6.4). The first three evaluated the impact of Philosophy for Children (P4C) while the last one evaluated another dialogic teaching model. The research findings suggest that dialogic teaching has a small positive impact on reading. Although the evidence base for dialogic teaching was relatively robust, its estimated impact appeared smaller and less promising compared to formative assessment and peer tutoring, with an overall estimated effect size of +0.10.

Table 6.4 Evidence on dialogic teaching

Report	Year group	Duration	Evidence strength	Estimated impact
Gorard et al. (2015)	5, 6 (UK)	One year	3 *	Reading +0.12, Writing +0.03
Lord et al. (2021)	6 (UK)	Two years	3 *	+0.01
Fair et al. (2015)	7, 8 (US)	24 weeks	2 *	+0.21
Jay et al. (2017)	5 (UK)	20 weeks	3 *	Cognitive Abilities Test +0.12, KS2 +0.03

Evidence for Writing about Values. The evidence on Writing about Values was slightly weaker than the first three strategies. Writing about Values is a self-affirmation intervention, which has been widely evaluated among Black students, mostly in the US but most research on it is weak. The evidence suggests that self-affirmation activities can have positive and long-term results improving academic achievement, especially for ethnic minority groups (e.g., Wu et al., 2021). Some studies have also

suggested that these effects persist through secondary school (Borman et al., 2021) and right up to university (Goyer et al., 2017).

A large scale randomised controlled trial conducted in England had shown that the Writing about Values is also somewhat effective in raising the attainment of low-SES students. See et al's (2022) study involving 11,000 students aged 14 to 16 in England also showed minimal positive effects on students' Attainment 8 scores (ES = +0.05) for those eligible for free school meals (Table 6.5). While this study was rated 4*, impact of Writing about Values on low-income and low-achieving students was not fully evaluated, especially in Chinese teaching context. Therefore, the evidence strength of Writing about Values in Table 6.1 remains at 2*.

Table 6.5 Evidence on Writing about Values

Report	Year group	Duration	Evidence strength	Estimated impact
See et al. (2022)	Year 10, 11 (UK)	One year	4 *	+0.05

Evidence for Growth mindset. The evidence for growth mindset is also weak (rated 2* in Table 6.1). Two large RCTs evaluated the impact of growth mindset (Foliano et al., 2019; Yeager et al., 2019). Yeager's work only consisted of two 25-minute sessions while Foliano's work provided more workshops to students and trained teachers to use growth mindset in daily practice. A smaller RCT by Rienzo (2015) had very high attrition rate and was thus rated as 1* (see Table 6.6). These three studies provided evidence that growth mindset is a promising approach for students' reading outcomes. However, as Yeager is a member of the developer, Dweck's research team, there is a conflict of interest, and the evidence is not considered as strong. Therefore, the overall evidence strength of growth mindset is 2*.

Table 6.6 Evidence on Growth mindset

	Year group	Duration	Evidence strength	Estimated impact
Foliano et al. (2019)	Year 6 (UK)	6 workshops for students; 2 half-day training for teachers	3 *	+0.01
Yeager et al. (2019)	Grade 9 (US)	2 sessions, 25 min each	3 *	Overall +0.11, Low achieving +0.33
Rienzo et al. (2015)	Year 5 (UK)	6 workshops and 4 further sessions for	1 *	Overall +0.18, Low-achieving +0.34

		students about malleability of intelligence		

This booklet provided research evidence-led teaching strategies to teachers and performed as a ladder to help teachers be aware of existing evidence. In the booklet, each of the strategies was described and the mechanism about how the strategy worked was briefly explained. Relevant high-quality research evidence was also presented to teachers to remind them of the idea of using research evidence. Free materials and online resources were also recommended, including structured guides and detailed supporting materials, especially activity sheets, example lesson plans, and checklists.

As part of the intervention, the steps required to implement the strategies were summarised for each strategy. These steps or protocols were taken from the high quality studies. Details of the activities to be used in the classroom, classroom observation forms, and checklists for teachers to evaluate the implementation of themselves or of their colleagues were adapted from books and literature published by the researchers of these individual strategies. However, teachers had the flexibility to select supporting materials relevant to their context, and make adjustments based on their teaching content. Some teachers voluntarily provided their teaching plans which are open-access resources for other treatment group teachers.

6.3.2 Teacher training presentation

Prior to the trial, one 90-minute teacher training session was held to introduce the evidence-informed strategies and demonstrate the use of research evidence. The presentation was delivered at the first two weeks of the intervention semester. The slides for the presentation could be found in Appendix C.

The researcher presented the first section of the presentation focusing only on the fundamental knowledge about educational research. The presentation started with simple examples of studies in education to dispel the notion that academic research was distant and overly complex for teachers. The introduction also encouraged teachers to approach research with a more critical and rational mindset, fostering a deeper understanding of its practical relevance. Teachers were shown the purpose of the new study, the research procedure, and the study timeline. The presentation was developed based on the topics: two dimensions of evidence-informed practice, access to research, judging research quality, and understanding research findings.

Participant teachers were introduced to research evidence using explicit and comprehensible language. It was explained that 'research evidence in education typically refers to findings generated by experts through studies that follow a formal research question and a systematic design, often assessing the effectiveness of educational programmes or approaches.' A distinction was made between research evidence and other forms of evidence, including students' actual test scores and teachers' reports of student performance.

The evidence-into-use process was categorised to two dimensions (Gorard et al., 2020):

- The level of evidence from plain evidence, modified evidence to engineered evidence
- Dissemination approach from passive, engagement, to interactive (see Section 2.2.2).

The presentation provided a simplified quality assessment tool (Gorard 'SIEVE') to help teachers make basic judgements of the research quality based on design, sample size, attrition, and measurement quality (see Section 5.6). To help teachers assess research trustworthiness in a real-world context, a simple exercise was provided featuring four brief statements evaluating the impact of graded reading on students' language skills. For example: 'Interview 40 students about their perceptions of graded reading after using related software for a period', and 'Randomly assign half of the classes to use the graded reading software, then measure and compare their reading test scores before and after the experiment'.

Teachers were introduced to common statistical indicators to help them interpret research findings, such as Cohen's d, Pearson r, and standard deviation, with an emphasis on how these measures compare students' average performance. The relationship between research quality and effect size was explained and distinguished. This session of the presentation ended up with slides showcasing example strategies that were informed by evidence.

The second section of the presentation was conducted collaboratively by the researcher and school leaders. It focused on presenting the research evidence behind the strategies selected by the school, encouraging teachers to assess the quality of the research and interpret the findings using the knowledge gained from the first session. The researcher introduced the materials that supported teachers to adapt a certain strategy such as example lesson plans, activity sheets, relevant books. Following this, the school research leads shared their understanding of the selected strategy and provided detailed guidance on adaptation in classroom practice, tailored to their specific school

context. They also set clear goals and expectations for teachers across different grade levels. The detailed implementation guide can be found in Appendix D.

6.3.3 Teacher workshops and follow-ups

Treatment teachers were expected to implement the selected strategy with fidelity in the classroom, adhering to the strategy's baseline requirements, such as frequency and specific activities. Monthly workshops were scheduled throughout the experiment semester, offering a variety of formats, including demonstration lessons, group discussions, and reading salons. Participants had opportunities to share their understanding of evidence-use in teaching practice, observe their peers' use of evidence-informed practice in the classroom, and discuss the problems that they encounter.

Importantly, the researcher provided only light-touch facilitation during the workshops. School research leads were primarily responsible for providing constructive feedback and guidance. When teachers needed clarification on the instructional materials, the researcher would step in to offer detailed explanations. Unless teachers made major implementation errors or completely failed to use any evidence-informed teaching strategies, the researcher did not interfere with their process of exploring and integrating research evidence. The researcher mainly served as a source of moral support, listening to teachers' experiences and challenges. Ultimately, teachers themselves found the solutions to the issues they faced.

6.4 Outcome measurement

The outcomes of this study were measured using students' Chinese literacy scores, the Teacher Use-of-Research-Evidence Survey, and the Student Enjoyment of Reading Survey.

6.4.1 Primary outcome

The primary outcome measure of this study was students' performance in the final Chinese literacy examination. This examination was developed in alignment with the national curriculum and assessed three core components, including vocabulary, reading comprehension, and writing skills. These components collectively reflect students' mastery of the Chinese language and literacy. Students' performance on this examination is referred to as their Chinese literacy outcome throughout this study.

The final examination was usually held by the district department for education. All primary schools in the same district participate in the same test at the same time. Teachers rated students' scores

based on the same standard, and therefore, the final test represented students' Chinese literacy outcomes with a high reliability. The final examination scores of Chinese literacy for the prior semester were collected as the pre-test results, and the final examination scores for the intervention semester were the post-test results.

6.4.2 Secondary outcome

Teachers' awareness and skills in research evidence were collected before and after the project through the Teacher Use-of-Research-Evidence Survey (see Appendix E). The survey was created and distributed online via *Qualtrics*. It was developed based on the NFER *Survey of Teachers 2010* and EEF *Research Use Survey (*https://x.com/educendowfoundn/status/1625819639635759106?s=46&t=hNyw4DDfn_r7gF3jQ0fn2g*).* The survey also referenced items in existing questionnaires about teachers' use of research (e.g., Cahill et al., 2015; Gleeson et al., 2023; Lysenko et al., 2014).

Existing surveys usually developed Likert scales with limited response options, interpretation issues, and difficulty in aggregation. However, the intervals between these options are not necessarily equal. For instance, the emotional distance between strongly disagree and disagree may not be the same as that between disagree and neutral. Therefore, in this survey, teachers rated each statement, from 0 to 10 to express their level of agreement from strongly disagree to strongly agree. This scale also allows the responses to be treated as numeric variables, thus facilitating analysis.

The Teacher Use-of-Research-Evidence Survey contained eight questions with 38 question items. Questions followed the order from less to more relevant to the topic about the use of research evidence. The first two questions tested whether teachers could identify trustworthy sources to make decisions about their teaching practice. The third question asked about teachers' confidence in their ability to understand and adapt appropriate research findings into teaching and their perceived usefulness of research findings to improve teaching. The next three questions (Questions 4, 5 and 6) assessed teachers' skills in accessing research findings, their understanding of impact, and ability to judge research quality. The final two questions aimed to find out about the conditions (e.g., workload, training, and materials) for teachers to better use research evidence in their practice and the awareness of using research evidence for decision-making.

This sequencing was intentionally designed to minimise potential bias in responses. Presenting research-related items frequently at the beginning of the survey might have influenced teachers'

initial judgments, for instance, when identifying the most trustworthy sources for professional decision-making. They might have rated research findings more highly if they assumed that such responses were expected or preferred by the researchers.

While the pre- and the post-tests had the same question structure, questions that measured teachers' ability to understand research impact, judge the research quality, and make evidence-led decisions were slightly different. This design aimed to reduce potential practice effects, where teachers might recall their responses from the pre-test and adjust their answers accordingly. Using varied but equivalent items also helped sustain teachers' engagement with the survey by avoiding unnecessary repetition while still measuring the same constructs.

Questions about the best description of research evidence (Question 2), understanding research impact (Question 5), judging research quality (Question 8), and evidence-informed decision making (Question 8) in the survey had 'preferred' answers. While these answers were considered more desirable, they were not the only correct options. All items were designed to be rating questions rather than multiple-choice or ranking formats. Teachers rated their level of agreement with each item on the same 0–10 scale. These items were intended to offer a relatively objective measure of teachers' ability to evaluate the quality of research and understand the potential impact of research evidence. The rationale for selecting these answers is provided below.

Question 2: Description of research evidence. How much do you agree with the following descriptions of what educational research evidence is?

Item 1: Books and articles by educators and psychologists disseminating their views.
Item 2: Knowledge from observations, information analysis, and discussions in my school.
Item 3: Findings from robust research conducted by experts in education.

A scoring protocol was developed. When teachers rated the third item as the highest, they got 2 points. If the third item was rated higher than any of the other two items, they got one point. Else, they got a zero.

Question 5: Interpretation of research findings. If education research shows that an intervention leads to a positive (or negative) impact on student attainment, what does this mean?

Item 1: The study was of high (or low) quality.
Item 2: The intervention was effective (or ineffective) in this research environment.
Item 3: All students in the intervention group performed better (or worse) than those in the control.

Item 2 was the preferable answer for this question, while item 1 mixed up research quality with positive impact, and item 3 ignored that the impact analysis was at the average level. Teachers could earn 2 points if they rated the second item as the highest. They got one point if the score of the second item was higher than either of the other two items. Otherwise, they earned 0 points. The wording of this question and the responding items had been slightly adjusted between the pre-test and the post-test to avoid a practice effect.

Question 6: Judgement of research quality. To assess teachers' skills in judging research quality, two different evaluation approaches for assessing the impact of a certain intervention on student reading were given. Teachers were asked to select the most appropriate out of two approaches. In the pre-test, teachers were asked to choose the most appropriate approach to investigate the impact of peer tutoring on student reading. In the post-test, they were asked to select the most appropriate approach to evaluate the impact of extracurricular reading.

Item 1 (pre-test): Selecting one class as a 'peer tutoring' class and comparing their reading scores with other classes.
Item 2 (pre-test): Interviewing 40 students about their experience of peer tutoring.

Item 1 (post-test): 1,023 students randomised to read extra books for 25 minutes per day, or not
Item 2 (post-test): Collecting reading scores and extracurricular reading time of 300 students for correlation analysis

Based on the Gorard SIEVE (2021, p.94), the first item in both the pre- and post-test surveys generally had a stronger design or more standardised outcome measurements. However, since the survey simplified the research descriptions so much, some key information for this audience group may have been omitted, potentially affecting teachers' ability to judge research quality accurately. Despite this limitation, the first item in each pair was likely a more rigorous evaluation of an intervention's effect. Teachers' skills to judge research quality was measured by the correct identification of high-quality research and perceived research quality difference between the pair.

If teachers could identity high-quality research (Item 1), they were awarded one point, otherwise they scored zero. However, this variable could not indicate the size of the difference the teacher perceived between higher-quality and lower-quality research. The measure of perceived research difference was calculated as the difference in scores that teachers give to high-quality and low-quality research. There is no correct answer.

Question 8: Evidence-informed decision-making. Four items were used to assess teachers' understanding of the conditions under which they should or should not use research evidence. In the pre-survey, the scenario described a circumstance where the school provided materials claimed to be evidence-based, yet the teacher observed no improvement among students. The options were:

Item 1: Stop using these materials immediately.
Item 2: Compare your use of materials with colleagues and ask for support.
Item 3: Look again at the research evidence and make a judgement on whether to continue.
Item 4: Trust the school and carry on using the materials as instructed.

Similarly, in the post-survey, the situation involved a famous educator advising against the practice of having students collect and correct their mistakes, despite the teacher having used this strategy for years.

Item 1: Trust your experience and carry on.
Item 2: Conduct your own research to decide whether to continue.
Item 3: Look again at the research evidence and make a judgement on whether to continue.
Item 4: Trust the expert and change your teaching approach

Teachers who rated items 2 and 3 higher than the average of these four items got two points, while those who rated one of the items higher than average got one point. The score represented teachers' awareness of making decisions based on research evidence.

6.4.3 Non-cognitive outcome

The Student Enjoyment of Reading Survey was designed based on the PISA 2018 student questionnaire and PIRLS student questionnaire (see Appendix F). The survey comprised students' favourite reading materials, reading attitudes, and Chinese homework time via five questions.

The survey started with the students' favourite reading materials, such as Chinese textbooks, Magazines, comic books, fiction books, non-fiction books, and social science. Students provided their frequency of borrowing books from the school library, reading online, and reading paper books. They were also asked about the frequency of reading for information and for enjoyment. These items did not require an accurate report of frequency but students' agreement with the statement that 'I often…'. This was for students to express the frequency easily. Students may find it difficult to provide an accurate number of the frequency. Students rated their agreement to the items from 0 to 10 to represent never to always.

Students' reading attitudes were measured using six items. Students' general attitudes towards reading, such as 'I read only if I have to', 'Reading is one of my favourite hobbies', 'For me, reading is a waste of time', captured students' overall dispositions (e.g., enjoyment or aversion) in a broad and abstract sense. The other three items were context-specific, including 'I read only to get information that I need', 'I like talking about books with other people', 'I would be happy if someone gave me a book as a present'. They help elicit more comprehensive responses linked to real-life experiences.

Five items were included to represent students' self-estimation of their Chinese reading ability. They started with a general question about their reading performance, asking, 'I usually do well in reading'. Then, students evaluated their learning experiences with Chinese reading. They were asked whether they found Chinese reading harder compared to their classmates and whether they considered Chinese easier to read than other subjects. This helped to gauge their perceptions of reading difficulty and their relative ease with learning Chinese. The survey ended up with two questions about reading comprehension, specifically in terms of how well students understood the text and answered related questions correctly.

In addition, the last question of the survey was about the average time for Chinese homework per day. This question was used to know if teachers' use of research evidence could alleviate students' learning burden.

6.5 Data analysis for impact evaluation

This section describes the impact evaluation analysis across the four RCT schools. The quasi-experiment school was excluded in the headline impact analysis.

6.5.1 Intention-to-treat analysis

The main analysis employed was intention-to-treat (ITT) where teachers were analysed according to the original group they were assigned regardless of whether they received the intervention or not. This preserves the original randomisation and minimises potential selection bias due to dropout or non-compliance. As teachers volunteered participation in this project, they could change their minds and drop out of the intervention before and during the trial. Excluding them from analysis may distort the results if those who dropped out were somehow different to those who stayed on. For example, teachers who dropped out maybe those who did not believe in research evidence, relying instead on their experience and knowledge of what worked. If they were not included in the analysis, the intervention effectiveness could be potentially overestimated.

The intention-to-treat analysis maintains the original comparability of the treatment group achieved after randomisation and reflects what would happen in real practice (Sedgwick, 2013). The study analysed the impact of the use-of-research-evidence intervention on teachers' perceptions and skills in research evidence and students' Chinese literacy outcomes according to their original assigned groups. The comparison was between the treatment group and the control group regardless of whether teachers participated in the training and implemented evidence-informed strategies. Even dropout teachers' survey data, corresponding student survey data, and attainment data were still collected and used in data analysis, if they were willing to provide.

6.5.2 Imbalance at baseline

This study employed a stratified randomised controlled trial to achieve a balanced teacher allocation in terms of grade and teacher assignment. However, teachers' characteristics such as teaching experience, degree qualification, and knowledge about research evidence and students' characteristics such as sex, reading habits, and Chinese literacy attainment might still be imbalanced before the intervention. Adjusting the allocation to balance after randomisation would be a potential threat to increase bias in comparison, which decreases the security of the project.

Baseline data is good practice for data analysis. The project reported characteristics of teachers and students from the control group and the treatment group. The Cohen's d 'effect' sizes for continuous variables and odds ratios (see Section 6.5.7) for categorical variables were calculated to estimate the differences between the treatment group and the control group at the beginning of the intervention. When analysing the outcomes, the gain scores were calculated to mitigate any imbalance in teachers' skills and perceptions of research evidence and students' prior Chinese literacy.

Additionally, for the situation that a strong imbalance was observed at the pre-test (e.g., teachers' awareness of evidence-led decision-making), a regression model was constructed to examine whether the intervention's impact on the post-test results persisted, after controlling pre-existing differences in pre-test. In this model, the post-test score served as the dependent variable, while pre-test scores and group allocation were included as independent variables.

6.5.3 Missing data

It is inevitable to have missing cases and missing values in social science research. Ignoring missing data and working only with complete cases reduces the number of valid cases and may lead to bias in research findings, especially when missing data exists non-randomly (Gorard, 2020). Compared with complex weighting responses and multiple imputation, some more feasible approaches are proposed such as acknowledging and recording missing data with a CONSORT flow diagram (http://www.consort-statement.org/consort-statement/flow-diagram), using sensitivity analyses such as the number of cases needed to disturb the difference (NNTD) (Gorard, 2023), and recoding missing as a category and replacing with default values.

This study reported the number of dropouts and their characteristics. The number of dropout students was compared to the number of counterfactual cases needed to disturb the finding which was calculated with the following formula (Gorard & Gorard, 2016):

$$NNTD = Effect\ Size \times N$$

where N is the number of cases in the smallest group in the comparison.

This study deals with two sorts of missing values: missing in students' final test results and missing in survey responses. A student whose prior test record was missing was excluded. Students who did not attend the final tests of the intervention semester were recorded for their reasons and excluded from further statistical analysis. The pre-test scores of these missing students were compared to estimate potential bias.

Teachers and students who did not fill in the surveys or blanks in almost all questions in the surveys were labelled as invalid cases. This study did not judge the invalid cases by 'low-quality' answers such as repeating the same choice for all questions. While it is possible that some participants may not have

answered in line with their true beliefs or just randomly entre some numbers, the researchers avoided making judgements, as doing so could introduce bias into the analysis.

The study reported excluded cases. The missing values from valid cases were reported and replaced with default values. For a categorical variable, this study created a new category of 'missing'. For a continuous variable, the missing values were replaced with the overall mean score. This approach ensured that the default values made no difference to the mean, and instances with default values did not exert excessive leverage on the findings (Gorard, 2020).

6.5.4 Primary outcome analysis

The primary outcome was students' performance in the final examination score of Chinese in the intervention semester. Due to differences in grading systems across districts and year groups, the project standardised students' scores according to the school average score and standard deviation with z-score to alleviate potential bias from grading systems. The average score and standard deviation in Chinese tests of each school in each grade were calculated. The standardised scores reflect students' Chinese literacy attainment in a certain grade and at a certain school. The standardisation was based on the following formula:

$$z = \frac{x - \mu}{\sigma}$$

where x is the raw score, μ is the average score, and σ is the standard deviation.

To evaluate student performance between the intervention group and the control group, the study compared the difference in final examination scores with Cohen's d effect size. The differences between pre- and post-test scores were employed to calculate gain scores.

Since schools selected different evidence-informed teaching strategies, the impact analysis also reported the impact of using each strategy. The effect size should not be simply considered as the estimate impact of each strategy. The true impact arose from the interaction between the use of research evidence and the particular strategies used.

In addition, the main experiment estimated the impact of teachers' use of research evidence on student sub-groups such as students at different reading proficiency levels, sex, rural Hukou students,

ethnic minority students, migrant students, and low-income students. This information was obtained from school records.

6.5.5 Secondary outcome analysis

The secondary outcomes were teachers' perceptions and basic knowledge about research evidence. This study compared post-test scores and gain scores of attitudes and knowledge with Cohen's d effect sizes to indicate the differences between the treatment group teachers and the control group teachers. The questions with preferable answers, particularly those reflecting teachers' skills of research evidence, were recoded into scores to estimate the impact on teachers' skills.

6.5.6 Non-cognitive outcome analysis

The non-attainment outcome referred to students' enjoyment of reading. In the analysis, answers to the inverse questions were converted to positive. For instance, 'For me, reading is a waste of time' was inverted to 'For me, reading is a valuable use of time'. The responses to these items were reverse-coded by subtracting the original score from 10 (i.e., $\text{recoded score} = 10 - \text{original score}$).

The effect sizes of post-test scores and gain scores were reported to estimate the impact of teachers' use of research evidence in the classroom on a wider outcome. A subgroup analysis was also conducted to examine the impact on students' reading enjoyment across different backgrounds, such as residency and migrant status.

6.5.7 Effect size

This study used Cohen's d effect size to measure the differences for each continuous variable between groups according to the following formula (Cohen, 2013):

$$Effect\ size = \frac{Mean\ score_{treatment} - Mean\ score_{control}}{Overall\ standard\ deviation}$$

Though randomisation leads to unbiased allocation, the groups can still be imbalanced (by chance). Gain scores, which refer to the difference between each student's pre- and post-test scores, were used to avoid any imbalance between groups in students' prior test scores. The effect size of the gain score was calculated as follows:

$$Gain\ effect\ size = \frac{Mean\ gain_{treatment} - Mean\ gain_{control}}{Overall\ standard\ deviation\ of\ gain}$$

An “effect” size is a simple and intuitive index that assesses the magnitude of a treatment effect, without requiring strong statistical assumptions. A positive effect size indicates that the intervention had a beneficial impact on the treatment group. This study did not employ ANOVA, ANCOVA, or similar significance tests because the data were not fully randomised due to missing values and clustering. Moreover, ANCOVA generally relies on the assumption of homogeneity of regression slopes, which is often difficult to satisfy in real-world data. In contrast, effect size measures remain the same regardless of clustering or intraclass correlation. Even with full randomisation, significance tests do not provide as straightforward an interpretation of the gain differences as Cohen’s d. Therefore, this study focused on effect sizes as indicators of intervention impact.

Effect sizes for categorical variables were estimated by odds ratio. There were no categorical variables in all outcome measurement, but some categorical variables existed in the background characteristics of teachers and students. The odds ratio was calculated to investigate the balance of the sample allocation based on the following formula:

$$Odds\ ratio = \frac{n(exposed\ cases)/n(unexposed\ cases)}{n(exposed\ non-cases)/n(unexposed\ non-cases)}$$

6.5.8 Regression-to-the-mean analysis

Although teachers were randomly allocated to the treatment group and the control group, the drop-out students and imbalanced allocation may disrupt the impact evaluation because of the potential explanation of regression-to-the-mean. Regression-to-the-mean (RTM) is a statistical phenomenon where if a particular sample of a random variable is extremely high or low, the next measurement of the same variable is likely to be closer to the average value (Barnett et al., 2005).

For instance, more students in the comparison group were in the highest 50% of pre-test scores, while more students in the treatment group were in the lowest 50%. Regression-to-the-mean might mean that the top 50% students tended to get lower scores while the bottom 50% students tended to get higher scores in the next examination. In this case, the treatment group had a higher probability of achieving greater scores in post-test scores, leading to more apparent progress, while the comparison group were more likely to obtain lower scores. This may result in a positive gain effect size in the impact analysis, but this effect would primarily be due to regression to the mean rather than the intervention itself.

The estimation of the RTM effect hypothesises that the gain in effect size was not due to the teachers' use of evidence-based interventions but regression to the mean. The RTM effect was estimated and compared to the overall impact to explore whether the impact was from the RTM or the intervention.

The RTM effect size among each group is calculated by the formular:

$$RTM\ effect = \frac{n \times |gain| + (N - n) \times overall\ gain}{N \times overall\ standard\ deviation}$$

where N is the total number of students in the top or bottom 50%, n is the difference between the numbers of students in treatment or the control group among the top or bottom 50%, and $|gain|$ is the absolute of the gain score of the top or bottom 50%.

6.6 Data analysis for process evaluation

The process evaluation provided formative evidence at every stage of the intervention, from teacher training to outcome evaluations (See et al., 2016). The process evaluation of the pilot study was conducted to improve arrangement in the main study. Limitations in the pilot study, such as the intervention organisation, design, and data collection methods were found. Feedback from teachers and headteachers was used to revise the training materials and supporting documents for evidence-informed teaching strategies.

In the main study, the process evaluation was conducted to better explain the impact results of the intervention. This evaluation assessed the fidelity of teachers' participation and the quality of delivering evidence-informed strategies in classroom practice. It reflected the efforts made by schools to support and facilitate teachers' use of research evidence. The process evaluation also captured the complexity of teachers' engagement with research evidence.

Data from the school level involved school scale and their research experience. School leadership engagement in the use-of-research-evidence intervention was based on the participation of the principal or research lead during researchers' school visits. Teachers' engagement included data were from teacher attendance at workshops and demonstration lessons, frequency of displaying demonstration lessons, and self-reported frequency of using evidence-informed teaching strategies.

Forty-four classrooms were observed during the intervention semester to investigate the process of using the recommended evidence-informed teaching strategies. Field notes were taken on classroom

activities. The researcher also talked to teachers informally to find out how the teaching strategies were implemented in the classroom and difficulties of practice.

6.6.1 Complier average causal effect analysis

Complier Average Causal Effect (CACE) analysis was employed to assess the impact of fidelity to treatment, based on teachers' attendance to the use-of-research-evidence intervention activities and the frequency of using evidence-informed strategies. This analysis addressed how variations in teacher compliance made a difference to the intervention's impact on students' Chinese literacy outcomes.

CACE is estimated based on observed treatment outcomes and the assumption that, due to randomisation, the proportion of compliers in the control group is equal to that in the treatment group. It is also assumed that the average outcome of non-compliers in the control group is the same as that of non-compliers in the treatment group. The calculation of CACE is shown in Table 6.7.

Table 6.7 CACE estimation formula

	Compliers		Non-compliers		All		Effect size
	N	Mean	N	Mean	Total N	Mean	
Treatment	A	m_1	B	m_2	C	m_T	$m_1 - x / S.D.$
Control	(D)	(x)	(E)	(y)	F	m_C	

Using the overall mean scores for both groups and the data for those in the treatment group who complied and who did not comply (A, B, m_1, and m_2), the average outcome for those in the control group who would have complied if given the treatment (*x*) could be calculated. The proportion in treatment group who complied is assumed to be A/C:

D will be $A/C \times F$,

E will be $F - A/C \times F$,

y is equal to m_2,

and x is calculated thus: $F \times m_c - E \times m_2 / D$.

The compliance to the use-of-research-evidence intervention refers to teachers who participated in all required intervention activities and achieved the baseline implementation frequency. Teachers who participated in the training session and two monthly workshops were categorised as 'compliers' with intervention activities. Meanwhile, the study encouraged teachers to apply evidence-informed

strategies with a baseline frequency. Once they practised the strategy in the classroom, they were invited to note down their implementation in their teaching log and reported in the post-test survey. Compliance with baseline implementation frequency was established differently for each teaching strategy, with consideration of recommended frequency in existing studies (Table 6.8).

Table 6.8 Compliance Criteria

<table>
<tr><td colspan="3">Attend:
• Use-of-research-evidence presentation
• Two monthly workshops</td></tr>
<tr><td colspan="3">AND</td></tr>
<tr><td>Strategy</td><td>Suggested using frequency</td><td>Compliance criteria</td></tr>
<tr><td>Formative assessment</td><td>Embedded</td><td rowspan="2">Once a week</td></tr>
<tr><td>Peer tutoring</td><td>3 times a week</td></tr>
<tr><td>Dialogic teaching</td><td>Once a week</td><td>Once a month</td></tr>
<tr><td>Writing about Values</td><td rowspan="2">2 workshops in total</td><td rowspan="2">Twice during the semester.</td></tr>
<tr><td>Growth mindset</td></tr>
</table>

To explore factors related to teacher compliance, this study focused on three key areas: school-level factors, individual teacher confidence and skills, and peer collaboration. At the school level, the average teacher compliance rate was examined in relation to school support, including leadership participation and research opportunities. Leadership participation was indicated by the proportion of school visits attended by school leaders. Research opportunities were calculated as the number of Chinese literacy-related projects conducted by schools per Chinese teacher.

At the teacher level, compliant and non-compliant teachers were compared in terms of their confidence and skills in using research evidence, as well as their frequency of participating in or delivering demonstration lessons. Cohen's d was used to estimate the 'effect' sizes of these differences.

6.6.2 Qualitative comparative analysis

Qualitative comparative analysis (QCA) is a method based on set theory to explore the relationships between conditions and the outcome (Ragin, 2009). This method concentrates on multiple conjunctural causation that different combinations of conditions achieve the same outcome among a set of cases with Boolean algebra. Although this approach is not strong enough to make a causal claim, it aims to find out and reduce the conditions for the outcome.

While QCA was first created for understanding macro-sociological change, researchers also use this method to investigate the success pathways for the process evaluation of an intervention (e.g., Kien et al., 2018; Warren et al., 2022). This study employed a Crisp-Set QCA to classify the key conditions for successful use of research evidence in classroom practice. The process followed instructions of Pappas and Woodside (2021).

Successful use of research evidence was defined as a class achieving higher average gain scores in Chinese literacy than the average gain scores of control group students. The analysis involved teachers' background information, workshop attendance, compliance with the baseline implementation of strategies, and skills in using research evidence.

The key conditions were transformed to binary variable for a crisp set qualitative comparative analysis. The cases for QCA were teachers allocated to the treatment group. All variables were calibrated to crisp sets with values of 0 or 1 which defined a case belongs to a group or not (Table 6.9). The success of implementation was defined as having a higher average gain score than the figure for the control group (average gain score = -0.02).

Teachers' attendance at the presentation of use-of-research-evidence and follow-up workshop were recorded by attendance sheets. Each participating school provided a list of teachers willing to give demonstration lessons in front of the researcher and colleagues, which represented the frequency of delivering demonstration lessons. Teachers' attendance at demonstration lessons, as well as their knowledge of research evidence use was collected via Teacher Use-of-Research-Evidence Survey in the post-test. Teachers who attended at least three demonstration lessons and provided at least one during the ten-week experiment were classified as full members. This classification was based on the average frequency of observing others' lessons and giving formal demonstration lessons. Teachers' skills to understand impact, judge research quality, make evidence-led decisions, and describe research evidence were recoded into binary variables by comparing with the average scores. The average scores in Chinese literacy of intervention classes were compared to that of the control group students.

Table 6.9 Calibration of conditions and potential outcome for success

Variable	Membership	Non-membership
Major in Teacher Education	1	0
Years of teaching	> 9	≤ 9
Attendance at presentation	1	0
Attendance at 1st workshop	1	0

Attendance at 2nd workshop	1	0
Compliance with strategy	1	0
Attending demonstration lessons	> 0.3	≤ 0.3
Providing demonstration lessons	> 0.1	≤ 0.1
Understanding impact	> 1	≤ 1
Perceiving difference in research quality	> 3.8	≤ 3.8
Evidence-led decision making	> 1	≤ 1
Describing research evidence	> 1	≤ 1
Outcome: Average gain score in Chinese literacy	> -0.02	≤ -0.02

The data analysis was conducted with the software fs/QCA version 4.1. Before performing the QCA, a necessity analysis was carried out, reporting the consistency and coverage of each variable to determine if any individual variable was a necessary condition for teachers' successful use of research evidence. QCA models were then built by adding conditions with high necessity, aiming to find a model that covered the most cases and clearly distinguished differences between conditions. Truth tables, including 2^n logically possible combinations of conditions were created where n variables were used in the model. The combinations were reduced to the recipes that were empirically observed, and the frequency of each combination was reported. All QCA solutions were presented in a table with '~' suggesting no existence.

6.6.3 Classroom observation and interviews

The classroom observation of experimental classes recorded how teachers explored and used a certain evidence-informed teaching strategy in real educational contexts. Classroom observational visits provided evidence of the implementation fidelity. The classroom observation form can be found in Appendix G. In addition, teachers and the researcher had informal discussion about their confusion, enthusiasm, and feedback. The informal interviews took place after the classroom observation and workshops.

The observation and interviews of the pilot study focused on the feasibility of teachers' use of research evidence. In total, 15 classes that used peer tutoring or dialogic teaching from the treatment group were observed. Three topics were discussed including teachers' understanding and awareness of research evidence, required materials for better practice, and reflection on teaching practice. The researcher recorded and summarised all information with classroom observation notes to adjust the intervention for the main study.

The main study concentrated more on teachers' classroom practice and students' behaviour with 44 classroom observations across all five schools, regardless of whether they were quasi-experimental or

RCT settings. During breaks, conversations with teachers and students were held to gather their perspectives and feedback. This study adopted a descriptive approach to summarise the observation and interview data to identify patterns in how the strategies were enacted across classrooms.

At the end of the experiment semester, interviews were conducted with headteachers and school research leads to reflect on the process of engaging teachers in the use-of-research-evidence intervention and integrating research evidence into teaching. The discussions covered the role of educational research in evidence-informed decision-making, the school's coordination and support for the project, observed changes among teachers in the treatment group, and plans for further research engagement. Teachers' interviews centred on their own practices, including their experiences, challenges, and any observed changes in their students. The interview questions could be found in Appendix H.

6.7 Ethics

The study was approved by the School of Education Ethics Committee at Durham University (Appendix I). The intervention in this study delivered research evidence-informed teaching strategies to teachers and provided materials and suggestions for implementation. Schools selected programmes from a menu and could adapt the provided materials to teaching plans, so this study did not make any major changes in teaching aims and plans. At the participant recruitment stage, headteachers and teachers were given a brief overview of the study. Teachers voluntarily participated in the study and could withdraw from the experiment at any time. The surveys were anonymous and all identifiable information was deleted after the data of individual teachers and students were linked.

6.8 Conclusion

This chapter clarified and justified the major components of conducting an RCT, including the development of the use-of-research-evidence intervention, sample, randomisation, and outcome measurement. It also described the methods of analysing different types of data for the impact analysis and the process evaluation.

Results

Chapters 7 to 15 report the findings of this study. According to the research cycle (Gorard, 2013, p. 14), this thesis begins by reporting existing evidence on teachers' use of research evidence. Chapter 7 is about the overall situation of teachers' access to research relevant information and predictors of teacher-perceived usefulness of research evidence. Chapter 8 discusses the most promising approach to facilitate teachers' use of research evidence with consideration of different forms of evidence and dissemination approaches.

Chapters 7 and 8 serve as a preliminary step in the primary research. The primary research undertaken in this study focuses on evaluating the use of research evidence in the Chinese teaching context. The results start from Chapter 9 which reports the lessons from the pilot experiment. Chapters 10 to 12 present the impact analysis of the main experiment on students' Chinese literacy outcomes, their enjoyment of reading, and teachers' perceptions and skills in using evidence. The in-depth process evaluation, covered in Chapters 13 to 15, reports the results of the compliance analysis and describes how teachers and schools incorporated the use-of-research-evidence intervention and evidence-informed strategies into their daily teaching.

Chapter 7 Situation of teachers' use of research evidence: results from secondary data analysis

This chapter illustrates the analysis results from TALIS 2008 to 2018 and the NFER Survey of Teachers 2010 to answer the following research questions:

1. What are the current situation and emerging trends in K-12 teachers' access to research information on education worldwide? (Sections 7.1 and 7.2)
    - What is the situation of teachers' access to research information on education around the world?
    - What is the trend of teachers' access to research information on education around the world?

2. How do teachers perceive and engage with research evidence to improve teaching practice? (Sections 7.3 to 7.5)
    - How do teachers value the usefulness of research findings?
    - What is the strongest predictor of teacher-perceived usefulness of research evidence for improvement of teaching?
    - To what extent may school encouragement and colleague discussion support teachers' confidence in using research evidence?

7.1 Teachers' access to educational information around the world

TALIS 2018 explored teachers' access to educational information relevant to research across 48 participant countries or economies. This was captured by two items: 'Education conferences where teachers and/or researchers present their research or discuss educational issues' and 'Reading professional literature'.

Among all participant countries, 53% of the teachers had attended education conferences or seminars and 73% of teachers read professional literature in the past 12 months on average. Overall, reading professional literature might be easier to get access for teachers rather than education seminars or conferences that require resources, time, and funding.

The variation in teachers' access to educational information was noticeable across the participating countries and economies. Shanghai (China) had the highest percentage of access to education seminars or conferences (85%). However, only a quarter of teachers from Georgia and Saudi Arabia attended education conferences or seminars. Vietnam and Shanghai (China) had the highest

percentage of access to professional literature (approximately 94%), while Italy le had the lowest figure (25%). In other words, whereas nearly every teacher in Shanghai had access to professional literature, in Italy, only one out of every four teachers reported the same.

Alberta (Canada), Australia, Croatia, Kazakhstan, Latvia, Lithuania, Russian Federation, Shanghai (China), Slovenia, and United Arab Emirates had a higher percentage than the world average percentage in both access to education conferences and professional literature. The percentage for these two items in 15 countries including one North American country (33%), four South American countries (100%), nine Europe countries (32%), and one Asia country (11%), were lower than the world average. Table 7.1 illustrates the percentage of teachers who reported having access to education conferences and professional literature.

Table 7.1 Access to educational information relevant to research, TALIS 2018

Country	Conferences or seminars	Professional literature
Alberta (Canada)	79	74
Australia	59	86
Austria	50	89
Belgium	49	65
Flemish Community (Belgium)	53	79
Brazil	47	59
Bulgaria	40	79
Ciudad Autónoma de Buenos Aires (Argentina)	51	66
Chile	37	43
Colombia	53	71
Croatia	72	92
Cyprus	57	58
Czech Republic	29	87
Denmark	45	67
England (United Kingdom)	36	66
Estonia	53	90
Finland	35	75
France	49	51
Georgia	24	85
Hungary	34	89
Israel	58	72
Italy	53	25
Japan	63	72
Kazakhstan	66	79
Korea	43	85

Country	Conferences or seminars	Professional literature
Latvia	72	82
Lithuania	60	94
Malta	43	45
Mexico	45	64
Netherlands	40	89
New Zealand	50	89
Norway	36	75
Portugal	47	58
Romania	43	60
Russian Federation	70	91
Saudi Arabia	26	44
Shanghai (China)	85	93
Singapore	59	56
Slovak Republic	27	80
Slovenia	57	92
South Africa	56	60
Spain	50	47
Sweden	47	79
Chinese Taipei	72	64
Turkey	56	72
United Arab Emirates	63	76
United States	55	69
Viet Nam	49	94
Total	53	73

7.2 Trend of teachers' access to educational research information

This section compares the trend in the proportion of teachers who reported having access to educational information relevant to research from 2008 to 2018, including education conferences, individual or collaborative research, and professional literature. In each cohort, the percentage of teachers who had access to educational information in each country or economy was recorded (Appendix J). Overall, the results suggest that in terms of these simple questions that could be related to research use, not much has changed over time (Table 7.2).

Professional literature remained the most accessible source of educational research evidence for teachers, whereas opportunities for conducting individual or collaborative research were the least available, both showing a slight downward trend over time. Around half of the teachers reported attending education conferences or seminars to access research-related information. This figure

declined to 43% between 2008 and 2013 but rebounded to 53% in 2018. Of course, the cases in each year are not the same. Therefore, at least some of the apparent difference over time could be due to sampling variation.

Table 7.2 Trends of access education information relevant to research

	2008	2013	2018
Education conferences or seminars	50	43	53
Professional literature	79	--	73
Individual or collaborative research	38	35	--

TALIS data has limitations. TALIS only collected data from some of the countries or economies, therefore, the data could not be generalised to all countries around the world. Regions such as Shanghai (China), England (United Kingdom), Alberta (Canada), and Belgium (Flemish Community) may not fully reflect the national context in these countries. For instance, although Shanghai (China) illustrated a large proportion of teachers who had opportunities to read professional literature and attend educational conferences relevant to educational research, this figure hardly represents the situation in China. In addition, as mentioned in Chapter 4, the items do not directly reflect teachers' use of research evidence in their practice.

To understand teachers' views on the usefulness of research findings with more detailed data, this study investigated the Survey of Teachers 2010 that was the only open access dataset presenting detailed insights into teachers' use of research evidence. Ideally, it would be helpful to examine trends in how teachers in England accessed educational information from 2008 to 2018, to estimate the continued relevance of the 2010 survey. However, England only participated in TALIS in 2013 and 2018. As such, this study can only consider broader international trends regarding how teachers access educational information. The analysis for TALIS data indicated that in terms of these simple items that could be related to research use, not much has changed over time, which provides part of the rationale for analysing the Survey of Teachers 2010. Though the Survey of Teacher 2010 data was collected a decade ago, its results are still relevant.

7.3 Do teachers find research evidence useful for improving teaching?

Among the six common approaches to improve teaching, using research findings was a less commonly experienced approach and had the lowest usefulness rate. While almost all teachers (99%) had experience to reflect on their practice, 24% of the teachers had no prior experience using research

findings, which was the second largest proportion, following the figure for doing their own enquiry or research (29%).

Teachers who had a chance to use research findings were not fully satisfied with its benefit to improve teaching, where about a quarter of teachers (24%) found it useless to enhance their teaching. Teachers either hold a critical view of research findings or have no experience in using research findings in their teaching practice. On the other hand, self-reflection in daily practice was highly valued by teachers, with more than 96% of teachers who had experienced it finding it useful. Table 7.3 illustrates how teachers valued different approaches to improving teaching practice.

Table 7.3 Usefulness of ways to improve teaching (percentage)

	No experience	Not useful	Useful	Useful rate
Reflecting on my own practice	1	2	97	99
Reviewing my practice as part of school self-evaluation	13	14	73	84
Working towards my performance management objectives	10	26	64	71
Doing my own enquiry or research	29	9	61	88
Using the findings of other people's research	24	18	57	76
Using teacher assessment data	10	18	71	80

7.4 What is the strongest predictor of teachers' perceptions of using research findings?

The perceived usefulness of research findings did not relate highly to their background characteristics but largely related to their confidence in understanding of research findings and teacher-perceived usefulness of conducting their own research.

7.4.1 Variables in the model

Teacher-perceived usefulness of research findings for teaching improvement did not have a substantial association with their current working status, sex, and work experience, with odds ratios of approximately 1 (Table 7.4). Teachers with full-time contracts, female teachers or teachers who had worked in Key stage 1 or 2 were 10% more likely to adopt a favourable stance towards research findings. There was no large difference in a positive perception of research evidence between categories of ethnicity and experience in Key stages 3 and 4.

Table 7.4 Odds ratios of finding research evidence useful

		Usefulness of research findings		
		No	Yes	Odds ratio
Ethnicity	BME	72	100	1.00
	White	1,539	2,144	
Current Working Status	Part time	490	610	1.15
	Full time	1,148	1,656	
Sex	male	359	454	1.11
	female	1,295	1,828	
Experience in Key Stage 1	No	1,167	1,549	1.12
	Yes	487	733	
Experience in Key Stage 2	No	982	1,300	1.10
	Yes	672	982	
Experience in Key Stage 3	No	907	1,265	0.97
	Yes	747	1,017	
Experience in Key Stage 4	No	927	1,292	0.97
	Yes	727	990	

A logistic regression model was constructed to investigate the strongest predictors for teacher-perceived usefulness of research evidence to improve teaching. This model compared 2,282 in-service teachers who reported the usefulness of research findings for teaching practice with 1,654 teachers who did not perceive the usefulness, either found research evidence useless, or had never used research findings. The receiver operating characteristic curves and adjustment of model cut-off point can be found in Appendix K.

The base percentage correct for the logistic regression model was 58%, indicating that by simply assuming all teachers found evidence useful, the model was able to predict a teacher's attitude towards the usefulness of research findings with a 58% accuracy rate. The model organised the variables into three blocks and applied a forward function. Overall, the model could predict 74.3% of teacher-perceived usefulness of research evidence or 39% of the otherwise unexplained variation.

The first block contained teacher background characteristics of sex, ethnicity, employment status, and key stages. However, these variables contributed only a marginal 0.2 percentage point increase in the accuracy of the model's predictions. Teachers who thought research findings were useful for their teaching did not vary by these characteristics.

The most substantial enhancement in predictive power, a noteworthy 11.6 percentage point, came from the block that included teachers' confidence and external support for using research evidence. The percentage correct increased by around five per cent by adding teacher-perceived usefulness of

conducting their own enquiries. This item was closely related to teachers' values about using research evidence, and therefore, added to the last batch. Table 7.5 presents the percentage of correct predictions and the incremental gains observed as additional blocks were added, providing a more interpretable measure of model performance.

Table 7.5 Percentage correct

Block	% Predicted correctly	Increase on previous figure
Baseline	58	--
Background	58.2	0.2
Confidence and support	69.8	11.6
Conducting own enquiries	74.3	4.5

7.4.2 Results of the logistic regression modelling

With the forward function, the model only included variables which increased the percentage correct of the predicted outcome. Teachers' background information was excluded in the regression model, which indicated the irrelevance of background information to predict the usefulness of research evidence.

Among all variables relevant to using other people's research findings, the finding suggested that teachers' confidence in understanding and transforming research findings into practice was the strongest predictor of the perceived usefulness of research evidence with a coefficient of 1.46. Teachers who were more confident in understanding and using research findings were more likely to view research evidence as useful. Specifically, each one-point increase in confidence was associated with a 46% rise in the odds of perceiving research as useful.

Encouragement to use research evidence from schools was another powerful predictor in this block with an odds ratio of 1.19. Access to research and discussion with colleagues were also predictors of usefulness but with much smaller predictive power (1.04 and 1.05 respectively). On the flip side, time constraints appeared to be an essential deterrent. When teachers perceived a one-point increased in time limitations, the likelihood of them considering research evidence useful took a hit, dropping by 14%.

In addition, teachers' perception of the usefulness of undertaking their own research was a powerful predictor. An odds ratio of 1.67 indicated that the likelihood of teachers finding research evidence useful increased by 67% with each incremental level of perceived usefulness of conducting their own research or inquiries. The coefficient of each included variable is shown in Table 7.6.

Table 7.6 Coefficient for each predictor

Blocks	Variables	$Exp(\beta)$
Block2	I know where I can find relevant research that may help to inform my teaching	1.04
	The research I have found I have been able to understand and use	1.46
	My school encourages me to use research findings to improve my practice	1.19
	I have frequent opportunities (at least termly) to discuss relevant research findings with my colleagues	1.05
	I do not have time to use research findings	0.86
Block3	Usefulness of doing my own enquiry or research	1.67
Constant		0.05

7.5 What is the role of support from schools and colleagues?

The logistic regression model revealed that teachers' confidence in understanding research evidence was the strongest predictor of their perception of its usefulness for improving teaching. However, the process from teachers becoming aware of how to access research evidence to developing the confidence to comprehend and apply it remains largely unexplored.

The path analysis focused on the procedure from finding research evidence to understanding research evidence with a specific focus on the support teachers receive from their schools and colleagues. This analysis illuminated the various steps involved and the role that external support played in helping teachers bridge the gap between accessing and understanding research evidence. Due to data limitations, this analysis could not fully capture the entire process from access to understanding research evidence, nor could it establish causal relationships between variables. It represented a relatively minor part of the thesis.

The model was built on the direct path from finding to understanding research evidence and contained the variables about colleague discussion and school support. Teachers' confidence in accessing appropriate research findings could independently explain 47.3% of the variance in their confidence in understanding and using research evidence. With the opportunity of colleague discussion and encouragement from schools, the model explanatory power rose to 50.2%. The small increase in R^2 suggested that the model offered a slightly improved explanation of the variance in teachers' confidence in understanding and using research evidence. However, the inclusion of these two variables was not sufficient to fully explain the entire process from accessing to understanding research evidence. Table 7.7 shows the model explanatory with the indicator R^2.

Table 7.7 Model explanatory power

Predictor	R^2
Finding research evidence	0.473
Colleague and school support	0.502

Access to research evidence was an essential step and almost a premise before understanding research evidence, as evidenced by a substantial coefficient of 0.598. When teachers were confident in knowing where to find relevant research that may help to inform teaching, they also felt confident in understanding and using that research evidence. In other words, the strong positive relationship between access to and understanding of research evidence suggested that ensuring teachers to easily reach relevant research findings was a critical factor in enhancing their confidence to comprehend and apply those findings in their teaching practice.

Encouragement to use evidence in practice from schools and opportunities to discuss research findings were positively related to finding and understanding research evidence. On average, teachers who knew where to get access to the research findings might receive more encouragement from schools and were more likely to get involved in discussions with their colleagues, with coefficients of 0.428 and 0.388 respectively. Furthermore, when teachers received more encouragement from schools and had a higher frequency to discuss research evidence with colleagues, they tended to be more confident in their ability to understand and use research evidence in their teaching practice. Figure 7.1 illustrates the standardised coefficient of each path which represents the shift in the outcome variable for a one-standard-deviation variance in the predictive variables.

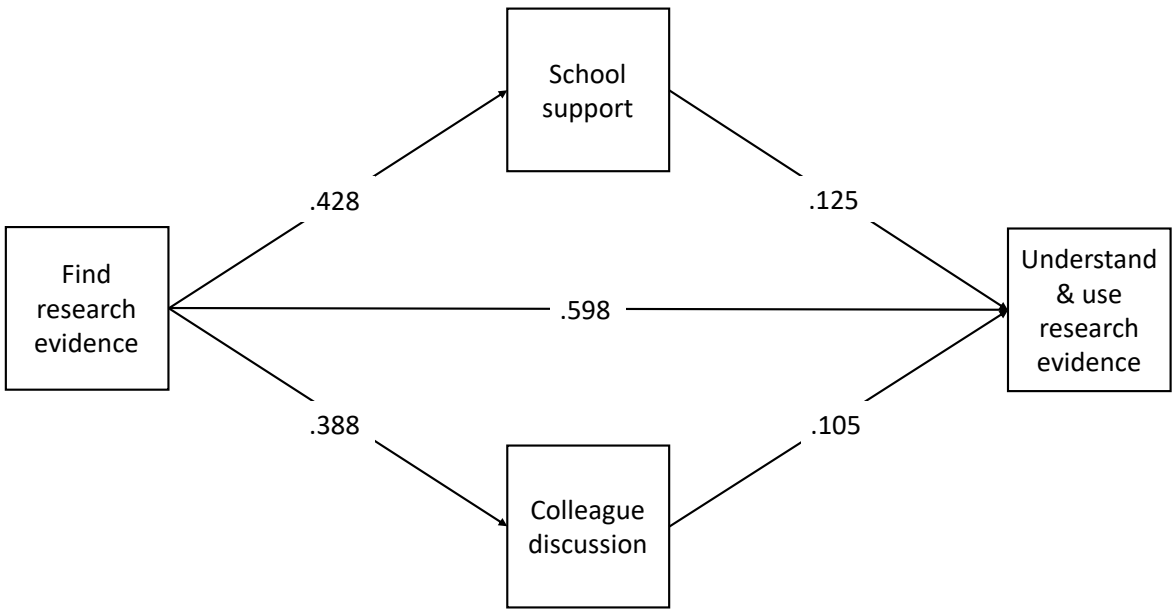


Figure 7.1 Coefficients for the path analysis

Table 7.8 represents the ratio of the indirect path 'effect' and the total path 'effect'. The ratio indicated the extent to which a supportive and collaborative environment within schools might make a difference in teachers' confidence in understanding and using research findings. Teachers' confidence in accessing research findings was strongly linked with their confidence in understanding and using research evidence with a total path 'effect' of 0.688. Of this, 6% was mediated through colleague discussion and 7.7% could be explained by a supportive school environment.

Table 7.8 Path 'effect' and ratio

Predictor	Path 'effect'	Ratio of indirect 'effect' over total 'effect'
Finding research evidence	0.688	--
Colleague discussion	0.041	6.0
School support	0.053	7.7
Total-mediation	--	13.7

The model still emphasised the importance of teachers' ability to find appropriate research findings to inform teaching, which was largely relevant to their confidence in understanding and using research. However, teachers did not navigate this journey alone. School climate of using research findings for teaching practice and active teacher collaboration were both essential components that may complement teachers' individual efforts in understanding and employing research evidence effectively.

7.6 Conclusion

This chapter introduced the data analysis results from TALIS and Survey of Teachers 2010. Some limitations of the datasets were mentioned. TALIS could only provide an insight to teachers' access to some research relevant information. And the data in some countries included a small region of that country such as Shanghai, China, and Alberta, Canada. The Survey of Teachers was conducted 15 years ago, which may not fully represent the current situation of teachers' attitudes towards research. Therefore, the results should be interpreted carefully.

Professional literature was a popular approach for teachers to get access to research information in education, comparing with educational conferences or seminars. Based on TALIS data from 2008 to 2018, there was not much change in the percentage of teachers who had access to professional literature, educational seminars or conferences, and individual or collaborative research.

Taking a further step about teachers' attitude towards research, the NFER Survey of Teachers 2010 suggested that approximately half of the in-service teachers in England did not think using research

evidence beneficial for their teaching or had no experience of using it. The strongest predictor was teachers' confidence in their ability to understand and use research evidence. Teachers who had a negative view on the usefulness of research were more likely to express the time limit for using evidence. In addition, 'research-rich' school environment and opportunities to discuss research evidence use with colleagues may potentially facilitate the process of accessing to understanding and use research evidence.

Chapter 8 The most promising approach to encourage teachers' use of research evidence: results of structured review

The structured review of prior research was taken to explore the most promising approach to encourage teachers to use research evidence and improve students reading achievement with the following research questions:

What is the best approach to facilitate teachers' use of research evidence, particularly in teaching literacy?

- What is the impact of teachers' use of research evidence on students' literacy outcomes?
- What is the most promising approach to encourage teachers' use of research evidence, in order to improve students' literacy outcomes?

The structured review screened 2,257 studies from five databases and finally includes 13 pieces of work that have a teacher level intervention about teachers' use of research evidence in mainstream schools, and a focus on literacy. The results are presented in this chapter.

8.1 Descriptive analysis

In total, 41,800 records were identified through ProQuest, Scopus, Web of Science, BEI, and ERIC, with more than half retrieved from ProQuest. However, most of these records were not directly relevant to the research focus. This is largely because terms related to 'evidence-based' practices are commonly mentioned in a wide range of publications, even when they do not specifically refer to interventions designed to support teachers' use of research evidence. Before deduplication, 28 potentially relevant studies were identified. The searching results are shown in Table 8.1.

Table 8.1 Searching results in each database

Database	Identified	Screened	Relevant
ProQuest	27,571	300	11
Scopus	3,675	400	2
Web of Science	1,163	496	10
BEI	82	82	2
ERIC	9,309	975	3
Total	41,800	2,253	28

Despite these electronic datasets, four reports about teachers' use of research evidence conducted by EEF were added to the records. Identified studies were sorted in order of relevance to the topic and screened by title and abstract. In total, 2,257 studies were screened. After deduplication, 27 studies met the inclusion criteria based on title and abstract. Twelve studies were excluded because they focused on special education students only, were not teacher intervention, or were not relevant to students' literacy outcomes. Only 13 were deemed relevant to the research questions and had met the inclusion and exclusion criteria. The process of this structured review is illustrated in Figure 8.1.

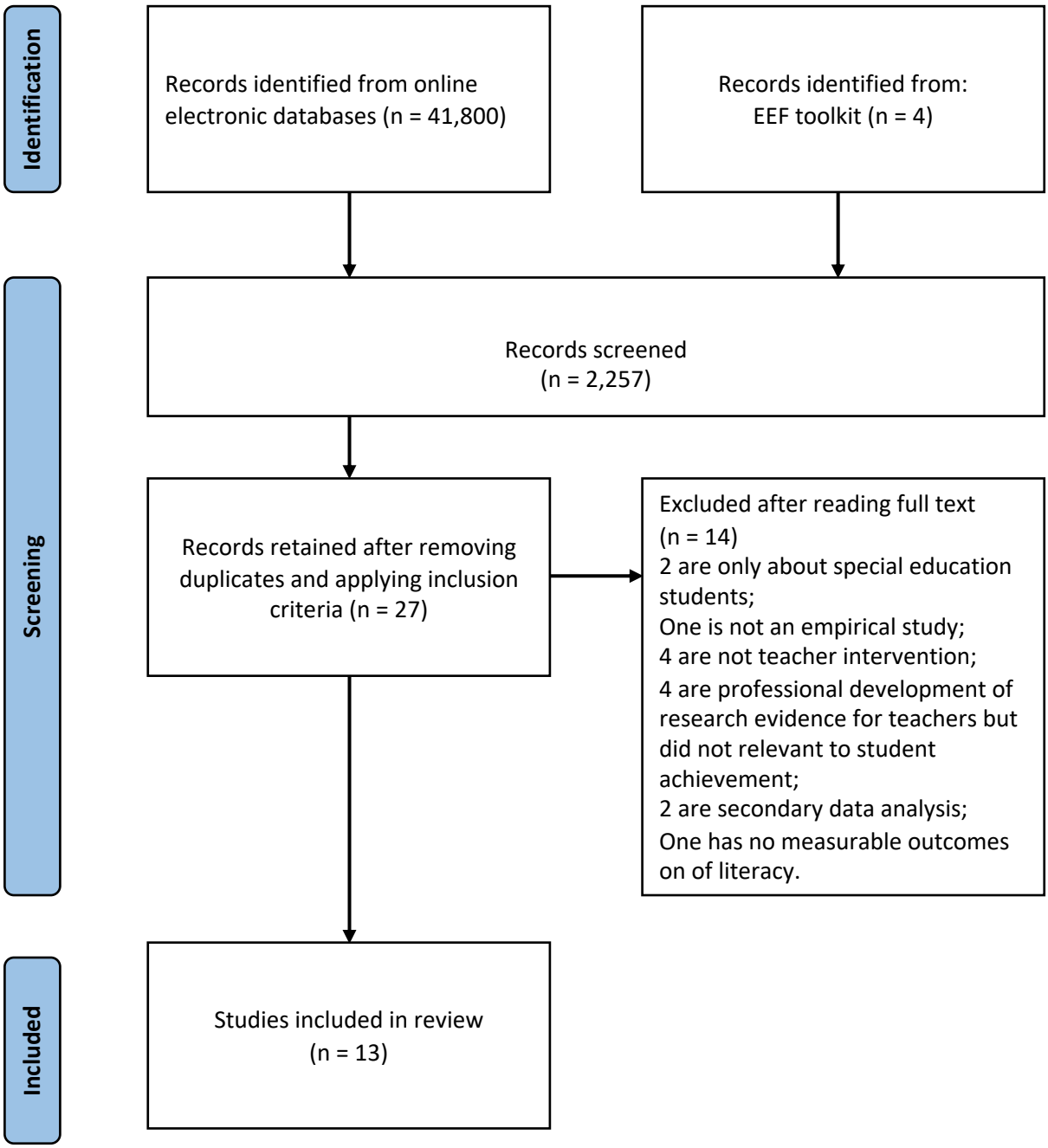


Figure 8.1 PRISMA flow diagram

This structured review evaluated the quality of findings from the aspects of research design, scales, missing data, and measurement quality, based on the 'SIEVE' (Gorard, 2021, p.94). Although knowledge on teachers' research engagement remains limited, recent studies indicate that encouraging teachers to use research evidence may have potential to improve students' literacy outcomes, though findings were often mixed or showed limited effects. Five of the thirteen studies illustrated a positive effect while eight studies illustrated a mixed or no impact. No negative impact was found. Table 8.2 is a cross-tabulation that summarises the impact of teachers' use of research evidence on reading attainment, categorised by quality. The detailed quality rating for each study can be found in Appendix L.

Table 8.2 Summary of impact and research quality

	Positive (5)	No impact/ Mixed (8)	Negative (0)
4 *	--	3	--
3 *	1	1	--
2 *	2	2	--
1 *	2	1	--
0 *	--	1	--

Among the most robust five studies that were rated with 3 or 4 asterisks, only one of them showed a small positive impact on students' reading. All of these high-quality studies were conducted in the UK. Seven studies were conducted in the US, and one was in Norway. Participants in these studies were teachers from kindergartens (1 study), elementary schools (12 studies), and secondary schools (2 studies). Two studies involved different Key Stages.

Overall, the current body of research indicates limited evidence on the effectiveness of teachers' use of research evidence. The most reliable studies (rated 3 asterisks or higher) mainly suggested a mixed impact on students' literacy outcomes. While the effect remained unclear, there was still potential for improvement, as no high-quality study reported any substantial negative impact.

8.2 Approaches for research evidence use

The review summarised approaches to encourage teachers' use of research evidence based on the two dimensions of practitioner's research evidence use (Gorard et al., 2020), as explained in Section 2.2.2. The framework classified studies into nine categories according to the format of presenting research evidence and user engagement in the dissemination process. Evidence can be in plain, modified, or engineered format and can be disseminated passively, engagingly, or (inter)actively to users. This review did not find any study that used an (inter) active transfer. Most studies in this review

had researcher-engagement in dissemination and provided teachers with engineered evidence such as evidence-based or research-based teaching materials.

8.2.1 Passive transfer of plain evidence

One study was found to disseminate plain research evidence passively, which employed a quasi-experiment design in Bexley, England to investigate the impact of teachers' evidence-use on Year 2–6 students' English attainment (See et al., 2016). The study recruited nine primary schools for the intervention and five local primary schools as the comparison group. Students' attainment was also compared with 49 other state-funded primary schools in the borough.

Since the treatment and control schools were not randomly assigned or matched, the comparison between two groups might be biased and imbalanced. The design of the study was not as strong as a randomised controlled trial for a causal question. Therefore, the design was rated with 3 asterisks. 1,677 students participated in the intervention while 3,364 students from local primary schools were in the control group. No school dropped out. Students' outcomes in Year 6 were measured by Key Stage 2 standardised tests. The quality of this research remained 3 asterisks.

The intervention delivered the meta-analysis conducted by Hattie and colleagues (Hattie & Timperley, 2007). Actions were taken by schools and teachers themselves, including the starter pack prepared by school project leads and training conducted by school leads. The implementation followed an action research model through four circles with different activities and aims. Teachers audited each other's lessons and received feedback from school research leads.

The results of the study focused on impact on students' literacy outcomes and feasibility of teachers' research evidence use. The study illustrated a mixed result on students' literacy outcomes. For Year 6 students, the effect sizes were +0.04 and +0.06, compared with 5 comparator schools and 49 state-funded schools respectively. Combining all intervention students, a small and negative impact was found with an effect size of -0.04 for reading and -0.05 for writing. The impact on FSM-eligible students were observable with effect sizes of +0.17 for reading and +0.12 for writing but the sample size of FSM-eligible students was limited. The passive dissemination of plain evidence brought teachers space to explore but also created challenges. The process evaluation of this study pointed out the risks of using plain evidence including teachers' ability to understand academic papers correctly, coherent interpretation of formative feedback, and implementation with consistent 'success criteria'. The summary of this approach is shown in Table 8.3.

Table 8.3 Summary: passive transfer, plain evidence

Reference	Research quality	Impact			
		Student literacy scores	Teacher knowledge/	Teacher attitude	Teacher practice
See et al. (2016)	3 *	?	--	--	Feasibility of evidence-use

Note: +positive impact, -negative impact, ?mixed impact, --not applicable.

8.2.2 Passive transfer of modified evidence

Two studies evaluated the effectiveness of passive transfer of modified evidence in the elementary phase. The high-quality study reported mixed effects, whereas the lower-quality study found positive impacts on both students' literacy and teaching practices.

The 'Literacy Octopus' Dissemination Trial had a clustered randomised controlled design and involved 12,500 primary schools in England with a low attrition rate below 1% (Lord et al., 2017a). The study rigorously evaluated the effectiveness of teachers' evidence-use.

'Literacy Octopus' involved four experimental arms and one comparison arm. The evidence-based literacy materials were prepared by four providers involving Durham University, University of York, ResearchEd and NatCen, and Teaching How2s. The evidence was selected, summarised, and organised in a more readable format such as booklets, posters, special magazine issues, presentations from high profile academics and practitioners, and websites. The modified evidence remained the noticeable elements of research. For example, the Institute for Effective Education at the University of York sent 'Best Evidence in Brief' that included research findings and links to relevant articles.

This research suggested a trustworthy finding that passive dissemination had almost no impact on students' reading achievement at Key Stage 2. The effect size was approximately +0.01 and the average scores of all four arms of dissemination were just slightly higher than the figure for the comparison group.

Taylor et al. (2005) proposed a CIERA School Change Framework and evaluated it during a two-year project. The study was rated with one asterisk mainly because of its pre-post design that had no proper comparison group. The study involved more than seven hundred students from 13 schools in Connecticut, North Carolina, Iowa, Minnesota, and California. The study reported that 733 students

completed both reading fluency assessments, 723 students for reading comprehension, and 577 students for writing.

Schools set up teacher study groups and discussed the materials provided by the project. The project created a CIERA School Change website including research summaries on effective reading instruction, effective schools, and effective school improvement. Some research articles were downloadable for teachers to discuss research-based reading practices relevant to the study group's focus area. The website also included video samples of effective practice and suggestions for study group activities. The activities were organised by school leadership teams, while researchers provided monthly monitoring meetings.

Students showed improvement in reading fluency, comprehension, and writing from pre- to post-test. The study revealed a positive correlation between the level of school reform effort and student outcomes in reading. For every 1-point increase in the reform-effort score, schools saw an average increase of 1.34 in students' comprehension scores and 4.87 in oral reading fluency scores. No strong association was found between school reform effort and writing performance. Teachers in high-reform-effort schools were also more likely to adopt research-based reading instructional practices and reported making greater shifts towards evidence-informed teaching compared to their counterparts in low-reform-effort schools. However, without a control group, it is difficult to attribute these gains directly to the CIERA School Change project. The research quality and impact are illustrated in Table 8.4.

Table 8.4 Summary: passive transfer, modified evidence

Reference	Research quality	Impact			
		Student literacy scores	Teacher knowledge	Teacher attitude	Teacher practice
Lord et al. (2017a)	4 *	?	--	--	--
Taylor et al. (2005)	1 *	+	--	--	+

Note: +positive impact, -negative impact, ?mixed impact, --not applicable.

8.2.3 Passive transfer of engineered evidence

One study estimated the impact of passive transfer of engineered evidence on students' literacy outcomes and teachers' instructional quality (Table 8.5). The impacts on student' literacy and teachers' practice remained unclear.

Edmonds' (2007) study was rated with one asterisk. The researcher conducted a school-level randomised controlled trial among five primary schools to explore whether increasing teachers' instructional knowledge could enhance students' academic success. Schools were randomly assigned to two groups using either comprehension or vocabulary instructional practice. Since schools all received one of the interventions, the comparison could not be fully fair. 16 teachers and 309 students from five primary school in the US participated in the study. 24% student data were missing. The sample size was too small to conclude the impact of teachers' instructional knowledge.

Comprehension instructional practices in this study contained three modules, providing teachers with training, lessons, and materials for offering students a new comprehension strategy. Teachers attended three 90-minute teacher study team meetings to discuss implementation. The team reviewed the literature on effective comprehension strategies to identify those that consistently produced exceptional outcomes and were best suited for content area education. The teaching materials were designed by researchers who had both classroom and research expertise in comprehension instruction.

The study constructed a hierarchical linear regression model to investigate the correlation between the intervention and students' comprehension achievement. It found that students' post-test results could be largely predicted by their pre-test results, while teachers' instructional quality and fidelity to the evidence-informed comprehension intervention had limited explanatory power. Instructional quality had slight positive relationship with mean classroom comprehension achievement. Since the average scores of the treatment group and the comparison group were not reported, the effect size could not be estimated for this review.

Table 8.5 Summary: passive transfer, engineered evidence

Reference	Research quality	Impact			
		Student literacy scores	Teacher knowledge	Teacher attitude	Teacher practice
Edmonds (2007)	1 *	?	--	--	?

Note: +positive impact, -negative impact, ?mixed impact, --not applicable.

8.2.4 Engagement in transfer of modified evidence

Five studies investigated the impact of using modified evidence with researchers engaging in transfer, with three studies rated with 3 asterisks or higher. Based on limited research about teachers' evidence-use, the impact of modified evidence with engagement in transfer was better established

compared with other formats. Overall, the impact on students' literacy scores was mixed while there might be positive impact on teachers' knowledge and views on using research evidence use.

Greenwood et al. (2003) provided 13 evidence-informed teaching practices such as phonics awareness, repeated reading, or writers' workshop, which could be adapted by teachers according to their needs in collaboration with researchers. Participant teachers were observed to use the evidence-informed strategies during the experiment. The study claimed success from school-wide use of an evidence-based literacy approach, particularly for silent reading. But the study was in only one school with 16 teachers and around three hundred students and compared the evidence-based approach with outcomes for students from a different grade level anyway.

Tortorelli and Bruner (2022) investigated the impact of a research-practice partnership to address teacher-generated questions related to how children learn to spell words. A pre- and post-test design was used, which lacked appropriate comparison group. The study was conducted with only 17 teachers, and the attrition rate was not reported clearly. To evaluate the impact of evidence-use, the quality of this study was too low to provide a robust result.

Studies by Rose et al. (2017), Lord et al. (2017b), and Wiggins et al. (2017) employed school-level randomised controlled trials with a large sample size. The result from Wiggins et al. (2017)'s study was less secure due to its high attrition rate of 9% to 13%. All three studies used national-wide standardised test such as Key stage 2 test and GCSE to measure students' reading attainment. Therefore, these studies had high quality and the results were more trustworthy.

The Research Learning Communities (RLC) project organised school social network over two years (Rose et al., 2017). There were 14 RLCs in total and each comprised Evidence Champions from up to five schools. An RLC held a one-day session four times a year to review research and evidence in an agreed-upon area of focus. Prior to the first session, the intervention team provided schools with a list of prospective subjects. Schools indicated from this list the possibilities that they were interested in studying, and a shared area which incorporated schools' interests was recommended by the intervention team. The Evidence Champions used the concepts from each workshop to develop, implement, and evaluate evidence-based school- or key-stage-wide improvement strategies. The goal was for school staff to undertake an improvement strategy and collect data to better understand its impact.

The effect size of RLC on Key Stage 2 reading was +0.02 with an effect size of +0.04 among Free School Meal students. When schools chose a literacy focused programme, the effect size was larger (+0.09). The impact on KS2 grammar, punctuation and spelling was +0.10. The intervention improved teachers' positive disposition to academic research in informing teaching practice and use of academic research to inform selection of teaching approaches, with an effect size around +0.20. Treatment teachers were more likely to find research evidence helpful and receive encouragement from schools. However, the intervention did not increase teachers' knowledge of research findings and methods with an effect size of +0.005.

The 'Literacy Octopus' Trial evaluated both active and passive research evidence dissemination interventions with 823 schools (Lord et al., 2017b). The interventions were provided by 4 academic institutions in England involving Institute for Effective Education (University of York), Teaching How2s, Centre for Evaluation and Monitoring (Durham University), and ResearchEd and NatCen. Schools receiving passive interventions were given evidence-informed materials through magazines, websites, literacy conferences, and booklets.

Despite the passive intervention materials, schools in the active arm were invited to evidence events and webinars. This active dissemination involved the teacher level engagement with evidence-based teaching strategies, the leader level decision-making about evidence-based projects, and peer-to-peer discussion about research evidence. For example, the CEM provided all schools in experiment groups with evidence-based materials, involving the *Improving Reading: A Guide for Teachers*, a '*Top Tips*' card, a staffroom poster, and monthly classroom posters. In the light active arm, there was a professional development session carried out by a researcher to introduce the booklet and its evidence base. Schools in the advanced active arm get an online diagnostic tool (InCAS) and extra peer-learning sessions to better understand students' learning status and discuss appropriate teaching strategies.

The overall study has a mixed impact on Key Stage 2 reading achievement. Interventions from IEE and How2s had slight negative impact on students' Key Stage 2 reading attainment, while active dissemination from CEM and ResearchEd resulted in small positive effect (+0.03 and +0.01 respectively). The 'Literacy Octopus' also measured teachers' and schools' degree of research use in terms of awareness, understanding, and action by research use survey. All active dissemination increased teachers' positive disposition to academic research in informing teaching practice and uses academic research to inform selection of teaching approaches with effect sizes varying from +0.11 to

+0.54. The effect size of teachers' knowledge about evidence-use was not reported but the mean scores of intervention groups and the comparison group were reported. On average, teachers from intervention groups had more knowledge about evidence-use. Findings about teachers' research use should be taken with caution because only a few teachers participated in the survey.

Wiggins (2019) proposed a research-informed school improvement model including deciding what you want to achieve, identifying possible solutions, giving the idea the best chance of success, evaluating the impact of decisions and identifying possible improvements for the future, and securing and spreading change. Apart from website resources, the project involves eight initial training sessions, follow-up meetings, and workshops for school leaders. The research randomly allocated 40 secondary schools into the control group and the treatment group. Positive effects were found for English (+0.05) and Mathematics (+0.09) in the first year. The effects diminished in the second year. The attrition rates were 9% and 13% in the first and the second year respectively, which makes the trial 3 asterisks. While the study investigated teacher outcomes based on the NFER Research Use Survey, the results were not reported due to the low response rate (18%).

Table 8.6 summarises the impact of delivering modified evidence with researchers actively involved in the transfer process.

Table 8.6 Summary: engagement in transfer, modified evidence

Reference	Research quality	Impact			
		Student literacy scores	Teacher knowledge	Teacher attitude	Teacher practice
Rose et al. (2017)	4 *	?	?	+	--
Lord et al. (2017b)	4 *	?	+	+	--
Wiggins et al. (2019)	3 *	+	--	--	--
Greenwood et al. (2003)	1 *	+	--	--	+
Tortorelli and Bruner (2022)	0 *	?	--	--	--

Note: +positive impact, -negative impact, ?mixed impact, --not applicable.

8.2.5 Engagement in transfer of engineered evidence

Four studies were identified that used engineered evidence to facilitate teachers' evidence-use in an engaged approach. Two of them claimed positive impact on students' literacy outcomes. However, all studies were rated with two asterisks or lower, which highlighted that the findings of teachers' use of engineered evidence with engagement in transfer based on these studies were not trustworthy (Table 8.7).

Benner et al. (2023) and Jacob (2017) conducted randomised controlled trials, while Wang et al. (2020) and Andreassen and Braten (2011) employed quasi-experiments to explore the impact of teachers' evidence-use. Though these studies had appropriate comparison groups, they were rated with 2 asterisks due to the small sample size or a high attrition rate.

Benner et al. (2023) constructed the Integrated Literacy Study Group to provide online training to teachers about effective reading instruction. Teachers were provided with personalised coaching. Although they particularly focused on students who risked of emotional and behavioural disorders, this study was not excluded because the sample students were not fully in special education need. The Hedges' g effect sizes on students' Dynamic Indicators of Basic Early Literacy Skills (DIBLS) Correct letter sounds and DIBLS Whole words recorded were +0.56 and +0.27 respectively. Although the treatment teachers did not show more knowledge of evidence-based reading than the comparison group, their self-efficacy and positive engagement in the programme were increased.

The Literacy Design Collaborative was a national community of educators, which providing teacher-designed and research-based framework, online tools, and resources for literacy. Wang et al. (2020) found no impact among primary school students but positive impact on secondary school students who performed 0.281 standard deviations above matched comparison students. The study claimed positive impact on teachers' practice and attitudes towards teaching skills according to self-report.

Jacob (2017) investigated the impact of Evidence-Based Literacy Instruction on primary school students' reading. The intervention involved a three-day in-person teacher training with EBLI's multisyllable strategy of segmenting, blending, and phoneme manipulation. After the training, teachers received 15 lesson plans, additional practice activities, and supplemental resources. Teachers were also supported by online consultation resources and three on-site instructional visits. The study found no difference in the attainment of students in grades 2 to 5 at seven primary schools in the US. There were no differences between treatment and control teachers in their self-reported confidence in use of evidence-based instructional practices. Treatment teachers used evidence-based practice for reading text and working sounds and words more frequently. They were also less likely to implement practices that were explicitly encouraged by EBLI. One of the problems faced was the difficulty for the teachers of implementing the approach. Many teachers then found it difficult to incorporate EBLI techniques into their reading instruction throughout the rest of the school year.

Similar issues of feasibility arose in Andreassen and Braten (2011). The intervention involved 5 teachers and 103 students. Teachers attempted to use four instructional practices, which were derived from the principles of relevant background knowledge, reading comprehension strategies, reading-group organisation, and reading motivation. No impact was found for students' reading comprehension. The effects of reading comprehension instruction may depend on how an intervention is implemented in classrooms. While teachers had appropriate implementation of relevant background knowledge and reading comprehension strategies, they poorly employed reading-group organisation and reading motivation.

Table 8.7 Summary: engagement in transfer, engineered evidence

Reference	Research quality	Impact			
		Student literacy scores	Teacher knowledge	Teacher attitude	Teacher practice
Benner et al. (2023)	2 *	+	?	+	--
Wang et al. (2020)	2 *	+	--	+	+
Jacob (2017)	2 *	?	--	?	?
Andreassen and Braten (2011)	2 *	?	--	--	?

Note: +positive impact, -negative impact, ?mixed impact, --not applicable.

These five studies illustrated researchers' engagement in the dissemination process of research evidence and interaction with research users. However, they did not provide specific training to increase teachers' knowledge about research evidence in education and therefore, there was no solid evidence on teachers' knowledge.

Teachers might be in the situation of using a certain strategy or material which was identified as evidence-based without consideration of their research basis. They still could not get access to research evidence, make judgement, and transfer research to practice independently. When the project is finished and teachers could not get extra support from researchers, they may not have the awareness of use research evidence to improve teaching.

8.3 Summary: A potential approach

This structured review illustrated different dissemination approaches and forms of research evidence, which suggested the feasibility of teachers' use of research evidence via various ways. Although there was no clear indication of the impact of teachers' use of research evidence on students' literacy

outcomes, it appeared to be easier to influence teachers' attitudes and even their behaviour. Among the studies that both met high-quality criteria (more than two asterisks) and assessed teachers' attitudes, most reported positive impacts. Some studies have demonstrated that it is possible to influence teaching practice through specific evidence-use interventions, including lists of evidence-based literacy approach (Greenwood et al., 2003), research papers (See et al., 2016), and collaborative partnerships between researchers and teachers (Wang et al., 2020). However, other studies have shown only small changes in teachers' instructional quality (e.g., Edmonds, 2007).

Three studies noted the issue of implementation quality (Andreassen & Braten, 2011; Jacob, 2017; See et al., 2016). Jacob (2017) surveyed teachers' satisfaction with training and the perceived usefulness of provided materials, revealing a low satisfaction rate of 55% and a neutral view on usefulness, with 60% agreeing. The studies of Andreassen and Braten (2011) and See et al. (2016) found that evidence-informed projects were poorly implemented in teaching practice.

These findings underscore the importance of carefully monitoring how teachers use evidence to ensure it is applied effectively. They raise the issue of whether using research evidence can ever be considered appropriate if it is not feasible for teachers to implement. Researchers may need to better understand teachers' needs and pain points in their teaching practice to improve the quality of training and intervention materials, making them more relevant and suitable for diverse teaching contexts. In this way, teachers might be more engaged with evidence-use and show higher degree of compliance.

Among the various approaches, the engagement transfer of modified evidence was the most frequently studied approach to support teachers' instructional practices. The high-quality studies disseminated modified research evidence and organised active workshops to facilitate evidence-use and engage teachers in the process of adapting research evidence. Most high-quality studies assessed its impact on students' reading attainment, with one reporting a positive effect. While there was no conclusive evidence of improvement in students' language learning in mainstream schools, the three high-quality studies (Lord et al., 2017b; Rose et al., 2017; Wiggins et al., 2019) showed the cost-effectiveness of the teachers' evidence-use projects. The RLC project costed £3.24 per student, while the RISE intervention costed £0.86 per student per year. The Literacy Octopus project, which involved various intervention suppliers, had costs ranging from £0.09 to £10.77 per student per year, with most interventions costing less than £4 annually per student. Importantly, these projects allowed schools

to choose their focus for implementing research evidence, integrating it into regular teaching with minimal disruption.

It may be easier to change teachers' perceptions and even behaviour compared to students' learning outcomes. Teachers can feasibly use research evidence through various approaches, but successful implementation often faces challenges, such as low teacher satisfaction and poor application in practice. Finding effective evidence for using research in teaching is urgently needed, especially given the current trend emphasising 'what works' and the push to 'use evidence' in education. Tailored support and understanding teachers' needs are crucial to enhancing engagement and compliance. Although the educational impact is mixed, some projects demonstrated cost-effectiveness, offering a practical option for schools and posing little to no harm to students.

Chapter 9 Lessons from the pilot study

The pilot study was conducted to assess the feasibility and implementation process of the use-of-research-evidence intervention, as well as to evaluate the clarity of the survey items and statements. It was carried out in two primary schools in Hefei, Anhui, China during May 2023 to July 2024. 46 teachers and 2,082 students from grades 3 to 6 participated in the trial. The flowchart of the pilot is illustrated as follows (Figure 9.1).

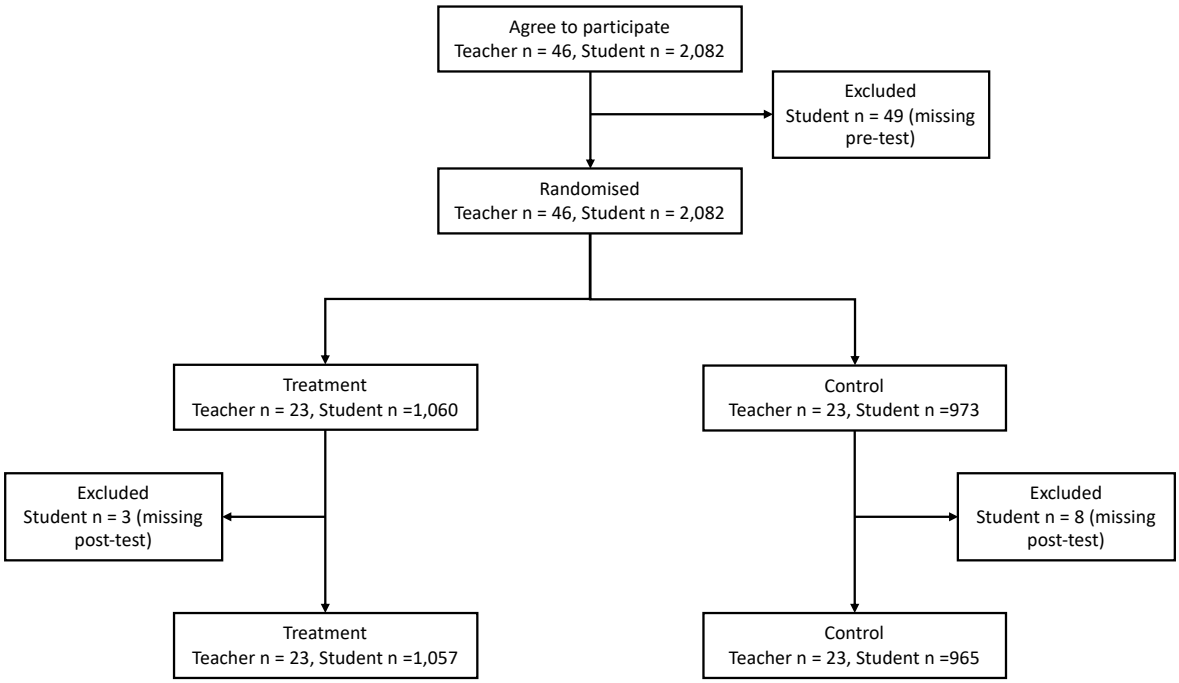


Figure 9.1 Flowchart of the pilot study

School leads selected the strategies to be implemented according to their interest and feasibility. School I implemented peer tutoring in grades 3 to 5 and Writing about Values in grade 6. School II chose dialogic teaching for grade 3 students, peer tutoring for grade 4, and Writing about Values for grades 5 and 6. The intervention presentations took place during the first week of May and the follow-up workshops were held at the beginning of June. The overall implementation was around six weeks. The pilot study found no impact on students' Chinese literacy. Detailed description of sample and impact analysis results of the pilot can be found in Appendix M. This chapter focuses on how teachers incorporate research findings into teaching practice in the pilot study and summarises the lessons learned and subsequent adjustments made for the main study.

9.1 Teachers' adaptation of research findings into teaching practice

This section presents teachers' exploration and adaptation of evidence-informed teaching strategies in their classroom teaching contexts. Eleven classes at School I and four classes at School II were observed during the intervention. Although most teachers adopted some evidence-informed strategies to varying extents, the quality of their implementation differed considerably. Using research evidence effectively requires teachers to adapt it flexibly to their specific teaching contexts.

9.1.1 Peer tutoring

Peer tutoring involved multiple classes and teachers, and therefore, teachers' implementation and feedback were diverse. Among the thirteen observed peer tutoring classes, all teachers arranged students to help each other in classroom activities. However, some teachers demonstrated notable fluency, successfully incorporating this strategy into their teaching environment, while others stiffly added peer tutoring to their teaching without amending their existing teaching plans. They left a few minutes for students to discuss before they announce the correct answer, so the duration of peer tutoring was lower than required.

Eleven teachers emphasised the peer tutoring rules at the beginning of their lessons. Teachers allocated students to different groups (2-4 students per group) and assigned students with a number. When they asked students to read the text aloud, for example, they required the student No.1 in each group to read first while other students listen carefully. During peer discussion, teachers invited one student from each group to summarise their group's response. Classroom observations revealed two main issues related to peer tutoring activities: students generally lacked a clear sense of time during these sessions, and there was considerable variation in students' peer tutoring skills. Teachers adopted different strategies to address these challenges.

To manage students' lack of time awareness during peer tutoring, teachers often took charge of setting and monitoring the duration of discussions. Some teachers directly stated, 'We have five minutes for discussion.' But the concept can be quite vague, especially if they are not keeping track of time or do not have access to a watch. A more effective approach observed in six lessons involved teachers using electronic blackboards or student iPads to display a countdown timer after announcing the discussion time. This method helped students visualise the time limit and supported better classroom pacing.

Another issue was that students demonstrated varying levels of peer tutoring skills, prompting teachers to tailor their guidance to different groups. While some students communicated fluently and offered constructive feedback to support their peers, others encountered difficulties. For example, students were asked to summarise the development of the story of Jiguang Huang in no more than 20 Chinese characters. Given the complexity of the narrative, this task proved challenging and led to intense group discussions. Students engaged in intense group discussions. One student proposed an answer, while another classmate remarked, 'I disagree, your summary lacks a plot point.' Another student suggested, 'Your answer exceeds the word limit, you can remove these two characters.' The teacher did not disrupt these students' discussions but listened to them carefully. In contrast, some groups struggled to complete the tasks. A few students remained silent throughout the discussion, unsure how to begin. Some students simply took turns reading out their responses, with no one offering feedback or challenging their peers' ideas. At this stage, the teacher gave them scaffolding supports, such as rephrasing questions, offering clues, and reminding them of relevant reading comprehension strategies.

To better monitor the quality of student discussions, teachers, especially those experienced in handling such situations, maintained a focus list of students needing extra support and paid closer attention to them. They didn't interrupt the discussions immediately but observed first. When they noticed students struggling, they stepped in with guiding questions to help them think through the problem step by step.

Techniques drawn from research could not always be directly applied in classroom settings. Some teachers actively revised the strategies to fit their teaching needs and created tools to support the peer tutoring process. For example, three teachers created a new activity which encourages students to discuss a challenging and open-ended question instead of the 'prediction relay'. According to the research lead at School II, this activity is more adaptative to the General Chinese lessons.

> We think that prediction relay is more suitable for reading a novel. But in the textbook, novels are one kind of reading material. We can implement peer tutoring to teach other non-fiction materials with the 'challenging and open-ended questions'.
> Once students get the Chinese book, they will read the novels first. So, most of they have already know what happened in the stories.

At School I, two teachers in grade 3 developed a task sheet for students. One teacher asked students to write down their peers' performance of reading aloud and rate them according to the national curriculum standard. At the end of the sheet, a blank was left for students to provide suggestions to their peers. Another teacher asked students to read the text, answer the questions on the sheet, and discuss their answers with peers.

Although some classes have successfully integrated peer assistance into their teaching, others treat it merely as a school requirement, lacking genuine engagement. Teachers fail to design meaningful central questions for peer tutoring activities, resulting in exercises that are unnecessary or superficial. For instance, during a reading session on Zikai Feng's work Fingers, students were asked to highlight descriptions of each finger and discuss which was the most beautiful. The question was expected to be open-ended with students' different ideas. However, all students highlighted almost the same sentences and provided standardised answers. Students expressed their ideas with the format 'I like the …, because…[citing the text].' The teacher demanded students to discuss but the whole process did not reflect any elements of peer tutoring because the answers were already evident. Moreover, the question itself was overly simple for students in grade 5, leaving little room for deep thinking or meaningful exchange.

9.1.2 Dialogic teaching (Philosophy for Children)

Only one teacher implemented P4C in two classes, which ensured a relatively consistent implementation process. For the two dialogic teaching sessions, Piggybook and Willy the Wimp were selected as reading materials. The lessons followed the ten steps of P4C, starting with sitting in a circle. The teacher first introduced the classroom rules such as listening to others carefully, keep thinking, and expressing loudly.

The picture book was quickly skimmed through on the electronic blackboard. Students first posed questions within their groups, and the teacher collected all the questions for whole-class discussion. Then, the teacher followed the guidance from the Philosophy for Children China website and used coloured cards for students to vote on the questions for further discussion. However, in the first lesson, the teacher noticed that some students changed their votes after seeing others' choices, which conflicted with the core principles of P4C. To address this, the teacher asked students to write the number of their chosen question on the card before voting. They also replaced paper cards with washable plastic ones for reuse in future lessons.

However, peer dialogue was limited, with most interactions occurring between students and the teacher rather than among students themselves. Although students were encouraged to repeat the previous speaker's opinion before sharing their own, few actively responded to or built on their peers' ideas. As a result, discussions often stalled, with students offering repetitive and isolated points. Only a small number of students were conscious about either supporting or disagreeing with others' viewpoints. Greater attention to peer responses could have helped deepen the conversation by encouraging agreement, disagreement, or elaboration.

The teacher's responses to students' ideas sometimes lacked openness and flexibility, which hindered the flow of discussion. When one student suggested that Willy was brave because he was rich, the teacher immediately challenged the view, questioning its alignment with moral values. This premature judgment embarrassed the student and disrupted the discussion. Only after another student offered supporting evidence did the teacher recognise its potential and use it to guide further dialogue. It might be better if teachers could devote less time to judging students' views.

9.2 Do teachers care about research evidence?

A majority of teachers misunderstood evidence-informed teaching at first and expressed their confusion. While most teachers showed interest in recent research findings, they paid little attention to issues of quality or how to apply the findings in practice. During the intervention presentation, most teachers became familiar with the concept of using research evidence but were more concerned with practical evidence-informed strategies and how to implement them. Through informal discussion, the presentation about using research evidence might be more beneficial for headteachers and school research leaders to make decisions and construct their own research projects rather than to classroom teachers.

Many teachers were interested in research evidence especially popular strategies in England. They were enthusiastic when talking about research resources.

> How did they implement this strategy in England?
> Could you find some research papers according to our interests please?

They requested more research materials to learn about education development in other countries. For them, the free accessibility to research papers might be a barrier.

> We want to access the research papers but most of them are not free.

However, it might be difficult for teachers to understand the idea of using research evidence at the first time of knowing it. Teachers easily confounded 'empirical study' and 'teaching based on research evidence'. For example, during the presentation about evidence-informed teaching, one teacher asked:

> Is what you are doing the same as empirical studies?

Some teachers conflated the quality of research evidence with the effectiveness of an intervention. They assumed that if a strategy was research-based, it would automatically produce positive outcomes in any context.

> … these strategies have been evaluated so we do not need to worry about their effectiveness.

Time and effort appeared essential for deepening understanding. For instance, teachers in the experimental group at School II carefully read the training booklet and organised a workshop to discuss it before the main presentation. They said that they had a better understanding of what to do after the use-of-research-evidence presentation. When teachers were asked to rate the quality of some example studies, they could quickly and accurately rate them and express the reason, which was encouraging.

In addition, headteachers and school research leaders may benefit more from the use-of-research-evidence intervention, particularly with evidence-informed decision-making. At the preparation stage, headteachers and research lead at School I selected the strategies which would be implemented in the following two months with consideration of the strength of research and the effect size. When discussing Writing about Values, their first reaction was suspicious and they hesitated.

> Students in grade six are busy passing the final examination. Would that take too much time? This strategy does not have a large effect size, but it is easy to implement. It provides detailed instruction and writing materials, so teachers do not need extra preparation.

The headteacher at School II had ample experience in conducting their own enquiry. From their perspective, the value of the current project is to help the school conduct high-level research.

It is a good opportunity to get access to world-advanced education research. We could also get inspiration from the project and conduct our own research and apply for funding.

Research design is essential for a school research project.

According to this headteacher, school research projects in China were mainly case studies. They develop a certain curriculum or perform a typical lesson without design or balanced comparison. They thought this project provided schools with challenges as well as a chance to improve their research. During the summer vacation, they applied for a school research project at the provincial level.

We would like to use the design 'randomised controlled trial' that you mentioned in your presentation. It is exciting to evaluate the real effectiveness of my project.

9.3 What do teachers need to apply evidence-led teaching strategies?

Three strategies were selected to be implemented in this pilot study, involving Writing about Values, philosophy for children, and peer tutoring. Different supporting materials were provided to teachers, as appropriate. Teachers requested more supporting materials and concrete examples, such as classroom videos, lesson plans, or opportunities to observe others' teaching, to facilitate implementation.

9.3.1 Peer tutoring

Fifteen of twenty-four (62.5%) treatment teachers used peer tutoring in this pilot study. This study introduced how teachers in Scotland implemented peer tutoring, including matching up students with their reading attainment and the rules of peer tutoring. Fuchs' work about peer-assisted learning was also included to provide example activities such as partner reading, paragraph 'shrinking', and prediction relay. Teachers' requirements could be classified into two aspects: example lessons and detailed peer tutoring standards for each grade. They required teaching examples to get visual experience of peer tutoring so that they could imitate the correct action according to the examples.

Do you have an example lesson or lesson plan so that we could organise the class according to it?

Will the school research lead provide a demonstration lesson using peer tutoring?

School research leads were invited to prepare example lessons in advance and the school research leads sent the class structure to the researcher before their implementation. However, only a few teachers in School I participated in the example lesson observation, which potentially influenced the implementation fidelity.

Experienced teachers noted that students' learning goals were different in each grade, so they wondered if the goal of peer tutoring would be adjusted according to their grades. They noted that the provided example activities may not adequately challenge students across all grade levels.

> Is there a standard for each grade?
> These example activities are too easy for grade five students.
> I think for each grade the goal of using peer tutoring is different. Can you provide such documental support?

Additionally, many teachers doubted that students could effectively manage learning tasks through peer tutoring and expressed concerns about the potential impact on students' learning outcomes. They may struggle to recognise the trustworthiness and value of research evidence, often relying more heavily on their own prior experience and intuition.

> I always want to monitor students' learning results. I encouraged students to learn from their peers, but I am still worried about whether they acquire the knowledge.
> I asked students to read aloud in groups, but the change is not observable. I found that students good at reading still present well but others remain in the same ability.

9.3.2 Dialogic teaching

The dialogic teaching was employed via philosophy for children. The researcher explained the ten steps of philosophy for children and attached the link to P4C China. However, they were still not sufficient for practice.

> The steps are vague. When you ponder each step, you will find more details to be considered.

During the investigation process, the teacher was consistently concerned about whether their responses in class were appropriate. This highlighted the potential value of providing teachers with a checklist to facilitate self-reflection on their classroom practices.

I am not sure if my reaction to students is correct.

I want to move forward with discussion but do not want to intervene too much.

I am not clear when I should give them support and instructions.

In addition, directly applying P4C in Chinese literacy classes presented certain challenges. This may be attributed to the traditional Chinese literacy teaching style, which often favours a single correct answer. Teachers tend to focus more on the transmission of knowledge, sometimes at the expense of fostering students' critical thinking skills. As a result, students are often unaccustomed to asking questions. The difficulties in implementing P4C suggested that some preparatory lessons on how to formulate high-quality questions and how to articulate one's opinions could be introduced beforehand.

Most of the questions posed by students are 'yes or no' questions and 'what' questions, which may be relevant to our daily practice. A few of them could pose 'why' questions and start a meaningful discussion.

In Chinese literacy lessons, students are not trained to pose questions. Even teachers are not good at posing good questions, …, I mean questions suitable for critical thinking and discussion.

9.3.3 Writing about Values

Writing about Values (WaV) provided teachers with a checklist and writing tasks. Before implementation, the presentation emphasised the key points be noticed. Some of them were astonished at first and repeatedly assured that the writing tasks would not be corrected. They were worried about the effect of this writing task without teachers' correctness or feedback.

If we do not correct students' writing, how could we know their progress and how to give them feedback?

But some students may write non-sense paragraphs. Do we control that?

Teachers were told that the task itself matters but not the writing content, which gave students a chance to reflect on themselves. Their mission was to make sure students participate in and complete the tasks. No teacher posed questions about the implementation process.

9.4 Adjustment for design and methods

The pilot study provided evidence and details to refine the main experiment, making it more rigorous. These adjustments included a more balanced randomisation process that accounted for teacher assignment and approaches to reduce the risk of intervention contamination. For data collection, unclear survey statements that could cause misunderstandings were revised. Additionally, classroom observation forms and an interview plan were introduced, and process data were collected from multiple dimensions to support the process evaluation. Moreover, to enhance the generalisability of the results, a more diverse sample should be recruited for the main experiment.

9.4.1 Randomisation

While the main study keeps the teacher-level randomised controlled trial design, more complex situations should be considered. The stratification should not only be at the school, and the grade levels but also consider teacher assignment. Teachers who teach more than one classes should be selected out and matched for randomisation, which guarantees a balanced class number and student number in each group.

In more extreme cases, a school may select a strategy that can only be implemented by a limited number of teachers, and only within specific reading lessons. To obtain an appropriate comparison group, the randomisation can be at the class level. Skilled and willing teachers are directly assigned to treatment classes for the specific reading lessons, with these teachers also participating in training sessions. The treatment classes receive the intervention in the specific reading lessons provided by the skilful teachers.

9.4.2 Contamination

The pilot study used teacher-level randomisation and paper-version materials to minimise any potential contamination. However, teachers in the control group may still get access to some of the materials such as inadvertent discussions with intervened teachers or seeing materials brought back to the office by teachers in the experimental group. Some schools may even disseminate the experimental materials as part of their own educational resources without the researchers' consent. In the main study, the researcher should emphasise and remind schools and treatment group teachers not to share the materials with others.

In addition, schools may have implemented some of the example strategies in their daily teaching, which may underestimate the effect of this project. The main study could advise schools to choose the strategy that they have not used yet or at least not in a well-structured way.

9.4.3 Survey revision

Teacher Use-of-Research-Evidence Survey and Student Enjoyment of Reading Survey provided essential insights into teachers' perspectives and competencies regarding evidence-use, as well as students' attitudes toward Chinese reading. Collecting survey data online can save time and effort compared to manual collection, potentially reducing processing errors.

In the pilot study, students provided various forms of identification beyond their class IDs, such as registration numbers. This variability made it challenging to link data consistently. Moreover, there was a possibility that students' class IDs were adjusted over time, and some students were not even aware of their registration numbers. To address these issues, the main experiment collected students' names in addition to their class IDs on the survey. This additional identifier could help ensure that students' background information, exam results, and survey responses can be accurately linked, thereby improving the reliability of the data analysis.

Several modifications were made to the student survey to improve clarity and relevance. The question 'How often do you read these materials because you want to?' was revised to 'Could you rate the following reading materials according to your preference?' to make the wording clearer. To gain a more comprehensive understanding of students' reading habits, a new question was added, covering aspects such as borrowing books, online reading, paper book reading, and reading purposes.

Questions about students' attitudes towards reading and their self-rated reading ability were separated to provide more distinct insights. The question on leisure reading time, 'How much time do you usually spend reading for enjoyment?', was removed, as students did not report their reading time accurately in the pilot, either due to fragmented reading patterns or a lack of time awareness. Lastly, to improve accuracy, the unit of time for Chinese homework was changed from hours to minutes, as students naturally tended to answer in minutes in the pilot.

The expression in the teacher survey should be revised to avoid misunderstandings regarding what the question asks. For example, the first item asked teachers to indicate how much they trust each source for decision-making by rating them from 0 to 10. However, when administered as a paper

questionnaire in one school in the pilot, some teachers only selected one or two sources they trusted the most. To clarify the intended meaning in the main experiment, the statement for this item should change from 'most trust' to 'how much you trust' each source. Similarly, in the second item, the phrase 'the best description' should be revised to 'how much do you agree with' to better capture the intended evaluation.

On the other hand, the main experiment also refined the sub-items used to measure teachers' skills to use research evidence. For instance, Question 6 originally aimed to assess teachers' ability to judge research quality by comparing two study designs: 'Collecting reading scores and peer tutoring experiences of 300 students' versus 'Interviewing 40 students about their experience of peer tutoring'. However, neither approach directly addresses the impact of peer tutoring on reading, making the distinction less meaningful. In the main experiment, this question was revised to compare 'Selecting one class as a peer tutoring class and comparing their reading scores with other classes' versus 'Interviewing 40 students about their experience of peer tutoring', ensuring a clearer contrast between the two approaches. Similarly, the items in Question 8 were adjusted to better align with the idea of using research evidence and to improve clarity.

All teachers and students should be required to respond to the survey once with their best answers in advance. When teachers or students submit additional responses, they should specify their reasons. Teachers should report these reasons to the school research lead, while students should inform their teachers. If they do not mention the reason, their first completed answers will be used in data analysis.

9.4.4 About the process evaluation

At first, a compliance analysis was planned according to workshop attendance sheet. However, there was not much difference about teachers' attendance to the intervention activities during such short-term intervention period. The process evaluation of this pilot study was mainly based on informal interviews and classroom observation. The main study should record more details and information such as teachers' attendance, frequency of using the strategy, and quality of implementation. These data could be used for compliance analysis to investigate teachers' compliance and the impact of use of research evidence on students' Chinese literacy achievement. The data also support qualitative comparative analysis in the main study to identify which combinations of teacher and implementation process characteristics play a role in successful implementation and how they interact.

9.4.5 Generalisation

The pilot study was based on two primary schools under the same local authority, and they are part of the same school union. Also, the two schools both have high reputation, which ensure good teaching condition and attracts high-quality teachers and students. The generalisation of findings of this study might be restricted and less warranted due to their homogeneous. Therefore, the main study should recruit more schools from different local authority and have different teaching conditions.

9.5 Adjustment of intervention

Based on the challenges teachers faced during the implementation and adaptation process in the pilot, the intervention and supporting materials could be refined to better address their needs.

9.5.1 Use-of-research-evidence intervention

The basic knowledge of research evidence might be expected to be combined with ample examples to help teachers understand. Description of research evidence should be clearly stated to avoid teachers' misunderstanding. Knowledge of understanding impact and judging research quality was challenging for teachers to understand correctly in the pilot, therefore the knowledge should be illustrated in a more attractive form. Interaction between teachers and the researcher might be necessary so that the researcher could know about teachers' understanding of research evidence and provide adequate assistance.

Teachers could also have opportunities to make decisions based on research evidence. In this study, while the school leaders selected the strategies, it would be important that teachers were aware these choices were grounded in both research and the specific needs of the school. Better organisation and more interactive activities could increase teacher engagement with the experiment, helping them see it not as a burden, but as a way to enhance their teaching.

In addition, this intervention could be linked with a school-level project so that teachers might feel a sense of belonging and participation, rather than passive users. The project could encourage schools to build up their own project and teachers to be researchers and reflective practitioners. Therefore, teachers may have a feeling that they are doing a high-quality project following school research leads and researchers. Their voice and classroom practice were highly valued in this project.

9.5.2 Materials for evidence-informed teaching strategies

No school chose computer-assisted formative assessment in this pilot because it largely depended on a certain software. The main study should reconsider this strategy and investigate alternative approaches that offer greater feasibility and versatility for schools.

The current implementation guide and materials were not enough for teachers to use this research evidence. Teachers needed more comprehensive and well-structured guides such as detailed teaching examples or plans, resources, and direction from researchers and school research leads to understand and use the recommended teaching strategies.

Teachers should be encouraged to collaborate with colleagues to share best practices and experiences with adapting research evidence into practice. In the pilot study, some teachers showed inspiring ideas to better implement the strategies in the classroom while others had no idea. Schools could foster a culture of collaboration among teachers. This collaborative environment allowed teachers to exchange ideas, broaden their perspectives, and collectively work towards improving their teaching methods.

9.6 Conclusion

This chapter reported lessons from the pilot study, providing valuable insights that helped refine the design and implementation of the main study. The research plan was enhanced with additional details based on practical experience to address potential complexities in the main experiment. Ambiguous survey items were revised, and process evaluation tools were added to strengthen data collection. The delivery process of the use-of-research-evidence intervention in the pilot study also prompted critical reflection. Throughout the process, the researcher gained a deeper understanding of teachers' needs and realised that the materials provided insufficient guidance. These insights were used to optimise the supporting materials for the main study.

Chapter 10 Impact evaluation: students' Chinese literacy scores

Following the pilot study reported in Chapter 9, the main experiment involved four primary schools with teacher-level randomisation and another school that only agreed to select some teachers to receive the intervention. This chapter focuses on the following questions:

- To what extent does teachers' use of research evidence improve students' Chinese literacy outcomes?
- Does the impact of teachers' use of research evidence on students' literacy outcomes differ according to students' demographic characteristics (sex, ethnicity, and hukou)?

Among the four RCT schools, one school did the randomisation by itself while teachers from the other three were randomly allocated to the treatment and control groups by the researcher (see Section 6.2.2). Participants from the RCT schools consisted of 155 teachers and 7,125 students with 153 teachers uploading 6,791 students' test results for the two semesters, forming the basis for the main impact analysis in this chapter. Results for sub-groups are also presented. Students who only completed the post-test were included in additional analysis for the post-test only. The impact analysis including the quasi-experiment school can be found in Appendix N.

10.1 Description of participant students

This section outlines students' background characteristics and baseline Chinese literacy scores. Missing cases in post-tests are discussed for the potential interference with the impact analysis. The section ends up with a flowchart illustrating the cases used in the randomised controlled trial.

10.1.1 Student background

7,125 students' names with the grades and classes were found in school records. Students' background characteristics including sex, ethnicity, Hukou, migrant status, and low-insurance status were also collected in school records. However, background information for 303 students was missing in the school system. These students were still included in the impact analysis if they were recorded in the pre-and the post-tests.

The participant students were from grades 3 to 6 in the middle and upper stages of primary education, with most students concentrated in grades 4 and 5. More male students took part in the study. Most students were Han ethnicity, held urban Hukou, had no migrant background, and did not come from

low-income insurance households. The allocation to groups was almost balanced in terms of sex, ethnicity, migrant status, and low-income insurance status. The percentage of rural Hukou students in the treatment group was slightly higher than the figure for the control group. The proportion of students whose background information was missing was slightly larger in the treatment group than the control group. The proportion of each characteristic was calculated as the number of students with the characteristic divided by the total number of students in the treatment or control group, as shown in Table 10.1.

Table 10.1 Student allocation and background characteristics

		Treatment (%)	Control (%)	N
Grade	3	21.4	26.4	1,703
	4	28.8	25.1	1,921
	5	30.0	27.6	2,052
	6	19.7	20.9	1,449
Sex	Female	43.7	44.2	3,130
	Male	51.3	52.3	3,692
Ethnicity minority		1.3	1.3	92
Rural Hukou		25.4	23.7	1,749
Migrant student		7.9	7.6	553
Receiving low-income insurance		0.4	0.4	28
System missing		4.9	3.6	303
N		3,562	3,563	7,125

10.1.2 Student baseline Chinese scores

As explained in Chapter 6, three of the four RCT schools from the same district administered the same final examination while the fourth school had its own tests. Z-scores were used to minimise differences in the metrics.

The Chinese scores in the final examination for the Autumn semester 2023/2024 were extracted for 6,904 students from school records and noted as the pre-test result. The pre-test record involved 3,396 students from the treatment group and 3,508 students from the control group. A small pre-existing difference between the groups was found (Table 10.2). On average, students in the control group performed slightly better in the pre-test than students in the treatment group. The minor difference between the treatment and the control groups was represented by an effect size of -0.04.

Table 10.2 Comparison of students' Chinese scores in the pre-test, all students

	N	Mean	Standard deviation	Effect size
Treatment	3,396	-0.02		-0.04
Control	3,508	0.02		

Total	6,904	0.00	1.00	

Table 10.3 indicates that the pre-existing difference was found in both researcher-randomisation and school self-randomisation settings. The gap in the three researcher-randomised schools (-0.04) was slightly smaller than that of the self-randomised school (-0.05). This gap suggests that the self-randomised school likely did not subvert the randomisation by selecting outstanding classes for the experiment group.

Table 10.3 Comparison of the pre-test, by trial settings

		Researcher-randomised schools	Self-randomised school
N	Treatment	1,810	1,586
	Control	1,556	1,952
	Total	3,366	3,538
Mean	Treatment	-0.02	-0.03
	Control	0.02	0.02
Standard deviation		1.01	0.98
Effect size		-0.04	-0.05

The differences were larger in the younger grades, and effectively non-existent in the upper grades (Table 10.4). The distribution of students' pre-test scores across grades 5 and 6 was almost balanced, with effect sizes of roughly 0. However, the pre-test scores of students in grades 3 and 4 were imbalanced. Before the intervention began, the negative effect size was -0.10. This imbalance among grades might be related to how teachers are allocated. Most schools tended to be more cautious in assigning their more experienced and skilled teachers to upper grades. As a result, teaching quality in the middle grades may vary more noticeably, which could contribute to differences in student abilities and increase the likelihood of imbalanced pre-test scores. Alternatively, this could be explained by a random occurrence.

Table 10.4 Comparison of the pre-test, by grades

		Grade 3	Grade 4	Grade 5	Grade 6
N		1,657	1,834	1,985	1,427
Mean	Treatment	-0.04	-0.05	-0.01	0.01
	Control	0.03	0.05	0.01	-0.01
Standard deviation		1.00	1.00	1.01	1.00
Effect size		-0.07	-0.10	-0.02	+0.01

The randomisation was conducted at the teacher level independently without matching students' pre-test scores. The observed differences may exist due to missing data in the pre-test or could simply be coincidental or have arisen by chance. This imbalance, though small, was important to note because

it might affect how the results of the post-test (after the intervention) would be interpreted. Therefore, gain scores are reported in the following impact evaluation to control for the initial difference.

10.1.3 Missing data

7,125 students' records were found including school registration ID, name, year of entry, and class. 6,904 student's Chinese scores for the autumn term were available from school records while 6,988 students participated in the final examination for the spring term. In total, 6,791 students completed the two final tests in the autumn term of 2023 and the spring term of 2024.

In the post-test, records of 113 students were missing. Reasons for missing Chinese test scores mainly involved authorised absence, transferring to different schools, and data loss due to teacher mobility. Six students transferred to other schools during the semester, 46 students were in a class where the teacher dropped out, and the remaining 61 students had authorised absence during the final examination of the intervention semester. Each of the treatment and control groups comprised no fewer than 3,396 students. The NNTD of this sample size to observe an effect size larger than +0.05 would be 170. This indicated that at least 170 students would need to score one standard deviation lower than the average score of the comparison group to noticeably alter the overall effect size. Therefore, the 113 students who were absent from the post-test are unlikely to have had a substantial impact on the results.

To assess whether missing data skewed or biased the results, the pre-test scores of students who did not complete the post-test are shown in Table 10.5. The mean pre-test score of students who missed the post-test (mean = 0.13) was higher than those who completed the post-test in the treatment group (mean = -0.03). On the other hand, the average pre-test of control students who missed the post-test was -0.56 as opposed to a mean of 0.04 for control students who did the post-test. The missing cases in the treatment group tended to be stronger than the missing cases in the control group. This indicated the possibility that the overall average of the treatment group (but not necessarily the progress) could have been higher if these 30 treatment students had taken the post-test. Also, the average of the control could have been lower if these 84 control students had completed the post-test.

Table 10.5 Comparison of the pre-test scores of students missing post-test

	N	pre-test of missing post-test	Standard deviation	N	pre-test of non-missing post-test	Standard deviation
Treatment	31	0.13	1.00	3,365	-0.03	0.98

Control	82	-0.56	1.83	3,426	0.04	0.96

Students who did not participate in either the pre-test or the post-test were excluded from further analysis. Considering the existing difference between the two groups in the pre-test, the impact of the intervention on students' progress was estimated based on the gain scores of students who fully completed both pre- and post-tests. In addition, since the intervention was delivered at the teacher level, all students in the class participated in the study, unless their teachers declined to provide the Chinese literacy scores for the entire class. The scores of all students who participated in the post-test were also included in the analysis to avoid bias in assessing the impact.

10.1.4 Flowchart for the main analysis

The flowchart for students and teachers included in the main impact analysis is presented in Figure 10.1. The overall attrition rate was less than 2%, which is reasonable. The missing cases mainly came from control group at the self-randomised school where a teacher dropped out and did not upload post-test score.

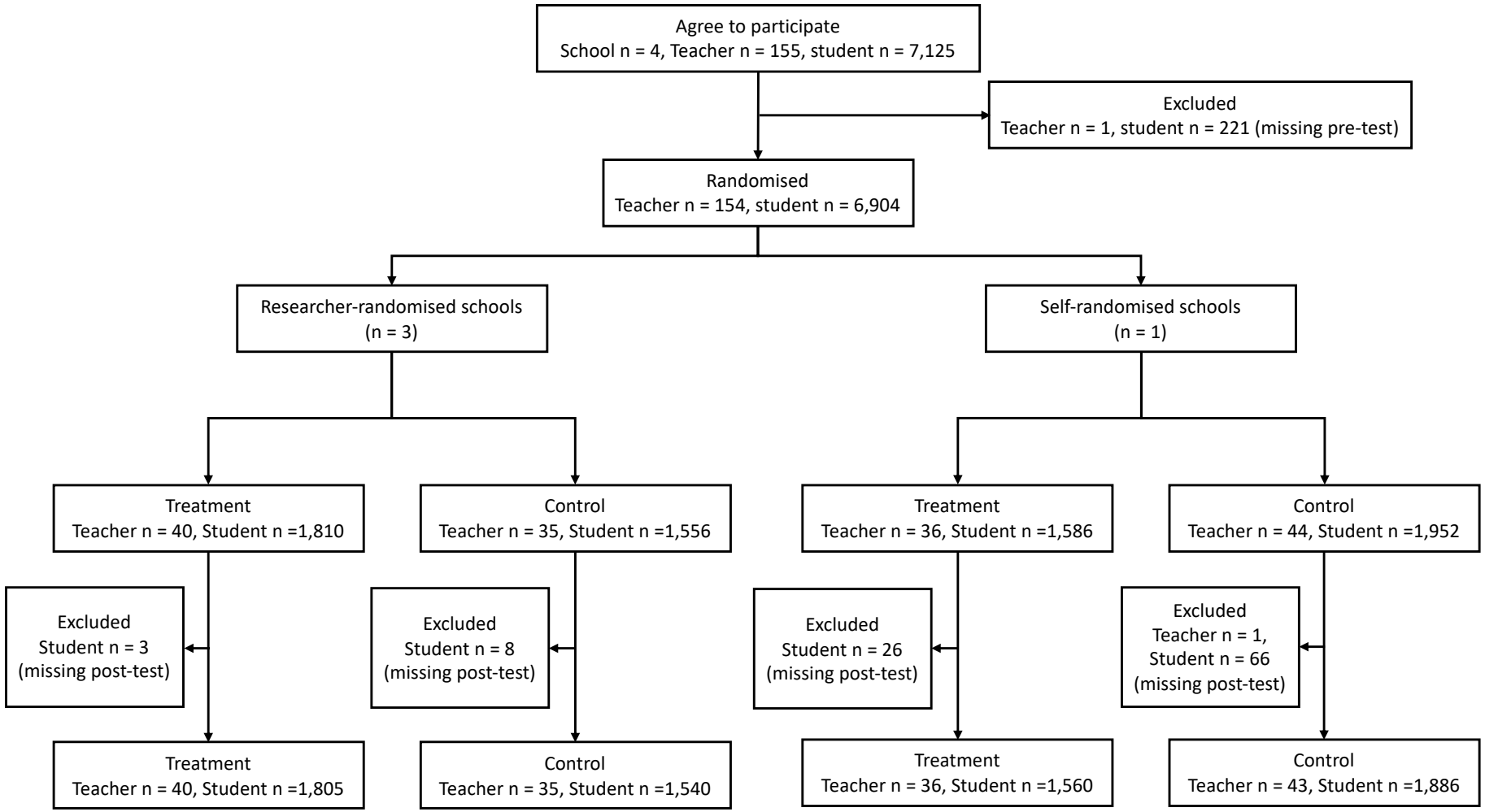


Figure 10.1 Flowchart of RCT schools

10.2 Impact on Chinese literacy gain scores

Students' Chinese literacy scores were the primary outcome for the impact analysis. Considering the small imbalance in the pre-test, students' gain scores were calculated to estimate the headline improvement during the intervention semester. A small but positive impact of teachers' use of research evidence was found on students' Chinese literacy test performance.

Treatment students did not perform better than the control group on average in the post-test, while they made slightly more progress than the counterfactual cases with an effect size of +0.07. Based on the EEF monthly progress table (Higgins et al., 2016), an effect size of +0.07 is considered equivalent to around an extra one month's reading age. The standardised pre-, post-tests, and gain scores, standard deviation of the post-test and the gain, and the effect sizes are reported in Table 10.6.

Table 10.6 Comparison of standardised pre-, post-tests, and gain scores, all students

	N	Mean			Standard deviation		Effect size	
		pre	post	gain	post	gain	post	gain
Treatment	3,365	-0.02	0.00	0.03			0.00	+0.07
Control	3,426	0.02	0.00	-0.03				
Total	6,791	0.00	0.00	0.00	1.00	0.75		

As explained in Section 6.5.7, statistical tests such as ANCOVA or similar significance tests were not appropriate for this study. However, as a sensitivity check, the number of cases needed to disturb the difference (NNTD) was calculated (see Section 6.5.3). The NNTD, based on an effect size of +0.07, was 236. This means that at least 236 individual cases would need to be reassigned to the treatment group with low gain scores of -0.78 (average gain score of the control group minus one standard deviation of gain) to eliminate the observed difference between the intervention and control groups. Therefore, although the effect size was relatively minimal, it still indicated potentially robust improvement in students' Chinese literacy. This result should be taken seriously.

The settings for school randomisation were slightly different, with one school doing the randomisation on its own and the other three randomised by the researcher. The researcher-randomised schools showed a stronger impact on students' post-test scores and gain scores. Treatment students from the three researcher-randomised schools performed slightly better than the comparison group in the post-test (+0.03) and demonstrated greater improvement (+0.08).

However, the effectiveness was somewhat less clear in the self-randomised school. Treatment students at the self-randomised school, still, performed worse than the control group in the post-test with a small negative effect size of -0.04. Though they made progress to catch up with the comparison group (+0.06), the progress was smaller compared to researcher-randomised schools. Variation between the two set-ups is illustrated in Table 10.7.

Table 10.7 Comparison of standardised post-test and gain scores, by trial settings

		Researcher-randomised school	Self-randomised school
N		3,345	3,446
Mean (post)	Treatment	0.01	-0.02
	Control	-0.01	0.02
Mean (gain)	Treatment	0.03	0.02
	Control	-0.03	-0.02
Standard deviation	post	1.00	1.00
	gain	0.78	0.72
Effect size	post	+0.03	-0.04
	gain	+0.08	+0.06

The difference between the two experimental settings indicates that the organisation of trials might influence the outcomes. First, the time for incorporating research evidence matters. While the experiment presentations were all held in March, the self-randomised school received the intervention two weeks later than others due to school arrangements and researchers' schedules. This school may get less exposure to the use of research evidence environment, potentially leading to smaller effect sizes compared to the schools that began earlier.

Second, the self-randomised school implemented the evidence-informed teaching strategies independently and used the provided materials as additional support. Additionally, the school organised the workshops involving both treatment and control group teachers, which may lead to contamination and a smaller effect size than 'real' impact. The researcher-randomised schools preferred to strictly follow the instructions from the provided materials and adjusted the implementation during practice. These are fully discussed in the process evaluation.

This study focused on students in grades 3 to 6, representing the middle and upper levels of primary school. Students at different grade levels have varying learning needs and may respond differently to changes in their teachers' daily practices. While treatment students at all grades showed some improvement during the intervention semester, students at grades 4 and 5 appeared to benefit the

most with gain effect sizes of +0.12 and +0.14 respectively. The impact on students' Chinese literacy attainment varied by grade, as shown in Table 10.8.

Table 10.8 Comparison of standardised post-test and gain scores, by grades

		Grade 3	Grade 4	Grade 5	Grade 6
N		1,645	1,818	1,929	1,399
Mean (post)	Treatment	-0.04	0.00	0.02	0.00
	Control	0.03	0.00	-0.02	0.00
Mean (gain)	Treatment	0.01	0.05	0.04	0.00
	Control	0.00	-0.04	-0.06	0.00
Standard deviation	post	1.00	1.00	1.00	1.00
	gain	0.76	0.76	0.76	0.84
Effect size	post	-0.07	-0.01	+0.04	0.00
	gain	+0.01	+0.12	+0.14	+0.01

10.3 Impact analysis for different strategies

The variations in performance between students of different grade levels could be related to the strategies they choose to use. Five evidence-led strategies were included in this experiment, each with a different estimated effect size. For instance, the estimated predicted impact of Writing about Values was +0.07, while the highest predicted estimated effect size was for peer tutoring at +0.30 (see Section 6.3.1).

While this study emphasised the use-of-research-evidence intervention as a unified whole, effect sizes for different strategies are reported as an exploratory insight to provide an overview of how each element might contribute to the overall result. The findings should be interpreted with some caution because this study was of a template and was not intended to evaluate individual strategies or the impact of any specific approach. In addition, the effect size for students who received Growth mindset was based on a small sample.

Compared to the other three strategies, embedding formative assessment showed positive effect sizes on both the post-test scores (+0.04) and the gain scores (+0.10). Students receiving peer tutoring and Growth mindset also made some progress during the intervention semester with gain effect sizes of +0.08. Students receiving Writing about Values performed better than their comparison group in the post-test (+0.05), but there was no evidence of their progress. Table 10.9 illustrates the effect sizes of the post-test score and the gain scores for each strategy.

Table 10.9 Comparison of standardised post-test and gain scores, by strategies

		Embedding formative assessment	Peer tutoring	Growth mindset	Writing about Values
N		2,005	3,958	134	694
Mean (post)	Treatment	0.03	-0.01	-0.08	0.02
	Control	-0.02	0.03	0.17	-0.03
Mean (gain)	Treatment	0.03	0.04	0.02	-0.02
	Control	-0.04	-0.03	-0.04	0.02
Standard deviation	post	1.00	1.00	1.00	1.00
	gain	0.63	0.76	0.71	0.97
Effect size	post	+0.04	-0.03	-0.25	+0.05
	gain	+0.10	+0.08	+0.08	-0.04

Teachers' implementation quality could have influenced the results. Whether teachers implemented the strategies correctly based on the materials provided by this study could make a difference. Some strategies such as peer tutoring and embedding formative assessment required frequent and consistent use. The implementation quality is analysed in the compliance analysis and the report for classroom observation (see Chapter 13 and Chapter 15).

10.4 Regression-to-the-mean effect

In Section 10.1.2, the average pre-test score of the treatment group was lower than that of the control group. This raised the concern that the statistical phenomenon of regression to the mean might be interfering with the impact evaluation. If students with low scores were more likely, or found it easier, to improve, then their slight prevalence in the treatment group could artificially increase the overall trial effect size.

Put another way, students with low pre-test scores may be more likely to show apparent improvement, posing the question of whether this improvement is due to the intervention itself or is simply a result of regression to the mean. This might suggest that the intervention has been successful, when in fact the improvement in the condition may just result from the fact that it is easier for students with low prior attainment to improve their score subsequently. Further analysis is advisable to distinguish the effect of the intervention or regression to the mean, particularly for extremely lower-attainment students.

The study first investigated the distribution of students' pre-test scores and categorised students into the highest 50% and the lowest 50%. The lowest 50% students in the pre-test made progress during the semester and performed better in the post-test compared to the pre-test. However, the average post-test score of the top 50% students decreased during the semester. The positive gain for the

lowest and the negative gain for the highest pre-test scorers suggested that regression to the mean may be interfering with the real impact of teachers' evidence use intervention. Table 10.10 presents pre-, post-, and gain scores for students at different pre-test levels regardless of the treatment or the control groups.

Table 10.10 Comparison of the lowest and the highest 50% students' gain scores (standardised)

	pre	post	gain
Lowest 50%	-0.59	-0.38	0.21
Highest 50%	0.59	0.39	-0.21

The estimation of regression to the mean hypothesises that the positive improvement in the treatment group is largely due to a shift in the uneven distribution: more students from the comparison group were in the highest 50%, and more students from the treatment group were in the lowest 50%. Therefore, due to the regression-to-the-mean, more control group students were likely to perform worse in the post-test, leading to a potential decrease in their average score, while the figure for the treatment group was more likely to increase.

Among the highest 50% of students, there were 146 more students in the comparison group. If the observed effect size were purely due to statistical tendency rather than genuine change, the intervention would appear effective simply because these 146 students in the control group had fewer opportunities to show further progress. Based on the hypothesis, the rest of 3,254 students should make no progress (as the mean overall gain score was effectively 0) during the semester while the 146 students in the comparison group had an average gain score of -0.21. According to the estimation formula in Section 6.5.8, the regression-to-the-mean effect size was +0.007. Likewise, the figure for lowest 50% students was +0.017.

The total regression-to-the-mean effect size was +0.02 which was smaller than the actual gain effect size (+0.07) in Table 10.6. This suggested that the gain effect size could not be explained solely or even largely by regression to the mean. The difference between the two figures, approximately +0.05, represented the portion of the observed improvement that could not be attributed to RTM. It is reasonable to infer that this difference (at least +0.05) likely reflected the contribution of the evidence-use intervention, though relatively modest. The effect of regression-to-the-mean was estimated, as shown in Table 10.11.

Table 10.11 Estimation of RTM effect

		N	gain	Standard deviation	RTM effect
Highest 50%	Treatment	1,627	-0.20		+0.012
	Control	1,773	-0.21		
	Total	3,400	-0.21		
Lowest 50%	Treatment	1,738	0.24		+0.008
	Control	1,653	0.17		
	Total	3,391	0.21		
Total students		6,791	0.00	0.75	+0.02

10.5 Subgroup analysis

The following analyses divided students into subgroups based on students' Chinese proficiency levels, sex, hukou, migration status, and low-income status.

10.5.1 Impact on students of different Chinese proficiency levels

This study divided students' Chinese proficiency levels into four groups involving the low achieving (lowest 30%), lower-middle achieving (lowest 30% to 50%), upper-middle achieving (50% to 70%), and high achieving (highest 30%). Treatment students at the lower-middle achieving level in Chinese showed improvement during the intervention semester (+0.16). High achieving students in the treatment group continued performing well in the post-test (+0.13), though only small progression was found (+0.04). No major impact was found among low achieving and upper-middle achieving students.

The headline impact appeared to be largely driven by gains among lower-middle achieving students. Variation across subgroups was inevitable anyways but the improvement for lower-middle achieving students was relatively consistent. The impact on students with different reading proficiency levels is illustrated in Table 10.12.

Table 10.12 Effect sizes of students at different reading proficiency levels

	Effect size	
	post	gain
Low achieving	+0.02	+0.04
Lower-middle achieving	+0.15	+0.16
Upper-middle achieving	-0.03	-0.04
High achieving	+0.13	+0.04

10.5.2 Impact for different sex

Regardless of their group allocation, female students (Mean = 0.15) performed better than male students (Mean = -0.12) in the pre-test. This advantage for female students remained and even

became larger in the post-test, indicated by a higher average score among female students compared to males. The standard deviation of female students in the post-test was smaller than that of male students, which suggested that the Chinese test scores of female students was more consistent and concentrated.

Students of different sex showed varying levels of academic achievement and progress during the intervention (Table 10.13). The intervention showed more positive impact for female students (+0.13), which was higher than the overall effect size (+0.07) and that of male students (+0.04). Negative impact on students whose background information was missing. The treatment group kept a lower attainment score in the post-test and did not make as much progress as the comparison group did.

Table 10.13 Comparison of standard pre, post-tests, and gain scores, by sex

		N	Mean			Standard deviation		Effect size	
			pre	post	gain	post	gain	post	gain
Female	Treatment	1,481	0.12	0.18	0.06			+0.03	+0.13
	Control	1,532	0.17	0.15	-0.03				
	Total	3,013	0.15	0.17	0.02	0.79	0.70		
Male	Treatment	1,720	-0.14	-0.14	0.00			-0.02	+0.04
	Control	1,793	-0.10	-0.12	-0.03				
	Total	3,513	-0.12	-0.13	-0.02	1.12	0.79		
Unknow	Treatment	164	-0.10	-0.09	0.01			-0.19	-0.21
	Control	101	-0.07	0.10	0.15				
	Total	265	-0.09	-0.02	0.07	1.01	0.66		

10.5.3 Impact on ethnicity minority students

Of the 92 ethnic minority students, 90 participated in both the pre- and post-tests. Overall, these students performed above the average in Chinese literacy, and they benefited slightly more from the intervention, with a gain effect size of +0.10 (Table 10.14).

Table 10.14 Comparison of standard pre, post-tests, and gain scores, ethnicity minority

	N	Mean			Standard deviation			Effect size		
		pre	post	gain	pre	post	gain	pre	post	gain
Treatment	46	0.23	0.20	-0.03				+0.02	+0.10	+0.10
Control	44	0.22	0.13	-0.09						
Total	90	0.23	0.16	-0.06	0.59	0.67	0.57			

10.5.4 Impact for rural Hukou students

Based on 1,749 identified rural Hukou students, 1,682 of them completed the pre-and the post-tests. Rural Hukou students had lower average scores in the pre-and the post-tests than overall average scores (around 0.00). They might meet more challenges when learning Chinese vocabulary, reading, and writing. For rural Hukou students, the average gain score of the treatment group was higher than the figure for the control group (+0.11). Teachers' use of research evidence and change in their teaching practice made more positive impact on rural Hukou students. Table 10.15 presents the average scores of them.

Table 10.15 Comparison of standard pre, post-tests, and gain scores, rural Hukou students

	N	Mean			Standard deviation			Effect size		
		pre	post	gain	pre	post	gain	pre	post	gain
Treatment	853	-0.14	-0.12	0.02				-0.07	+0.01	+0.11
Control	829	-0.06	-0.13	-0.06						
Total	1,682	-0.10	-0.12	-0.02	1.04	1.12	0.79			

10.5.5 Impact for migrant students

A total of 520 identified migrant students had records for both final examinations. The number of migrant students was balanced between the treatment and the control groups. However, migrant students in the treatment group had a much lower average score in the pre-test (mean = -0.26) than the comparison group (mean = -0.01). Migrant students from the treatment group were facing worse situation in the pre-test, either compared to migrant students in the control group or the performance among all students.

As shown in Table 10.16, the treatment group made much more progress than the comparison group with a gain effect size of +0.20. The intervention of teachers' evidence-use may have made a crucial difference to them. Although they still gained less points in the post-test, the difference between the two groups decreased.

Table 10.16 Comparison of standard pre, post-tests, and gain scores, migrant students

	N	Mean			Standard deviation			Effect size		
		pre	post	gain	pre	post	gain	pre	post	gain
Treatment	255	-0.26	-0.23	0.04				-0.24	-0.07	+0.20
Control	265	-0.01	-0.14	-0.13						
Total	520	-0.14	-0.18	-0.05	1.04	1.24	0.83			

10.5.6 Impact on low-income students

Only 27 students were identified to receive low-income insurance. The sample size was too small to make strong conclusions about the impact for them. In contrast to rural Hukou students and migrant students, low-income students in the treatment group had a similar average score compared to all students. They kept performing much better than the control group before and after the experiment semester. The effect size of the gain scores was +0.35 (Table 10.17).

Table 10.17 Comparison of standard pre, post-tests, and gain scores, low-income students

	N	Mean			Standard deviation			Effect size		
		pre	post	gain	pre	post	gain	pre	post	gain
Treatment	12	0.02	0.14	0.12				+0.31	+0.46	+0.35
Control	15	-0.24	-0.73	-0.49						
Total	27	-0.13	-0.34	-0.22	0.83	1.89	1.75			

In brief, students with rural Hukou or migrant status appeared to benefit more from the intervention with gain effect sizes larger than the overall gain effect size (+0.07). The largest effect size was observed among low-income students whose families were registered for low-income insurance, though this effect size was less trustworthy due to the limited sample size. Besides, background information for 303 students was missing, which could impact the reliability of the subgroup analysis results.

10.6 Additional analysis for post-test only

Additional analysis for post-test only results was performed by including students who did not have pre-test records due to teacher replacement or absence for the examination but participated in the post-test and their teachers did not claim to dropout. For RCT schools, 221 students' records were missing either due to teacher adjustment or absence in the pre-test. Specifically, 94 students were missing due to absence, 22 students transferred to other schools before the pre-test, and 12 students had just transferred to their current school and did not participate in the pre-test. Test records for two classes, comprising 93 students, were lost due to their teachers either moving to different positions or being persistently absent. Out of the 221 students whose records were missing during the pre-test, 199 participated in the post-test.

The post-test only analysis was conducted to see if these students differed in any way from their peers in the other groups. The control group students who missed the pre-test obtained a much lower score in the post-test compared to those participated in the pre-test. This average score (mean = -0.47) was

also lower than the figure for students missing the pre-test in the treatment group (mean = -0.16). These control group students tended to perform worse in the pre-test. Although the number of students missing pre-test scores was small, it is still possible that the gain effect size slightly underestimated the intervention effect. Table 10.18 compared the post-test scores for students who participated in the post-test but missed the pre-test and students completed both tests.

Table 10.18 Comparison of the post-test scores of students missing pre-test

	N	post-test of missing pre-test	Standard deviation	N	post-test of non-missing pre-test	Standard deviation
Treatment	153	-0.16	1.13	3,396	0.00	1.01
Control	46	-0.47	1.36	3,508	0.00	0.98

The result of post-test only analysis is shown in Table 10.19, which kept most of the participants and maintained the randomisation. Based on 6,988 students who participated in the post-test, there was no major difference between the treatment and the control groups.

Table 10.19 Comparison of post-test only, all students

	N	z-post	Standard deviation	Effect size
Treatment	3,518	0.00	1.01	-0.01
Control	3,470	0.00	0.98	
Total	6,988	0.00	1.00	

10.7 Conclusion

This chapter reported the impact of the main experiment on students' Chinese literacy attainment. The main analysis involved a large sample size and low dropout, which increased the trustworthiness of the findings. Since treatment group students performed worse than the control group in the pre-test, a regression-to-the-mean analysis was done to explore whether the RTM may interfere with the estimated effect.

Overall, treatment students made more progress during the experiment semester with a small positive gain effect size of +0.07. No major difference was found in students' post-test performance. But the headline finding is that the intervention as a whole showed signs of promise. Treatment students at researcher-randomised schools showed better performance in the post-test and progressed more compared to the treatment group at self-randomised school. Different selected strategies may result in different effect sizes. In this experiment, students receiving embedding formative assessment showed better performance in the post-test and their Chinese ability got more

improvement. The subgroup analysis suggested that the intervention might be more beneficial for lower-middle achieving and high achieving students. Positive gain effect sizes were found among rural Hukou students and migrant students.

Chapters 11 and 12 reports on the impact on students' enjoyment of reading as non-cognitive outcome and teachers' perceptions and knowledge about research evidence use as teacher-level outcome.

Chapter 11 Impact evaluation: students' enjoyment of reading

Students' Chinese literacy scores in the final examination partially represent their native language skills, including vocabulary, reading, and writing. Chapter 10 suggested that teachers' evidence-use had the potential to increase students' Chinese attainment. However, performance in the examination does not test students' reading habits or attitudes. This chapter considers whether teachers' evidence-use and adjustment in teaching practice may make a difference to students' enjoyment of reading with the following question:

- Is there an impact of teachers' use of research evidence on student's enjoyment of reading?

The Student Enjoyment of Reading Survey was developed based on the PISA 2018 student questionnaire and PIRLS student survey. The adapted survey involved students' preference for Chinese reading materials, reading media and purpose, attitudes towards reading, and self-reported reading ability. This chapter starts with a description of students who completed the survey and reports the impact on the aforementioned four aspects respectively based on the four RCT schools. As with Chapter 10, the impact on students from the quasi-experiment school is presented in Appendix N. In addition to students' self-reported enjoyment of reading and time spent on Chinese homework, informal discussions with teachers also revealed perceived changes in students, which are included in this chapter.

11.1 Students in the student survey

The pre-test for reading enjoyment at the end of the autumn semester 2023/2024 included 2,446 students from the treatment group and 2,019 students from the control group. At the end of the spring semester of 2023/2024, 2,193 treatment students and 1,664 control students answered the survey as the post-test. In total, 3,216 students completed the two surveys, 1,886 of whom were in the treatment group. The response rate was about 54% in the post-survey. Among all respondents, 47 students were from ethnic minority backgrounds, 870 held rural Hukou, 225 were identified as migrant students, and 12 were eligible for low-income insurance. More students in the treatment group completed the surveys, which might cause bias in the ensuing results.

The major dropout for the student survey came from the self-randomised school as illustrated in Table 11.1. 94% of the students at the three researcher-randomised schools replied to the pre-test survey in the autumn semester. 87% of students responded to the reading enjoyment survey at the end of

the experimental semester. In contrast, the response rate at the self-randomised school was quite low with 23% of students completing the post-survey.

Table 11.1 Number of students in the survey, by trial settings

	Researcher-randomised schools			Self-randomised school		
	pre	post	gain	pre	post	gain
Treatment	1,806	1,741	1,659	640	452	227
Control	1,470	1,295	1,223	549	369	107
Response rate (%)	94	87	83	33	23	9
Total N			3,481			3,644

The differences may stem from variations in how the school research leads planned the experiment timeline. The three researcher-randomised schools had well-established timelines and allocated sufficient time to collect student survey data. They organised the students efficiently and facilitated the survey in a structured manner at school, ensuring both a high response rate and data quality. This arrangement also prevented parents from filling it out on behalf of their children, which presumably increased the data accuracy and consistency.

In contrast, the research lead at the self-randomised school did not allocate sufficient time or provide adequate access to the computer room for students to complete the survey in an orderly manner. Although this lead proactively communicated with researchers and shared the survey link within the staff group, only a few teachers ensured that students participated, highlighting issues with effective communication between the lead and the teachers.

While the response rate may interfere with the impact analysis, over three thousand students from both the experimental and control groups completed the survey twice. In addition, students who participated in the survey responded to all survey questions. Thus, there were no missing values within the survey data. Analysing their responses can still provide meaningful insights into the effects on students' reading enjoyment.

11.2 Impact on reading material preference

The survey mentioned six Chinese reading materials that were frequently accessible to primary school students, such as Chinese textbooks, magazines, comic books, fiction, science, and social science. Students rated these reading resources based on their preferences. Overall, there is no solid evidence

regarding the impact on students' preferences for reading materials, except for a decrease in interest in comic books and science (Table 11.2).

Fiction was the most popular reading resource among all students, achieving the highest average score across all groups and both surveys, while magazines and newspapers were the least favoured reading resource for all students. Treatment students demonstrated less interest in comic books (-0.13) and science (-0.10). No major difference was found in students' preference for Chinese textbooks, magazine, fiction, and social science.

Table 11.2 Comparison of means in reading material preference, all students

Items	Treatment			Control			Standard deviation		Effect size	
	pre	post	gain	pre	post	gain	post	gain	post	gain
Chinese textbook	6.91	6.52	-0.39	6.92	6.52	-0.40	3.04	3.30	0.00	0.00
Magazine, newspaper	6.09	5.98	-0.11	6.13	6.21	0.08	3.42	3.67	-0.07	-0.05
Comic book	6.39	6.26	-0.13	6.24	6.62	0.38	3.64	3.80	-0.10	-0.13
Fiction	7.48	7.35	-0.13	7.38	7.20	-0.18	2.93	3.33	+0.05	+0.02
Science	7.06	6.80	-0.26	7.02	7.08	0.06	3.11	3.25	-0.09	-0.10
Social science	6.98	6.73	-0.25	6.98	6.82	-0.16	3.15	3.39	-0.03	-0.03

Students' survey responses were linked to their background information. Since there were very few students from ethnic minority backgrounds or low-income families, and some data were missing for these groups, the subgroup analysis of students' survey data focused primarily on students with rural Hukou and those identified as migrant students. Consistent with the overall results, no major differences were observed among treatment students with rural Hukou status. They showed slightly less preference for science books (–0.11) but a small increase in interest in social science materials (+0.09).

More noticeable differences emerged among migrant students. Those in the treatment group reported increased interest in nearly all types of reading materials compared to the comparison group, particularly in Chinese textbooks (+0.18), comic books (+0.21), and social science (+0.20). This suggests that the teachers' use-of-research-evidence intervention may have contributed to greater engagement with a variety of reading materials among migrant students. The corresponding effect sizes are presented in Table 11.3.

Table 11.3 Effect sizes of reading material preference, by student characteristics

Items	Rural Hukou		Migrant	
	post	gain	post	gain
Chinese textbook	-0.06	+0.05	+0.13	+0.18
Magazine	+0.06	+0.03	+0.03	+0.06
Comic book	+0.13	-0.02	+0.35	+0.21
Fiction	0.00	-0.02	-0.05	-0.04
Science	-0.09	-0.11	-0.01	+0.14
Social science	-0.02	+0.09	0.00	+0.20

11.3 Impact on reading media and purposes

At primary school, students are generally provided with hard copy books and can borrow books from the school library. Some schools offer library reading lessons, designed to help students select appropriate reading materials and encourage regular borrowing. In addition, information technology enables students to access online reading resources. A primary goal for these schools is to help students find enjoyment in reading, and extract valuable information from reading materials.

The survey collected students' reported frequency of using school libraries, reading online, reading paper books, reading for information, and reading for enjoyment. It investigated how teachers' use of research evidence may influence students' reading habits, specifically their choice of reading media and reading purposes. However, students in the treatment group kept a similar reading habits of media and purpose as the comparison group. Treatment students showed small declining frequency of borrowing books from a school library compared to the control group with a gain effect size of -0.05.

The results might be explained by the use-of-research-evidence intervention being focused on evidence-informed teaching strategies in Chinese literacy lessons, which may not directly change students' wider reading media and purposes. Additionally, students' reading media and purposes may relate to their prior reading habits and available resources and therefore, were hardly changed. The comparison of the two groups is shown in Table 11.4.

Table 11.4 Comparison of means in reading media and purposes, all students

Items	Treatment			Control			Standard deviation		Effect size	
	pre	post	gain	pre	post	gain	post	gain	post	gain
borrow books from school library.	6.43	5.91	-0.52	6.65	6.30	-0.35	3.46	3.61	-0.11	-0.05

Items	Treatment			Control			Standard deviation		Effect size	
	pre	post	gain	pre	post	gain	post	gain	post	gain
read online.	3.10	3.08	-0.02	3.46	3.30	-0.16	3.52	4.00	-0.06	+0.03
read paper books.	7.91	8.07	0.16	7.94	8.06	0.12	2.89	3.40	+0.00	+0.01
read for information.	5.50	5.13	-0.37	5.56	5.28	0.33	3.42	4.30	-0.04	-0.02
read for enjoyment.	6.96	7.17	0.21	6.76	7.09	0.33	3.20	3.89	+0.02	-0.03

While the reading habits were similar between the treatment and the control groups, those with rural Hukou and migrant backgrounds in the treatment group showed increased frequency in all forms of reading and for different purposes (Table 11.5). Rural Hukou students in the treatment group read online and for information more frequently (both +0.10). Migrant students showed greater increases in reading online (+0.38), for information (+0.20), and for enjoyment (+0.10). The teachers' use of research evidence may have helped to promote more frequent reading among students from rural and migrant backgrounds, regardless of the type of reading material or the purpose of reading.

Table 11.5 Effect sizes of reading media and purposes, by student characteristics

	Rural Hukou		Migrant	
	post	gain	post	gain
borrow books from school library.	-0.19	+0.04	-0.16	+0.11
read online.	+0.01	+0.10	+0.15	+0.38
read paper books.	+0.03	+0.05	-0.03	+0.08
read for information.	+0.05	+0.10	+0.15	+0.20
read for enjoyment.	+0.12	+0.06	-0.01	+0.10

11.4 Attitudes towards reading

Treatment students expressed a slightly more positive attitude towards reading compared to those in the control group. The treatment group expressed stronger agreement on statements of 'Reading is one of my favourite hobbies' (+0.08), and 'I read for more than just information' (+0.05). While the difference between the two groups was minimal, they demonstrated a somewhat more positive attitude towards reading as part of daily life. They were more likely to view reading as something they did voluntarily, not just out of necessity, and for various purposes whether for enjoyment, entertainment, or learning. No major difference was found in other statements. Table 11.6 presents the reading attitudes of students from the treatment and the control groups.

Table 11.6 Comparison of means in reading attitudes, all students

Items	Treatment			Control			Standard deviation		Effect size	
	pre	post	gain	pre	post	gain	post	gain	post	gain
I read not only if I have to.	6.98	7.23	0.25	6.80	7.00	0.20	3.38	4.12	+0.07	+0.01
Reading is one of my favourite hobbies.	7.84	7.58	-0.26	7.84	7.34	-0.51	3.08	3.20	+0.08	+0.08
I like talking about books with other people.	7.03	6.91	-0.13	6.94	6.72	-0.22	3.21	3.66	+0.06	+0.02
I would be happy if someone gave me a book as a present.	7.88	7.48	-0.39	7.88	7.42	-0.45	2.99	3.17	+0.02	+0.02
For me, reading is a valuable use of time.	8.96	9.00	0.05	8.91	8.89	-0.02	2.46	3.19	+0.05	+0.02
I read not only for information.	7.58	7.74	0.16	7.51	7.46	-0.05	3.14	3.79	+0.09	+0.05

Treatment students with rural Hukou or migrant status may benefit more from teachers' use of research evidence by developing more positive reading attitudes (Table 11.7). Rural Hukou students in the treatment group showed stronger agreement with statements such as enjoying reading as a hobby (+0.09) and feeling happy when receiving a book as a gift (+0.14). Migrant students rated higher in enjoying discussing books with others (+0.18), appreciating books as gifts (+0.24), and viewing reading as a valuable use of time (+0.16).

Table 11.7 Effect sizes of reading attitudes, by student characteristics

Items	Rural Hukou		Migrant	
	post	gain	post	gain
I read not only if I have to.	+0.04	+0.01	-0.13	-0.04
Reading is one of my favourite hobbies.	+0.08	+0.09	-0.08	-0.03
I like talking about books with other people.	+0.04	-0.02	+0.30	+0.18
I would be happy if someone gave me a book as a present.	+0.09	+0.14	+0.22	+0.24
For me, reading is a valuable use of time.	+0.09	+0.05	+0.15	+0.16
I read not only for information.	+0.13	+0.05	+0.07	+0.02

11.5 Self-assessment of Chinese reading ability

Students' performance in the final examination may not fully represent students' Chinese reading ability. This survey also gathered information on students' self-assessed Chinese reading ability, to understand how the intervention affected their perceived reading skills using five items. Most students in this study were confident in their reading abilities. They all supposed that they could do

well in reading. They were neutral about how easy learning Chinese was compared to other subjects, with average scores around 5.

Treatment students appeared to be slightly more confident in their Chinese reading ability. They found it easier to understand texts without needing to reread, and their confidence in their reading comprehension increased during the experiment (+0.06). However, they were less confident in making correct responses to reading comprehension questions (-0.09). The comparison of the treatment and the control groups is illustrated in Table 11.8.

Table 11.8 Comparison of means in perceived Chinese reading ability, all students

Items	Treatment			Control			Standard deviation		Effect size	
	pre	post	gain	pre	post	gain	post	gain	post	gain
I usually do well in reading.	7.05	6.76	-0.29	6.88	6.64	-0.24	2.89	3.11	+0.04	-0.02
Reading is easier for me than for many of my classmates.	4.98	5.36	0.38	4.80	5.04	0.24	3.33	4.02	+0.09	+0.03
Chinese is easier for me than any other subject.	4.93	5.03	0.09	4.92	4.96	0.04	3.31	3.89	+0.02	+0.01
I do not have to read a text several times before completely understanding it.	5.98	6.21	0.23	5.88	5.88	0.01	3.35	4.03	+0.10	+0.06
I find it easy to answer questions about a text.	3.52	3.29	-0.23	3.52	3.63	0.11	3.29	3.10	-0.11	-0.09

The use-of-research-evidence intervention may enhance overall confidence in Chinese reading among students with rural Hukou and migrant backgrounds, with effect size gains of +0.14 and +0.09, respectively (Table 11.9). Treatment students with Rural Hukou were more likely to agree that Chinese was easier than other subjects (+0.14). However, the intervention did not appear to significantly improve students' confidence in answering actual reading comprehension questions, particularly among migrant students.

Table 11.9 Effect sizes of perceived Chinese reading ability, by student characteristics

Items	Rural Hukou		Migrant	
	post	gain	post	gain
I usually do well in reading.	+0.13	+0.14	+0.07	+0.09
Reading is easier for me than for many of my classmates.	-0.01	+0.01	-0.16	+0.01

Items	Rural Hukou		Migrant	
	post	gain	post	gain
Chinese is easier for me than any other subject.	+0.13	+0.14	-0.09	-0.08
I do not have to read a text several times before completely understanding it.	+0.04	+0.01	-0.10	-0.07
I find it easy to answer questions about a text.	-0.07	-0.07	-0.02	-0.14

11.6 Time for Chinese literacy homework

This study also collected the average time spent on homework of Chinese learning from a student survey, reflecting the quantity of homework and learning efficacy. 3,120 students completed the student survey before and after the experiment. While the result should be treated with caution due to the missing cases, the difference in time for Chinese homework provided an insight into students' learning burden. Compared with the control group, the increase in homework time for students in the treatment group was about two minutes less, with an effect size of -0.06 (Table 11.10).

Table 11.10 Comparison of time for Chinese literacy homework

	N	mean (min/day)			Standard deviation		effect size	
		pre	post	gain	post	gain	post	gain
Treatment group	1,836	44.71	48.32	3.51			-0.02	-0.06
Control group	1,284	42.97	48.87	5.86				
Total	3,120	44.00	48.54	4.48	31.81	36.65		

Although this difference was small, teachers' use of research evidence in their daily teaching still demonstrated the potential to enhance teaching effectiveness and reduce students' learning burden. Teachers can achieve the same teaching outcomes with a smaller amount of homework, or students can complete assignments more quickly. During the experiment, more than one teacher observed that they had not added to students' workload. Instead, their teaching had become more efficient and less stressful than before.

One teacher who incorporated formative assessment remarked.

> This term, I've given less homework than before because I understand my students better through this project, and the assignments I set are more targeted. When my colleagues left more homework to their students, I found it not necessary. And sure enough, students' test results proved. (Teacher Sy, School B)

This process might represent a positive feedback loop: by using evidence-informed teaching strategies, teachers optimised their approach, making it easier to engage students. In turn, this lightened their teaching workload, which may encourage their continuation of using research evidence.

11.7 Other changes in students perceived by teachers

Treatment group teachers found 'evidence' of change in students through interactive responses, attitudes towards learning, and test performance. During workshops and informal interviews, eight teachers provided detailed descriptions of these changes. Overall, teachers reported improvements in their students' Chinese learning enthusiasm and engagement.

Teachers Yl at School B, and Jsz and Pxh at School C, integrated peer tutoring into their daily teaching practice and observed increased student engagement. Students developed a greater reported awareness of collaboration and communication through peer tutoring activities. Even typically reserved students found it easier to participate naturally in group tasks.

> Students have developed a sense of working with their desk partners, improving not only their collaboration skills but also their communication abilities, which have noticeably progressed compared to before. (Teach Yl, School B)

> It's heartening to see that students who are usually reserved have gradually gained a sense of belonging in their groups, becoming more confident and enthusiastic participants. (Teacher Pxh, School C)

Teacher Jsz shared two specific examples from their class, noting their increased classroom participation and engagement.

> Shi, previously struggled with reading aloud, frequently mispronouncing words and pausing incorrectly. Now, he makes far fewer mistakes and is much more eager and confident when asked to read. Another student, Heng, initially failed to engage actively in group work and was criticised by the group lead. In a recent lesson involving a collaborative, role-based reading task, Heng took on a crucial role as the narrator. (Teacher Jsz, School C)

Teacher Zl at School B who integrated computer-assisted formative assessment, stated students had more opportunities to engage actively in learning activities, allowing the teacher to hear from them

more frequently. With regular, immediate feedback, students gained a clearer understanding of their own progress and were more enthusiastic during lessons.

Growth mindset and Writing about Values were not designed to directly impact students' learning outcomes, but to support their mindsets and self-affirmation. Teachers noted that students began to respond more positively to mistakes made in exams and proactively set learning goals. They observed reduced anxiety among students and a growing sense of determination. Students seemed more focused on doing their best rather than feeling discouraged or believing they could never outperform others.

> Students preferred to say, 'I'll correct them, and next time, I'll get it right', rather than 'What should I do? I made so many mistakes. How can I pass the exam?' They have a consistent attitude of believing in their effort and staying positive, which makes me feel hopeful. Even though we've been doing a lot of practice and writing exercises, the students' overall mindset has remained strong and optimistic. (Teacher Zjj, School A)

> This activity provided a valuable entry point for us as educators. I could see a spark in their eyes. That's what I find most gratifying. (Teacher Wrh, School B)

Routine homework and unit tests provided teachers with accessible evidence of students' progress, and some reported noticeable improvement, particularly among previously lower-performing students.

> The accuracy of my students' homework seems higher than before, and those who were struggling are now writing more neatly and diligently correcting their work. (Teacher Hym, School A)

> The grades of those weaker students are noticeably higher than last term. (Teacher Sy, School B)

On the other hand, Teacher Sy also pointed out a few students who remained unresponsive despite various efforts.

> There's one student who seems impervious to change, and there are many reasons behind this, of course—family background, the child's lack of self-motivation, and the limited time I can dedicate to them. (Teacher Sy, School B)

Rigorous research findings often come from large-scale trials, which reflect average effects across groups but do not guarantee that every student who receives the intervention makes progress. Some students may require more personalised or intensive support to see meaningful progress.

11.8 Conclusion

The Student Enjoyment of Reading Survey reflected Chinese reading preferences, media, purposes, attitudes, and perceived reading ability based on self-report. The data collection may have involved some subjectivity. Students who did not respond to the survey increased the risks of bias. However, most participating students completed the survey at school during the same week of the semester, which helped ensure the consistency and reliability of the data, supporting the accuracy of the analysis results.

The teachers' evidence-use intervention has the potential to encourage more positive reading attitudes and improve their confidence in understanding texts, thereby enhancing the likelihood of a long-term impact from evidence-use. The evidence on students' reading preferences is mixed, but the findings revealed that treatment students developed a stronger preference for fiction over magazines and comic books. They were more likely to find reading a favourite and meaningful hobby, not just a source of information. Based on their perceptions, they believed that reading was easier for them compared to their peers and felt confident in understanding the meaning of texts. According to observations from treatment group teachers, students appeared to be more proactive and engaged in their learning than before. These may serve as early indications of the potential long-term impact of teachers' use of research evidence on students' Chinese literacy learning.

Furthermore, this potential may demonstrate even greater advantages for students with rural Hukou or migrant students, which may enhance education equality. They showed increased preferences for various types of reading materials, and their frequency of reading across different media and for different purposes had risen. This indicates that they were making use of available resources to engage with reading. Their attitude towards reading had also become more positive.

Chapter 12 Impact evaluation: teachers' perceptions and skills in evidence-use

The use-of-research-evidence intervention provided a teacher training session regarding teachers' knowledge of research evidence and how to transfer research evidence into teaching, which can affect students indirectly. The impact of the intervention on teachers was considered a secondary outcome, focusing on their perceptions, confidence, and skills in using research evidence. This chapter focuses on these two questions:

- What impact does the use-of-research-evidence intervention have on teachers' perceptions of research evidence?
- To what extent does the use-of-research-evidence intervention make a difference to teachers' skills of using research evidence?

A teacher survey was developed to assess teachers' perceptions and skills for using research evidence (See 6.4.2). This chapter reports participant teacher characteristics and their change in attitudes, confidence, and knowledge of evidence-use. The main analysis focused on 154 teachers from the four RCT schools. Of the participating teachers, 116 completed the pre-test, while 128 took part in the post-survey. A total of 106 teachers responded to both the pre- and post-intervention surveys. Results of the quasi-experiment school are reported in Appendix N. Additionally, interviews with school leads also revealed perceived changes in teachers, which are included in this chapter.

12.1 Teacher characteristics

Overall, 154 teachers from 4 primary schools who taught Chinese literacy participated in the randomised controlled trial, with School D randomising teachers by itself. 76 of these 154 teachers were allocated to the treatment group and received the intervention during the semester. The majority of teachers in the randomised controlled trial were females and of Han ethnicity, typically holding a Bachelor's degree. Most of them had permanent employment, reflecting a stable and well-qualified teaching workforce. The proportion of each background characteristic was calculated as the number of teachers with the characteristic divided by the total number of teachers in the treatment or control group respectively.

The allocation was nearly balanced, particularly regarding sex, ethnicity, employment status, and administrative role. The percentage of teachers with a Master's degree or higher was lower in the treatment group compared to the control group. The proportion of teachers who had experience in

leading school research was higher in the treatment group, as those designated as school research leads for this project were also counted as teachers engaged in school research work. Table 12.1 illustrates teachers' background and allocation.

Table 12.1 Teacher background information and allocation

		Treatment (%)	Control (%)	N
School ID	A	7	5	9
	B	24	21	34
	C	22	18	31
	D	47	56	80
Grade	3	22	27	38
	4	29	26	42
	5	29	27	43
	6	20	21	31
Sex	female	91	94	141
	male	9	6	12
Ethnicity	Han	100	97	152
	Minority	0	1	1
	Missing	0	1	1
Highest degree qualification	Junior college diploma	3	3	4
	Bachelor	91	85	136
	Master or higher	5	15	13
	Missing	1	1	2
Permanent employment status		100	94	149
Research work		9	4	10
Administrative role		7	6	10
Total		76	78	154

The study involved teachers with varying levels of teaching experience, ranging from those with no teaching experience (0 years) to those with extensive teaching experience (up to 30 years). The distribution by year of teaching is illustrated in Figure 12.1. The horizontal axis represents the years of teaching while the vertical axis refers to the number of teachers who have taught for this length of time. The treatment teachers had slightly more teaching experience, with an average of 10.29 years compared to 8.75 years for the comparison group. The median teaching experience in the treatment group was 8 years, whereas the median for the control group was 7 years.

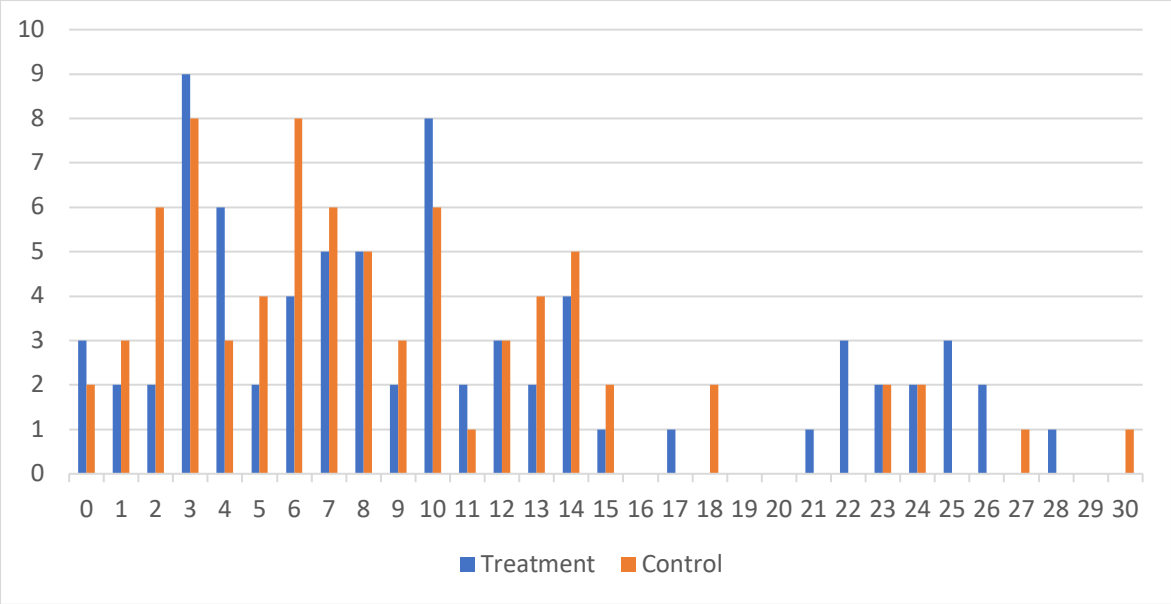


Figure 12.1 Years of teaching

Teachers reported their major in their highest degree qualification, choosing from a list that included teacher education (Chinese), teacher education (other), curriculum and pedagogy (Chinese), curriculum and pedagogy (other), general education, Chinese literacy, and applied linguistics. They could select multiple subjects and specify any unlisted majors in an open-text field.

The treatment group included more teachers with majors in Teacher Education (Chinese) than the control group (Table 12.2). The proportion of teachers with other majors, such as Chinese language and literature or general education, was relatively similar across groups. However, this comparison should be interpreted with caution, as a substantial number of teachers in the control group (32%) did not provide information about their major.

Table 12.2 Teacher major subjects for their highest degree qualification

	Treatment (%)	Control (%)
Teacher Education - Chinese	43	22
Teacher Education - Other	7	4
General education	13	10
Chinese language and literature	18	22
Curriculum and Pedagogy - Chinese	0	4
Others	11	13
Missing	16	32

12.2 Missing cases in teacher survey

The pre-test response rate was 75%, while the post-test response rate was 83%. In total, 12 teachers from the treatment group and 26 from the control group did not complete the pre-test. Non-responders were mainly from the self-randomised school (N = 36). Among the 116 who did the pre-test, 5 from each group did not respond to the post-test. Overall, 106 teachers (68.4%) completed both surveys. All teachers who completed the survey answered every question, resulting in no missing values. Due to the missing cases, the results of the impact on teachers' attitudes, confidence, and knowledge of evidence-use should be treated with some caution.

12.3 Does the intervention change teachers' views on research evidence?

The survey started with asking about the most trustworthy resource that helped teachers make decisions, followed by teacher-perceived usefulness of research evidence and facilitators of using research evidence in teaching.

12.3.1 Resources for decision making

The survey included six common sources for decision-making, including projects and materials from schools or local systems, colleagues, education figures, academic researchers, DfE guidance, and educational articles. The higher they rated an item, the more they trusted the resource. Overall, treatment teachers became more suspicious of all these sources, particularly of materials from schools and local systems and educational articles (Table 12.3). Academic researchers were not the first choice of teachers (-0.04). There was a slight increase in ratings for colleagues (+0.14) and high-profile education figures (+0.12) in the treatment group. Apart from the listed resources, teachers nominated some other approaches to help their decision making such as teaching practice and reflection, and example lesson videos.

Table 12.3 Comparison of resources for decision-making, all teachers

Items	Treatment			Control			Standard deviation		Effect size	
	pre	post	gain	pre	post	gain	post	gain	post	gain
Promising projects and materials from your school, or the local system	7.59	7.37	-0.26	7.74	7.70	-0.04	2.05	2.07	-0.16	-0.11
Colleagues e.g. informal discussion, classroom observation feedback	7.22	7.24	0.02	7.34	7.06	-0.28	2.25	2.11	+0.08	+0.14
High profile education figures e.g. authors of edu-books, education consultants	7.47	7.54	0.07	7.70	7.47	-0.23	2.08	2.42	+0.03	+0.12
Academic researchers	7.34	7.20	-0.20	7.70	7.59	-0.11	2.05	2.04	-0.19	-0.04
China Education Guidance from Department of Education	8.25	7.98	-0.26	8.21	8.11	-0.07	1.96	2.36	-0.07	-0.08
Educational news reports/ articles e.g. Teachers in primary schools in China	7.84	7.53	-0.32	8.13	8.02	-0.11	2.02	2.26	-0.24	-0.09
None of these sources	4.09	4.25	0.16	3.69	4.85	1.17	2.80	3.61	-0.22	-0.28
Other	4.07	4.81	0.73	4.08	5.51	1.43	1.98	2.95	-0.35	-0.24

Table 12.4 illustrates the outcomes for different experimental settings. Treatment teachers from the researcher-randomised schools were more judicious in their choice of resources for decisions. They became much critical about all sorts of resources, especially educational magazines, education guidance from the government, school materials, and academic researchers with gain effect sizes around -0.25. The trustworthiness for colleagues and high-profile education figures increased. However, the treatment group at the self-randomised school presented increasing trustworthiness to all resources mentioned in the survey. The gain effect sizes of colleagues, academic researchers, guidance from the DfE, and educational articles were higher than +0.30.

Table 12.4 Effect sizes of resources for decision-making, by trial settings

Items	Researcher-randomised schools		Self-randomised school	
	post	gain	post	gain
Promising projects and materials from your school, or the local system	-0.28	-0.22	+0.16	+0.12
Colleagues e.g. informal discussion, classroom observation feedback	+0.01	+0.05	+0.34	+0.34
High profile education figures e.g. authors of edu-books, education consultants	+0.02	+0.06	+0.11	+0.24
Academic researchers	-0.28	-0.20	+0.09	+0.35
China Education Guidance from Department of Education	-0.18	-0.28	+0.20	+0.41
Educational news reports/ articles e.g. Teachers in primary schools in China	-0.30	-0.32	-0.07	+0.33
None of these sources	-0.36	-0.28	+0.22	-0.22
Other	-0.52	-0.26	+0.05	-0.13

12.3.2 How do teachers find research evidence useful for improving teaching?

The statement 'I find research evidence is useful for improving teaching' was proposed in the survey for teachers' agreement. Before the intervention, treatment teachers were less likely to find research evidence useful for their teaching compared with the control group. The use-of-research-evidence intervention increased teacher-perceived usefulness of research evidence with a gain effect size of +0.26. Teachers in the treatment group hold a greater awareness of the usefulness of research evidence to improve teaching (Table 12.5).

Table 12.5 Comparison of perceived usefulness of research evidence, all teachers

	Mean			Standard deviation		Effect size	
	pre	post	gain	post	gain	post	gain
Treatment	6.56	6.85	0.29			+0.08	+0.26
Control	7.06	6.68	-0.38				

Total	6.78	6.77	-0.01	2.15	2.58		

Treatment teachers across different experimental settings consistently reported increased perceptions of the usefulness of research evidence (Table 12.6). The value placed on research evidence by treatment teachers at the self-randomised school was particularly stronger than the comparison group with a gain effect size of +0.42.

Table 12.6 Effect sizes of perceived usefulness of research evidence, by trial settings

	post	gain
Researcher-randomised schools	+0.06	+0.18
Self-randomisation school	+0.12	+0.42

12.3.3 What is required to improve evidence-use?

Teachers' needs for external support to facilitate use research evidence were categorised into practical resources (e.g., evidence-based syllabi, free access to research), collaborative learning (e.g., guidance from researchers and peers), and workload adjustments. On average, treatment teachers' requirements for external supports decreased, which indicated that the use-of-research-evidence intervention provided adequate support and guidance to facilitate teachers' use of research evidence.

Treatment teachers expressed a slightly stronger need for observing other teachers' evidence-based practice (+0.04). This increase may stem from the fact that the gap between research and practice cannot be bridged solely by extracting materials from research. Example lessons and teacher collaboration may provide a more concrete understanding of how to adapt research evidence effectively. However, they reported a much lower need for evidence-based syllabi (-0.09) and access to education research evidence (-0.15). Their requirements for workload reduction and guidance from researchers decreased during the intervention semester with gain effect sizes of -0.04 and -0.10 respectively. Table 12.7 presents external support required by treatment and comparison teachers.

Table 12.7 Comparison of external support for evidence-use, all teachers

Items	Treatment			Control			Standard deviation		Effect size	
	pre	post	gain	pre	post	gain	post	gain	post	gain
Evidence-based syllabi from schools	6.92	6.83	-0.08	7.17	7.28	0.11	1.84	2.10	-0.24	-0.09
Less managerial or teaching workload	7.17	6.90	-0.27	7.02	6.87	-0.15	2.45	2.74	+0.01	-0.04
Free access to education research evidence	7.05	6.71	-0.34	7.04	7.09	0.04	2.28	2.46	-0.17	-0.15
Training and instructions from researchers	7.46	7.03	-0.42	7.28	7.06	-0.21	2.11	2.11	-0.01	-0.10
Observation of evidence-based practice	7.34	7.15	-0.19	7.28	7.00	-0.28	1.98	2.09	+0.07	+0.04

Teachers in different trial settings demonstrated varying needs for external support, which may be related to differences in how schools planned and monitored the integration of the use-of-research-evidence intervention. Treatment teachers from the researcher-randomised schools reported much lower requirement of external support than the comparison. Workload and access to latest research evidence could hardly be an excuse for them to refuse adapting research evidence into practice with gain effect sizes lower than -0.30.

However, teachers in the self-randomised controlled trial expressed a need for various types of support, particularly for workload reduction, and this requirement increased during the intervention semester (+0.56). This increase might be because the school organised the trial largely on their own, which resulted in infrequent contact with the researcher. Their implementation is discussed in later chapters for compliance analysis and process evaluation. The effect size of requirements for external support was also reported by trial settings, as shown in Table 12.8.

Table 12.8 Effect sizes of required external support, by trial settings

Items	Effect size			
	Researcher-randomised schools		Self-randomised school	
	post	gain	post	gain
Evidence-based syllabi from schools	-0.27	-0.17	-0.14	+0.04
Less managerial or teaching workload	-0.13	-0.35	+0.34	+0.56
Free access to education research evidence	-0.24	-0.32	+0.02	+0.21
Training and instructions from researchers	-0.06	-0.27	+0.14	+0.33
Observation of evidence-based practice	+0.03	-0.12	+0.21	+0.37

In brief, this section reported teachers' perceptions of research evidence and discussed their needs for facilitating evidence-use. Treatment group teachers were more suspicious about all resources used for decision-making. They showed increasing awareness of the usefulness of research evidence during the intervention semester. The intervention provided most teachers with adequate external support to use research evidence in their teaching, leading to a reduction in their expressed needs by the end of the intervention. Variations in teachers' needs across experimental settings were largely attributable to differences in the implementation processes.

12.4 To what extent do teachers feel confident in their ability to use research evidence?

The impact of the use-of-research-evidence intervention on teachers' confidence in using research evidence was mixed. As reported in Table 12.9, treatment teachers improved their level of confidence particularly in adapting research for teaching practice (+0.12) and conducting their own research

(+0.19). Throughout the intervention semester, they had extensive opportunities to explore, practise, and reflect on their use of research evidence, which likely boosted their confidence and enthusiasm for applying research findings and conducting their inquiries. However, they were less confident than the control group in understanding the findings of the research (-0.10) and judging the quality of research (-0.20). They may have found these aspects complex and exceeding their prior knowledge. Moreover, this knowledge was not always directly applicable to teaching practice.

Table 12.9 Comparison of confidence in research evidence-use, all teachers

Items	Treatment			Control			Standard deviation		Effect size	
	pre	post	gain	pre	post	gain	post	gain	post	gain
I can understand the findings of research.	6.71	6.75	0.03	6.81	7.09	0.28	2.03	2.54	-0.17	-0.10
I can decide whether the findings of research are trustworthy.	6.80	6.69	-0.10	6.70	7.11	0.40	2.00	2.44	-0.21	-0.20
I can select appropriate research evidence to support my teaching practice.	7.03	7.07	0.03	6.91	6.87	-0.04	1.87	1.95	+0.10	+0.03
I can adapt research for my teaching practice.	6.59	6.93	0.34	6.81	6.87	0.06	2.13	2.35	+0.03	+0.12
I can conduct my own robust research.	6.27	6.58	0.31	6.38	6.19	-0.19	2.46	2.61	+0.16	+0.19

Across all trial settings, treatment teachers consistently reported greater confidence in conducting their own research. At researcher-randomised schools, treatment teachers showed a strikingly mixed pattern. While their confidence in understanding research findings (−0.17) and judging research quality (−0.35) was noticeably lower than that of the comparison group, they were markedly more confident in adapting research to their teaching (+0.18) and conducting their own research (+0.14). In contrast, teachers at the self-randomised school showed slight gains in confidence regarding their knowledge and selection of research evidence (about +0.05), though this did not extend to classroom application. The impact on the confidence of teachers who participated in different settings is illustrated in Table 12.10.

Table 12.10 Effect sizes of confidence in using research evidence, by trial settings

Items	Researcher-randomised schools		Self-randomised school	
	post	gain	post	gain
I can understand the findings of research.	-0.15	-0.17	-0.19	+0.04
I can decide whether the findings of research are trustworthy.	-0.22	-0.35	-0.13	+0.07
I can select appropriate research evidence to support my teaching practice.	+0.11	+0.01	+0.06	+0.06
I can adapt research for my teaching practice.	+0.09	+0.18	-0.13	-0.06
I can conduct my own robust research.	+0.13	+0.14	+0.24	+0.24

12.5 Does the intervention increase teachers' skills in using research evidence?

This section investigates the impact of the use-of-research-evidence intervention on teachers' knowledge and skills in using research evidence. It covers their skills to describe and access research evidence, understand research findings, evaluate research quality, and make decisions based on research evidence.

12.5.1 How do teachers describe research evidence?

The survey asked teachers to select the best description of research evidence of the three statements and rate how well each item defined research evidence based on their agreement (Table 12.11). Although the third statement is not a standard definition of research evidence, it is the best description of research evidence compared to the other two. In the post-test, both groups rated 'Findings from robust research conducted by experts in education' the highest, with average scores of 7.61 (treatment) and 7.91 (control), which suggested that most of the teachers had an appropriate understanding of what constituted strong research evidence.

Table 12.11 Means of the best description of research evidence, all teachers

Items	Treatment			Control		
	pre	post	gain	pre	post	gain
Books and articles by educators and psychologists disseminating their views.	7.34	7.15	-0.19	7.19	7.66	0.47
Knowledge from observations, information analysis, and discussions at school.	7.54	7.58	0.00	7.34	7.64	0.30
Findings from robust research conducted by experts in education.	7.49	7.61	0.12	7.66	7.91	0.26

The analysis combined the scores of three items to indicate whether teachers could figure out the best description of research evidence. The treatment group teachers demonstrated better skills in correctly describing research evidence compared to the comparison group, with a gain effect size of +0.21. The intervention possibly enhanced their sense of a correct description of research evidence (Table 12.12).

Table 12.12 Comparison of teachers correctly describing research evidence, all teachers

	Mean			Standard deviation		Effect size	
	pre	post	gain	post	gain	post	gain
Treatment	0.51	0.54	0.03			+0.01	+0.21
Control	0.70	0.53	-0.17				
Total	0.59	0.54	-0.06	0.70	0.96		

Teachers in different trial settings showed similar improvements in finding out the best description of research evidence (Table 12.13).

Table 12.13 Effect sizes of correct description of research evidence, by trial settings

	post	gain
Researcher-randomised schools	+0.10	+0.16
Self-randomised school	-0.21	+0.28

12.5.2 Access to research evidence

Six potential approaches for teachers to get research evidence were posed and rated based on their frequency of use. Table 12.14 reports that treatment teachers' reliance on general search engines for locating research evidence reduced (-0.21). Instead, treatment teachers increased their use of reports, books, and articles published by government bodies, schools, or teachers, as well as social media platforms like TED, WeChat blogs, and Zhihu. There was not much change in the frequency of searching academic databases, reading reports, books, and attending professional development conferences.

This may imply that during the experiment, teachers' attention to research evidence increased, leading them to avoid sifting through ambiguous information on general search engines. In addition to academic databases, social media may represent the most cost-effective and easily accessible options for teachers to access research information but also take the risks of non-accurate information.

Table 12.14 Comparison of frequency of accessing research evidence, all teachers

Items	Treatment			Control			Standard deviation		Effect size	
	pre	post	gain	pre	post	gain	post	gain	post	gain
Search engines e.g. Baidu	7.53	7.32	-0.20	7.51	7.79	0.28	1.89	2.24	-0.25	-0.21
Academic Databases e.g. China National Knowledge Infrastructure	7.10	7.14	0.03	7.09	7.15	0.06	2.10	2.27	0.00	-0.01
Reports, books, and articles published by the government, schools or teachers	6.97	7.20	0.24	7.02	7.09	0.06	2.03	2.42	+0.05	+0.07
Research articles or summaries published by researchers	6.92	7.00	0.08	7.15	7.26	0.11	1.99	2.32	-0.13	-0.01
Professional development conferences	6.42	6.46	0.03	6.70	6.64	-0.06	2.39	2.69	-0.08	+0.03
Social media e.g. TED, WeChat blogs, Zhihu	7.12	7.25	0.14	7.13	6.89	-0.23	2.12	2.34	+0.16	+0.16

Small variances in frequently used approaches were found in different trial settings (Table 12.15). The treatment group at the researcher-randomised schools got access to academic databases, educational reports, books, and articles more frequently than the control group, with gain effect sizes around +0.10. They also used social media for the latest research much more often than the comparison group with gain effect size of +0.23. However, treatment teachers from the self-randomised school had access to most approaches for research less often than the comparison group, except for attending professional development conferences (+0.24). The reported frequency of using other approaches decreased over the intervention semester, especially for academic databases, which showed the smallest negative effect size of -0.32.

Table 12.15 Effect sizes of approaches for research evidence, by trial settings

Items	Researcher-randomised schools		Self-randomised school	
	post	gain	post	gain
Search engines e.g. Baidu	-0.21	-0.22	-0.33	-0.25
Academic databases e.g. China National Knowledge Infrastructure	+0.11	+0.11	-0.24	-0.32
Reports, books, and articles published by the government, schools or teachers	+0.03	+0.12	+0.15	-0.10
Research articles or summaries published by researchers	-0.10	+0.07	-0.16	-0.22
Professional development conferences	-0.17	-0.08	+0.15	+0.24
Social media e.g. TED, WeChat blogs, Zhihu	+0.24	+0.23	-0.04	-0.05

12.5.3 Impact on teachers' skills in understanding research findings

Teachers' confidence in their ability to understand research findings, judge research quality, and select evidence-based teaching strategies was investigated based on their report in Section 12.4. However, the reported confidence in their ability could hardly represent their knowledge about research. This survey designed three questions with preferable answers to indicate teachers' abilities in a more subjective way in terms of understanding research findings, judging research trustworthiness, and making decisions with consideration of research evidence.

To estimate teachers' ability to understand results of research, teachers were provided with a situation where a study suggested that an intervention had a positive or negative impact of an intervention on students' reading attainment and were asked to explain its meaning. Treatment teachers scored the second item with the highest average score in both the pre-and post-test, which indicated that they agreed with the second item more. In contrast, the control group shifted their

preference from the second item to the first item as the best explanation of the impact in the post-test survey (Table 12.16).

Table 12.16 Means of understanding research findings, all teachers

Items	Treatment			Control		
	pre	post	gain	pre	post	gain
The study was of high/low quality.	7.32	5.32	-2.00	7.49	6.40	-1.09
The intervention was (in)effective in this research environment.	7.42	5.41	-2.02	7.51	6.11	-1.40
All students in the intervention group performed better/worse than those in the control.	6.83	4.92	-1.92	7.02	5.94	-1.09

The analysis combined and scored the three items into a composite variable to represent teachers' ability to interpret research findings, particularly those related to the impact of an intervention. Although treatment teachers had lower confidence in understanding research findings, their actual skills surpassed those of the control group, with a gain effect size of +0.15. The use-of-research-evidence intervention enhanced teachers' critical understanding that high-quality studies may yield small effects, and that effect sizes represented average impacts, not guaranteed benefits for every student (Table 12.17).

Table 12.17 Comparison of ability to understand research findings, all teachers

	Mean			Standard deviation		Effect size	
	pre	post	gain	post	gain	post	gain
Treatment	0.42	0.46	0.03			+0.16	+0.15
Control	0.45	0.36	-0.09				
Total	0.43	0.42	-0.02	0.65	0.79		

However, this improvement was not shown at all schools, as shown in Table 12.18. The treatment group from the researcher-randomised schools had enhanced performance in interpreting impact and their ability was enhanced with a noticeable effect size of +0.29. However, the average score of treatment teachers at the self-randomised school was lower than the control group in the post-survey and decreased during the intervention semester (-0.14).

Table 12.18 Effect sizes of skills in understanding research findings, by trial settings

	post	gain
Researcher-randomised schools	+0.28	+0.29
Self-randomised school	-0.12	-0.14

12.5.4 Impact on teachers' skills in judging research quality

Teachers were provided with simplified situations to judge whether the research approach evaluated a specific intervention robustly (Table 12.19). In the pre-test survey, the average score from both groups of teachers of the first item was higher than the second. However, in the post-test, the average score of the second item overtook the figure for the first, which suggests that items in the post-test might be more challenging to judge.

Table 12.19 Means of rating research quality, all teachers

	Items	Treatment	Control
pre	Selecting one class as a 'peer tutoring' class and compare their reading scores with other classes.	6.95	7.28
	Interviewing 40 students about their experience of peer tutoring.	6.90	7.09
post	1,023 students randomised to read extra books for 25 minutes per day, or not.	6.69	7.11
	Collecting reading scores and extracurricular reading time of 300 students for correlational analysis.	6.80	7.23

Teachers' skills in judging research quality were measured by whether they accurately identify the higher-quality research and the extent to which they recognise variations in research quality. Table 12.20 illustrates the comparison of teachers' ability to identify high-quality research. This measure reflects whether teachers can accurately identify the higher-quality research. There was no evidence that the use-of-research-evidence intervention might improve teachers' skills in judging the quality of primary research. The treatment group had a lower correct rate in the post-test survey and showed lower skills to make correct judgements of research quality compared with the control group (-0.18).

Table 12.20 Comparison of research quality identification, all teachers

	Mean			Standard deviation		Effect size	
	pre	post	gain	post	gain	post	gain
Treatment	0.27	0.19	-0.08			-0.10	-0.18
Control	0.21	0.23	0.02				
Total	0.25	0.21	-0.04	0.41	0.58		

Both the treatment and comparison groups showed negative gains, which may suggest that the post-test was more challenging than the pre-test. In the post-test, comparison teachers showed a larger absolute difference between the two options. However, since this difference was negative, it suggested that while they noticed differences in research quality, their judgments were inaccurate. In contrast, treatment teachers were more cautious in their evaluations, avoiding extreme differences

between the two items. The gain effect size for this measure was +0.07, indicating that teachers in the treatment group were slightly more proficient at perceiving research quality differences (Table 12.21).

Table 12.21 Comparison of perceived research quality difference, all teachers

	Mean			Standard deviation		Effect size	
	pre	post	gain	post	gain	post	gain
Treatment	0.05	-0.10	-0.15			+0.02	+0.07
Control	0.19	-0.13	-0.31				
Total	0.11	-0.11	-0.23	1.51	2.33		

Combining the two indicators of teachers' skills in judging research quality, while treatment teachers may not always correctly identify which study is of higher quality, they have a more refined perception of research quality differences overall. In other words, rather than making extreme or overly confident judgments, they appear to be more cautious and discerning in their evaluations.

The use of research evidence may not improve teachers' ability to judge research quality. One possible explanation is that the test items themselves contained inherent ambiguity, suggesting that modified research designs might make evaluating evidence even more challenging. Additionally, while the intervention introduced basic tools for assessing research quality, applying these skills in real-world judgment scenarios may still be difficult.

The impact on teachers' skills in identifying high-quality research and distinguishing between research quality varied across different trial settings (Table 12.22). Treatment teachers at researcher-randomised schools showed little improvement in judging research quality. At the self-randomised school, the control group outperformed the treatment group in making accurate judgments after the intervention semester (effect size = -0.50). However, treatment teachers at this school demonstrated stronger skills in perceiving differences in research quality, with a gain effect size of +0.10.

Table 12.22 Effect sizes of skills in judging research quality, by trial settings

	Identification		Perceived difference	
	post	gain	post	gain
Researcher-randomised schools	+0.20	0.00	-0.04	+0.05
Self-randomised school	-0.77	-0.50	+0.13	+0.10

12.5.5 Impact on teachers' awareness of evidence-led decision-making

At the end of the survey, a scenario was presented to capture teachers' preferences for evidence, big names, school plans, and their experiences in a conflicting teaching situation. In the pre-survey, the

scenario described a circumstance where the school provided materials claimed to be evidence-based, yet the teacher observed no improvement among students. In the post-survey, the situation involved a famous educator advising against the practice of having students collect their mistakes, despite the teacher having used this strategy for years.

In both pre-and post-test surveys, changing teaching practices without consideration was always not the best choice for all teachers. In the post-test survey, treatment teachers expressed their strong preference for evidence with higher average scores in conducting their own research (mean = 6.81) and looking again at the research evidence before making (mean = 7.02). The comparison group showed a similar preference for research evidence (mean = 6.94) and their own experience (mean = 6.91). Teachers' responses to each action are displayed in
Table 12.23.

Table 12.23 Means of rating actions for the conflict teaching situation, all teachers

	Items	Treatment	Control
pre	Stop using these materials immediately.	5.32	6.11
	Compare your use of materials with colleagues and ask for support.	7.37	6.96
	Look again at the research evidence, and make a judgement on whether to continue.	7.30	7.13
	Trust the school and carry on using the materials as instructed.	6.10	6.11
post	Trust your experience and carry on.	6.25	6.91
	Conduct your own research to decide whether to continue.	6.81	6.91
	Look again at the research evidence, and make a judgement on whether to continue.	7.02	6.94
	Trust the expert and change your teaching approach.	3.97	5.06

The study recoded the four items into an evidence-led decision-making indicator to clearly illustrate teachers' differences in action before decision-making. Teachers in the experiment group had strong recognition of evidence-led decision-making before and after the intervention. Comparing the average post-test score to the comparison group, the impact of the use-of-research-evidence intervention was notable with an effect size of +0.43. However, no improvement in their skills of deciding with research evidence was observed (Table 12.24).

Table 12.24 Comparison of evidence-led decision-making, all teachers

	Mean			Standard deviation		Effect size	
	pre	post	gain	post	gain	post	gain
Treatment	1.24	1.32	0.08			+0.43	0.00
Control	0.85	0.94	0.08				
Total	1.07	1.15	0.08	0.89	1.08		

Due to the imbalance in teachers' awareness of evidence-led decision-making at the pre-test, it is difficult to determine whether the large, positive effect size observed at the post-test can be attributed to the intervention. In this case, a linear regression model was conducted with the following result (Table 12.25).

When the teacher allocation variable was added to the model, the explanatory power increased slightly from 0.074 to 0.10. The estimated 'effect' size of teachers' awareness of evidence-led decision-making was around +0.03. This suggests that there was no strong evidence supporting the impact of the use-of-research-evidence intervention on teachers' awareness. The observed post-test effect size could be attributed to other factors such as pre-test scores and teachers' background rather than the intervention itself.

Table 12.25 Adjusted effect size of evidence-led decision-making

	Items	Standardised β	R^2
Block 1	Evidence-led decision-making (pre)	0.27	0.07
Block 2	Evidence-led decision-making (pre)	0.24	0.1
	Treatment or control	0.16	

The use-of-research-evidence intervention improved the evidence-led decision-making ability of treatment teachers at researcher-randomised schools (+0.09). They were more committed to considering all available evidence before making decisions. However, treatment teachers at the self-randomised school did not perform stronger intention to consider about research evidence before decision-making with a gain effect size of -0.15 (Table 12.26).

Table 12.26 Effect sizes of skills in evidence-led decision-making, by trial settings

	Effect size	
	post	gain
Researcher-randomised schools	+0.63	+0.09
Self-randomised school	-0.07	-0.15

To sum up, treatment teachers were more cautious about the approaches to research evidence and showed increasing interest in research findings disseminated by social media. The intervention had a positive impact on teachers' skills in describing research evidence correctly and interpreting research findings. The impact on teachers' ability to judge research quality and evidence-informed decision-making was mixed. Teachers' ability to evidence-use varied from different experiment settings. Treatment teachers from researcher-randomised schools benefitted more from the intervention to improve their evidence-use knowledge.

12.6 Change in teachers' professional engagement

According to observations from headteachers and school research leads, teachers showed growing enthusiasm for engaging in the use-of-research-evidence project. Their passion for teaching increased, along with their confidence and critical judgement in selecting the most appropriate and promising classroom strategies.

12.6.1 Views of Headteachers

Three of the four headteachers highlighted that teachers had become more energetic about improving teaching practice. In this use-research-evidence intervention, treatment teachers were required to carry out additional preparation and adjust their teaching practice. While some mid-career teachers had shown limited engagement in previous school activities, they demonstrated unexpected compliance in this project, fully meeting expectations and even guiding early-career teachers. This shift reflected a meaningful change in their professional engagement.

> I found that teachers have become more passionate and energetic. (Headteacher, School A)

> Teachers have been fully motivated. Some of the more experienced, mid-career teachers are even willing to act as role models, inspiring younger teachers to improve together. (Headteacher, School B)

> Teachers are now more engaged in their teaching and eager to showcase their work and have developed a strong sense of urgency about their professional growth. It's a particularly brave move for mid-career teachers who previously refused most school activities. But they have been actively participating in this project and provided good examples for early-career teachers. (Headteacher, School C)

The headteacher at School A shared the example of Teacher Hym, a treatment group teacher who explored Growth mindset strategies during the intervention semester. Teacher Hym was usually reserved to take on new tasks without detailed instructions and preferred to avoid unnecessary distractions. However, the headteacher noticed a clear shift in their attitude during the intervention. They believed that when teachers found the project meaningful, they were more willing to invest effort and even share their experiences with colleagues.

> Her two Growth mindset class workshops were exceptional. She creatively added new content, indicating great preparation and effort behind the scenes, which was completely different from her usual approach. I think the support and concrete examples provided may have eased her anxiety about adapting research evidence. (Headteacher, School A)

The headteacher at School D noticed a shift that teachers appeared to adopt a more rigorous and logical approach to their work, which could bring long-term benefits. Teachers were no longer approaching problems purely based on experience and intuition but increasingly turned to evidence and systematic thinking. The headteacher shared an example from a Year 6 lesson planning workshop.

> When discussing curriculum standards, teachers would reflect on questions such as, 'What are we aiming for? How should we design the lesson to achieve our goals? What kind of learning outcomes are we expecting? How should we validate these outcomes in the end?' (Headteacher, School D)

12.6.2 View of school research leads

Two school research leads observed that teachers learned new teaching techniques or enhanced the quality of their routine implementation of strategies, while also experiencing a shift in their thinking and engaging in deeper reflections on their teaching practices.

> I believe the impact goes beyond simply adding a new strategy or skill to the teachers' repertoire. It felt more like a 'qualitative' transformation, almost a metamorphosis. (Research lead 2, School B)

> Our initial aim of participating in the use-of-research-evidence project was to bring a new climate into their teaching and professional attitudes. I believe we've achieved that. (Research lead, School C)

School research leads guided teachers in using research evidence and worked as classroom teachers themselves. The school research lead at School A noted, based on their own experience, that the project helped teachers engage with teaching strategies more critically, choosing only those that were both trustworthy and practical for their classrooms.

> Before the experiment, I never used to think about which strategies were effective, and which were not. But now, I get access to all kinds of teaching strategies more rationally, focusing on doing what is right. I no longer waste time and energy on methods that lack evidence of effectiveness. (Research lead, School A)

12.7 Conclusion

The impact of the use-of-research-evidence intervention on teachers' perceptions, confidence, and skills to use research evidence was mixed, whereas school leaders observed that teachers in the treatment group became more enthusiastic about improving their teaching. Results should be interpreted carefully due to the missing cases.

The use-of-research-evidence intervention fulfilled nearly all external support needs of evidence-based syllabi, free access to research, guidance from researchers, and workload adjustments. Appropriate training on evidence-use can enhance teacher-perceived usefulness of research evidence. They grew confidence in incorporating research into their teaching and became more eager to conduct their own research. Teachers reported increased access to research findings through social media, while their use of search engines became less frequent. The intervention demonstrated potential to enhance teachers' skills in accurately describing research evidence and interpreting research findings, though there was no clear improvement in identifying high-quality research and deciding with research evidence.

Surprisingly, treatment teachers were more suspicious of all resources for decision-making. They express lower confidence in understanding research findings and judging research quality. Rather than relying on surface-level interpretations or unquestioned sources, they became more aware of the complexity of evaluating educational research. This increased awareness might have temporarily reduced their confidence, not because they became less capable, but because they became more cautious about making judgments.

Chapter 13 Process evaluation: compliance analysis

The impact analysis was correctly based on an intention-to-treat approach, which strictly followed teacher allocation, regardless of their actual participation and practice. However, the teachers' use-of-research-evidence intervention largely relied on teachers' autonomy to use and adapt the materials. The way that schools organised the experiment could directly affect teachers' participation rate and quality of use. This chapter outlines the planning and monitoring approaches used by each participating school and presents the results for the complier average causal effect. It examines factors potentially related to teacher compliance at both the school and individual levels. Additionally, a crisp set qualitative comparative analysis was conducted to identify the pathways for successful implementing evidence-informed teaching strategies.

13.1 Description of school-level planning

All participant schools organised one presentation session and two workshops during the experiment semester, in line with the requirements of the use-of-research-evidence intervention. Each school also developed a research plan that outlined its selected strategies and goals. However, the planning and management of these activities varied across schools, resulting in differing levels of teacher exposure to the use-of-research-evidence environment.

School A was a small size school and participated in both the pilot and the main studies. Through the pilot, they had already built up a collection of peer tutoring resources and gained experience in organising activities for Writing about Values. During the main experiment phase, the researcher directly worked with participant teachers, allowing for close collaboration and effective communication. Teachers at School A followed a more flexible timeline and took initiative in organising demonstration lessons and arranged workshops in response to their own progress and needs in implementing evidence-informed teaching strategies. The planning process was described by the school research lead.

> We categorised the selected teaching strategies and assigned them to different grades, so teachers know exactly which strategy they need to use. Additionally, we created an evidence-use project plan that outlines who is responsible for each strategy, collects process-related research materials, and includes relevant notes. I remind teachers at each key point to help them implement the plan as intended. (Research lead, School A)

In contrast, Schools B, C, and D scheduled the demonstration lessons and workshops after confirming the participating teachers. Schools B and C also organised reading salons, initial workshops and example lessons at the start of the intervention, offering treatment teachers a more concrete understanding of how to use research evidence in practice and set a benchmark for implementation. At the end of the month, teachers gathered to discuss challenges they encountered and to reinforce key implementation requirements, such as the intended purpose and frequency of activities.

Adjustments were made to better accommodate activities during the intervention semester. For instance, the research lead at School B modified the schedule for the Writing about Values activity based on their final exam timeline.

> At the start, I set a timeline for two writing activities and made sure to print enough writing paper and envelopes. The first activity took place right after the semester began, and the second was initially planned for May. However, we were later informed that the graduation exams would be on June 20th, so we rescheduled the second activity for June 15th. (Research lead 3, School B)

While all schools fulfilled the basic requirement of holding two workshops, Schools A, B, and C scheduled additional activities to further support teachers' use of research evidence. School A organised follow-up discussion workshops after each demonstration lesson. Schools B and C held two demonstration lessons per week, and School C also provided a workshop every two weeks throughout the intervention period. Teachers in these schools were more consistently and extensively exposed to the evidence-use environment.

However, due to the large scale of School D, it required each teacher to attend only one of the two scheduled workshops. This was a lighter requirement compared to other schools, where teachers were expected to attend both. The presentation on using research evidence at School D was held in late March, about two weeks later than in other schools. Therefore, teachers at School D had limited exposure to the intervention.

Headteachers considered teachers' workload and employed different approaches to alleviate the burden and encourage participant teachers. At Schools A and B, participation in this project was recorded as part of the teachers' involvement in school-level research, allowing the treatment group to be exempt from other projects during this semester. At School C, experimental group teachers were

not required to attend regular lesson planning workshops. School D integrated the intervention activities into regular grade-level workshops, although this approach may have risked contamination.

Leadership at Schools A, B, C demonstrated stronger preparation before the experiment began. They proactively organised meetings involving the research leads, headteacher, and researcher to establish a school-level research protocol and to gain a clear understanding of the focus of the use-of-research-evidence intervention. Headteachers and research leads maintained this level of active engagement throughout the experiment. They always participated in school visits and audited demonstration lessons to support their participant teachers. However, the headteacher at school D did not closely follow up on the project. Although the research lead was briefed on the study in advance, their role was largely limited to assisting with data collection and workshops.

As the foundation for implementing the use-of-research-evidence intervention, schools were responsible for organising and overseeing related activities. Variations in their planning may relate to the intervention fidelity. The following complier average causal effect analysis provides a closer look at this issue focusing on teachers' compliance.

13.2 To what extent does teachers' compliance make a difference to the intervention effectiveness?

The complier average causal effect (CACE) analysis was based on compliance criteria that were established before the start of the intervention, with details provided in Section 6.6.1. The use-of-research-evidence intervention required teachers to attend a presentation, participate in at least two workshops during the semester, and meet baseline usage frequency targets. The CACE analysis was not to estimate the impact of teachers' use of evidence intervention but to highlight the importance of the implementation process.

The attendance sheet indicated that no teachers from the control group participated in the intervention activities. Almost all teachers in the treatment group took part in the use of evidence presentation at the beginning of the experiment (n =73). Of the 76 teachers, 50 attended the first workshop and 38 the second, with only 26 fully participating in all required intervention activities. None of them from the self-randomised school.

The study tracked how frequently teachers used selected strategies over ten weeks to assess whether they met the suggested usage (Table 13.1). On average, teachers who used embedding formative assessment and peer tutoring implemented these strategies frequently, with average usage rates of

2.0 and 1.8 times per week. Growth mindset and Writing about Values required teachers to organise relevant activities twice during the semester. Forty-three treatment teachers met the baseline of usage frequency.

Table 13.1 Average frequency of using strategies

	Frequency (times/week)
Embedding formative assessment	2.0
Peer tutoring	1.8
Growth mindset	0.2
Writing about Values	0.2

In total, 24 teachers were labelled as compliers to the use-of-research-evidence intervention. 1,100 students were taught by teachers who completed the intervention activities and reached the baseline frequency of implementation. Using the overall standard deviation of the gain score from Table 10.6, the complier average causal effect based on full compliers was +0.22 (Table 13.2). While the headline effect size on students' gain scores was +0.07, the complier effect presented a larger effect size, which indicated that teachers' compliance could have made a difference to the effectiveness of the intervention.

Table 13.2 Complier average causal effect, full compliance

	Compliers		Non-compliers		All		Effect size
	N	Mean	N	Mean	Total N	Mean	
Treatment	1,098	0.05	2,267	0.01	3,365	0.03	+0.22
Control	1,118	-0.11	2,308	0.01	3,426	-0.03	

Compliant teachers may obtain more opportunities to have a deeper understanding of adapting research evidence into teaching. Through discussions with colleagues and school research leads during the workshops, they may have received clearer guidance and adjusted their evidence-informed practices. Or they may simply be more motivated teachers.

Since full compliance with the use-of-research-evidence intervention requires both participation in all intervention activities and meeting the baseline implementation frequency, the CACE for each partial compliance criterion is reported separately in Table 13.3 and Table 13.4. Meeting the minimum participation requirement in intervention activities led to greater student benefits, even when teachers did not fully implement the teaching strategies with required frequency (CACE = +0.20).

Table 13.3 Complier average causal effect, intervention activity

	Compliers		Non-compliers		All		Effect size
	N	Mean	N	Mean	Total N	Mean	
Treatment	1,188	0.06	2,177	0.01	3,365	0.03	+0.20
Control	1,210	-0.09	2,216	0.01	3,426	-0.03	

On the other hand, as shown in Table 13.4, using evidence-led strategies could be beneficial for students anyway but may be with minimal impact (CACE = +0.12). Basic knowledge of research evidence and the opportunity to collaborate with colleagues were important for adapting research findings into teaching and working on students' attainment.

Table 13.4 Complier average causal effect, implementation frequency baseline

	Compliers		Non-compliers		All		Effect size
	N	Mean	N	Mean	Total N	Mean	
Treatment	1,928	0.03	1,437	0.02	3,365	0.03	+0.12
Control	1,963	-0.06	1,463	0.02	3,426	-0.03	

13.3 Variables relevant to teacher compliance

Teachers' compliance with the use-of-research-evidence intervention may be shaped by both school-level factors (e.g., leadership participation and school research opportunity), individual-level factors (e.g., confidence and research-use skills), and interactive factors (e.g., demonstration lessons).

13.3.1 Leadership participation and research opportunity

Possible patterns related to school-level support and research opportunity can be cautiously observed from the data presented in Table 13.5. It is notable that no teachers at School D were overall compliant with the intervention requirements. As discussed earlier (see Section 13.1), this was largely due to the school's planning approach, which did not ensure full teacher participation in all required activities.

Schools with a stronger tradition of conducting their own research projects appeared to demonstrate higher levels of overall compliance. Teachers at School A had most opportunities to conduct and participate in research. And they showed the highest overall compliance rate. On the other hand, Schools C and D, both relatively new with weaker research backgrounds, had lower compliance rates. Schools with more experience in research might be better positioned to organise and support teacher engagement in research-related activities.

A potential relationship was also observed between the proportion of teachers meeting the implementation baseline and leadership participation. Headteachers (and/or research leads) at Schools B and C were actively involved in all workshops and classroom observations. The headteacher at School D did not attend any sessions, and the research lead participated in only one workshop, even though the school having a lighter intervention schedule. While these patterns were not conclusive, leadership participation may encourage teachers' engagement with the interventions. The role of leadership support is further explored in Chapter 14.

Table 13.5 Teacher compliance, leadership, and school research

Schools	Percentage of compliers			Leadership participation rate	School research opportunity (per teacher)
	Intervention activity	Implementation baseline	Overall		
A	80	80	80	75	0.16
B	61	94	61	100	0.09
C	65	76	53	100	0.03
D	0	26	0	50	0.02

13.3.2 Confidence and skills

Compliant teachers were more likely to show confidence in understanding research findings (+0.07) and selecting adequate research evidence for teaching (+0.19) (Table 13.6). However, they did not demonstrate increasing awareness of the potential value of research evidence for enhancing their teaching practice than their non-compliance peers (-0.26). Confidence and perceived value of research evidence alone may not be sufficient to motivate teachers to actively engage in a use-of-research-evidence intervention.

Table 13.6 Comparison of teachers' confidence and compliance

	Understand	Trustworthiness	Select	Adapt	Conduct	Usefulness
Compliance	0.08	-0.08	0.21	0.33	0.21	-0.13
Non-compliance	-0.09	-0.20	-0.17	0.26	0.29	0.46
Standard deviation	2.44	2.27	1.99	2.06	2.29	2.27
'Effect' size	+0.07	+0.05	+0.19	+0.03	-0.03	-0.26

While confidence and perceived value may not fully explain teachers' engagement, skills in using research evidence appeared to be a more distinguishing factor. As shown in Table 13.7, compliant teachers expressed stronger skills in using research evidence, particularly in understanding research findings (+0.28), identifying high-quality research (+0.45), and making evidence-informed decisions (+0.41). With these skills, teachers may have felt more capable of navigating and applying research materials to their classroom practice.

Table 13.7 Comparison of teachers' skills and compliance

	Describe	Understand	Identify	Distinguish	Decide
Compliance	0.13	0.17	0.08	-0.37	0.38
Non-compliance	-0.03	-0.06	-0.20	0.00	-0.11
Total Standard deviation	0.89	0.83	0.62	2.70	1.21
'Effect' size	+0.18	+0.28	+0.45	-0.14	+0.41

13.3.3 Support from peers

Demonstration lessons may serve as a platform for teachers to discuss challenges and share experiences in incorporating research evidence into their practice. Peer communication and collaboration played a role in supporting compliance with the use-of-research-evidence intervention. Teachers identified as compliant participated more frequently in others' demonstration lessons and delivered them more often than non-compliant teachers, with 'effect' sizes of +0.84 and +1.16 respectively (Table 13.8).

Table 13.8 Comparison of demonstration lesson opportunity and compliance

	Frequency of attending demonstration lessons	Frequency of providing demonstration lessons
Compliance	0.60	0.11
Non-compliance	0.27	0.03
Standard deviation	0.40	0.07
'Effect' size	+0.84	+1.16

Teachers' frequency of attending and delivering demonstration lessons was relevant their individual autonomy and school arrangements. At Schools A, B, and C, treatment group teachers each delivered at least one demonstration lesson during the intervention, with some voluntarily presenting multiple sessions. In contrast, School D provided fewer opportunities for teachers to showcase their lessons to researchers and colleagues, limiting their chances to receive feedback and share experiences with peers.

13.4 What are the key conditions for successful use of research evidence in teaching practice?

The investigation for success conditions involved 64 teachers from the treatment group who had complete records of attendance, frequency of using evidence-informed strategies, skills in research evidence use, and average gain score. The rest twelve teachers with missing values were excluded from this analysis. Since all 64 teachers participated in the intervention presentation, the attendance of the presentation could not serve as a distinguishing condition. Of the 64 treatment teachers, 36

were categorised as successful users of evidence-informed teaching strategies, as their classes achieved average gain scores higher than the overall average of the comparison group.

Table 13.9 illustrates the necessary condition analysis with consistency and coverage, where consistency refers to the frequency with which the condition appears in all successful cases, while coverage reflects the proportion of successful cases among all cases where the condition is present. Attendance at the second workshop and demonstration lessons, compliance with the usage baseline, and perceiving differences in research quality appeared in more than 70% of the successful cases. Each of these conditions also accounted for more than 56% of the successful cases in which they were present. Therefore, they could be potential necessary conditions for teachers to effectively use research evidence. On the other hand, understanding research findings showed low consistency (0.38) but high coverage (0.80). Although it was not consistently present in successful cases, it was often associated with success when it appeared. Therefore, this condition will be further explored in subsequent analyses.

Table 13.9 Necessary condition table

	Consistency	Coverage
First workshop	0.67	0.52
Second workshop	0.76	0.76
Compliance to usage baseline	0.71	0.68
Teacher education	0.38	0.50
Frequency of attending demonstration lessons	0.71	0.56
Frequency of providing demonstration lessons	0.62	0.59
Understanding research findings	0.38	0.80
Perceiving difference in research quality	0.95	0.63
Describing research evidence	0.62	0.54
Evidence-led decision making	0.14	0.43

The crisp-set qualitative comparative analysis was initially designed to include all cases, as they were uniformly involved in the use-of-research-evidence intervention. However, the results of the truth table were highly complex for straightforward interpretation (Appendix O). The complexity of the configurations may suggest that the implementation quality was difficult to describe. Consequently, the conditions were discussed based on different strategies to identify the simplest combination of conditions that best explain the successful use of research evidence. This adjustment does not indicate a hierarchical structure within the use-of-research-evidence intervention but identifies ‘successful conditions’ in a clearer and focused approach.

13.4.1 Success combinations for embedding formative assessment

Fourteen teachers were included in the analysis to identify combinations of factors contributing to the successful adaptation of embedding formative assessment during the experimental semester, with ten teachers classified as successful users. Variables considered were teaching experience, possession of a teacher education degree, workshop attendance, compliance with embedding formative assessment practices, attendance of demonstration lessons, and understanding research findings. Table 13.10 shows the truth table for embedding formative assessment.

Table 13.10 Truth table, Embedding formative assessment

Experience	Teacher education	Compliance		Attend lesson	Understand findings	N	Outcome
		Workshop	Strategy				
1	1	0	0	1	0	2	1
0	1	0	1	0	1	1	1
1	1	0	1	0	1	1	1
0	0	1	1	0	1	1	1
1	0	1	1	0	1	1	1
1	1	0	0	1	1	1	1
0	1	0	0	0	0	1	1
0	0	1	1	0	0	1	1
1	0	1	1	1	0	1	1
1	0	0	0	1	1	1	0
0	1	1	1	1	1	1	0
1	0	0	0	0	0	1	0
1	0	0	1	0	0	1	0
0	1	0	1	0	0	1	0
0	1	0	0	1	0	1	0
0	1	1	1	1	0	1	0

Six configurations summarised successful implementation of formative assessment. The ‘success’ pathways for early-career teachers and experienced teachers were different. The early-career teachers required more support and guidance from workshops, whereas observation of other teachers’ practice was more essential for experienced teachers. Table 13.11 illustrated the success combinations.

Table 13.11 Success combinations, embedding formative assessment

Combinations of conditions	raw coverage	unique coverage	consistency
1.Teacher education* ~Workshop* Strategy* ~Lesson * Understanding	0.2	0.2	1
2.~Teacher education* Workshop* Strategy* ~Lesson* Understanding	0.2	0.1	1
3.~Experienced* ~Teacher education* Workshop* Strategy* ~Lesson	0.2	0.1	1
4.~Experienced* Teacher education* Workshop* ~Strategy* ~Lesson* ~Understanding	0.1	0.1	1
5.Experienced* Teacher education* ~Workshop* ~Strategy* Lesson	0.3	0.3	1
6.Experienced* ~Teacher education* Workshop* Strategy* Lesson* ~Understanding	0.1	0.1	1

Configurations 1 and 2 offered pathways for teachers with limited opportunities to observe their peers' practice, regardless of their experience. These cases underscored the importance of compliant with baseline implementation frequency and skills in interpreting research findings.

For early-career teachers with limited opportunities to observe demonstration lessons, full attendance at the use-of-research-evidence workshops was necessary. Early-career teachers, without a teacher education degree and sufficient opportunities to observe others, achieved successful use by fully attending the workshops and adhering to the baseline implementation frequency of embedding formative assessment (Configuration 3). Early-career teacher with a teacher education degree could also achieve successful use through full workshop attendance, even if they implemented formative assessment less frequently (Configuration 4).

For experienced teachers, observing demonstration lessons appeared to be a key factor. Experienced teachers with a teacher education degree could successfully adapt embedding formative assessment by observing others' classrooms, even without attending workshops or strictly adhering to the strategy (Configuration 5). In contrast, configuration 6 showed experienced teachers without a teacher education degree achieved successful adaptation by attending workshops, meeting the baseline implementation frequency, and observing demonstration lessons.

13.4.2 Success combination for peer tutoring

This analysis involved 36 teachers who adapted peer tutoring as exploration of evidence-use. Teachers had a diverse demographic background from different schools, varying in teaching experience and majors. Eight variables were considered to cover all cases and summary them briefly, involving

teaching experience, possession of a teacher education degree, workshop attendance, compliance with peer tutoring requirements, attendance of demonstration lessons, and ability to understand research findings and describe research evidence. Table 13.12 illustrates the truth table for peer tutoring.

Table 13.12 Truth table, peer tutoring

Experience	Teacher education	Workshop		Compliance with strategy	Attend lesson	Understand findings	Describe research evidence	N	Outcome
		1st	2nd						
0	0	0	1	0	1	0	0	2	1
0	1	1	0	1	1	0	0	2	1
0	0	1	1	1	1	0	0	2	1
1	0	1	0	1	1	0	1	1	1
0	0	0	1	0	1	1	1	1	1
1	0	1	1	1	1	1	1	1	1
0	1	1	1	1	0	0	0	1	1
1	1	1	1	1	0	0	0	1	1
0	1	0	1	0	1	0	0	1	1
0	0	0	1	1	1	0	0	1	1
0	1	1	1	1	1	0	0	1	1
1	1	1	1	1	1	0	0	1	1
0	0	0	1	0	0	1	0	1	1
1	1	1	1	0	0	1	0	1	1
0	0	0	0	1	0	1	0	1	1
0	0	1	1	1	0	1	0	1	1
0	0	1	0	1	1	1	0	1	1
1	0	1	1	1	1	1	0	1	1
0	1	1	0	0	0	0	0	2	0
0	1	1	0	0	1	0	0	2	0
1	0	0	1	1	1	0	1	1	0
1	0	1	1	1	1	0	1	1	0

Experience	Teacher education	Workshop		Compliance with strategy	Attend lesson	Understand findings	Describe research evidence	N	Outcome
		1st	2nd						
0	1	1	1	1	1	0	1	1	0
1	1	1	1	0	1	1	1	1	0
0	1	1	0	1	0	0	0	1	0
0	0	0	0	0	1	0	0	1	0
0	0	1	0	0	1	0	0	1	0
1	0	1	0	1	1	0	0	1	0
1	1	1	0	1	1	0	0	1	0
1	0	1	1	1	1	0	0	1	0
0	0	1	0	0	1	1	0	1	0

The qualitative comparative analysis, based on the truth table, summarised twelve potential pathways that teachers effectively adapted peer tutoring in daily teaching practice, as illustrated in Table 13.13. The findings suggest that structured support, such as workshops, peer observations, and adherence to implementation requirements, plays a role in helping teachers adapt evidence-informed strategies, while skills of understanding research findings can serve as an alternative pathway to success.

Table 13.13 Success combinations, peer tutoring

Combinations of conditions			raw coverage	unique coverage	consistency
1.Teacher education* 1st* 2nd* Strategy* ~Understand* ~Describe			0.19	0.14	1
Early-career	~Understand* ~Describe* Lesson	2.~1st* 2nd* ~Strategy	0.14	0.14	1
		3.Teacher education* 1st* Strategy	0.14	0.10	1
		4.~Teacher education *2nd* Strategy	0.14	0.14	1
	~Teacher education* Understand	5.~1st* 2nd* ~Strategy* ~Lesson* ~Describe	0.05	0.05	1
		6.~1st* 2nd* ~Strategy* Lesson * Describe	0.05	0.05	1
		7.1st*~ 2nd* Strategy* Lesson* ~Describe	0.05	0.05	1
		8.~1st* ~2nd* Strategy* ~Lesson* ~Describe	0.05	0.05	1
		9.1st* 2nd* Strategy* ~Lesson* ~Describe	0.05	0.05	1
Experienced	10.Teacher education* 1st* 2nd* ~Strategy* ~Lesson* Understand* ~Describe		0.05	0.05	1
	~Teacher education* 1st* Strategy* Lesson	11.2nd *Understand	0.10	0.10	1
		12.~2nd * ~Understand* Describe	0.05	0.05	1

The first configuration, which had the largest coverage, outlined a pathway for teachers with a teacher education degree but limited skills in describing and interpreting research findings, regardless of their teaching experience. Successful use was achieved through full compliance with the intervention and implementation frequency. This finding provides substantial support for the importance of the use-of-research-evidence intervention.

For early-career teachers, the pathways to success were multiple and complex. Configurations 2 to 4 represented 60% of successful cases within this group, regardless of whether they had a teacher education degree. Attending at least one workshops and observing others' demonstration lesson was

essential for teachers who have difficulties in their abilities of understanding research findings and describing research evidence. Configurations 5 to 9 reflected pathways for early-career teachers without a teacher education degree and with limited exposure to workshops or demonstration lessons. Their success, driven by a strong understanding of research findings, underscores the importance of this skill.

For experienced teachers, the analysis suggested three pathways. Experienced teachers with a teacher education degree could be successfully users by actively engaging in all intervention activities and having a strong understanding of research findings, even with limited usage frequency, peer observation, and difficulty defining research evidence (Configuration 10). For teachers without a teacher education degree, but attending the first workshop, meeting the strategy's requirements, and observing demonstration lessons were essential for successful adaptation. They should at least be good at one key ability, including understanding research findings (Configuration 11) and correctly describing the definition of research evidence (Configuration 12).

13.4.3 Success combination for Growth mindset

Only two teachers implemented Growth mindset during the experiment semester, one of which succeeded in the adaptation, reflected by an increase in their students' Chinese literacy test result. While the two teachers had similar background of teaching experience, education degree qualification, and compliance with the intervention and implementation, the successful user had better ability in understanding research findings and using research evidence for decision-making.

13.4.4 Success combination for Writing about Values

Nine teachers adopted Writing about Values in the experiment with four successful users. All of them completed two activities of Writing about Values and the majority were experienced teachers. Also, the intervention did not require demonstration lessons for Writing about Values. It was hard to identify teachers based on their compliance with implementation, teaching experience, and attendance of demonstration lessons. Variables used in this analysis involved possession of a teacher education degree, workshop attendance, understanding research findings, distinguishing research quality, and evidence-led decision-making. The truth table was illustrated in the following Table 13.14.

Table 13.14 Truth table, Writing about Values

Teacher education	Workshop	Understand	Distinguish	Decide	N	Outcome
1	0	0	1	0	1	1
1	1	0	1	0	1	1
1	1	1	0	1	1	1
1	1	0	1	1	1	1
1	0	0	0	0	1	0
0	0	1	0	0	1	0
0	1	1	0	1	1	0
0	0	1	1	1	1	0
1	0	1	1	1	1	0

The truth table analysis suggested three combinations of success conditions, as shown in Table 13.15. All successful users were majored in teacher education. Each teacher demonstrated at least one key skill: understanding research findings, distinguishing difference in research quality, or making evidence-informed decisions.

Table 13.15 Success combinations, Writing about Values

Combinations of conditions	raw coverage	unique coverage	consistency
1.Teacher education* ~Understand* Distinguish* ~Decide	0.50	0.25	1
2.Teacher education * Workshop* ~ Understand * Distinguish	0.50	0.25	1
3.Teacher education * Workshop* Understanding * ~Distinguish* Decide	0.25	0.25	1

Configuration 1 described teachers who were less proficient in understanding research findings and making evidence-led decisions had a strong skill in distinguishing difference in research quality. Configurations 2 and 3 consisted of teachers who voluntarily attended workshops. In Configuration 2, teachers demonstrated a strong ability to judge research quality but had weaker skills in understanding research findings. In Configuration 3, teachers excelled in understanding research findings and making evidence-informed decisions but were less capable of evaluating research quality.

13.5 Conclusion

The use-of-research-evidence intervention was largely based on support from participant schools and teacher autonomy. This chapter discussed the school-level planning for the intervention, highlighting differences in how schools organised the experiment.

To assess the impact of compliance, the Complier Average Causal Effect (CACE) was estimated, reflecting the likely effect size if all teachers had adhered to the compliance criteria. The results indicated that students taught by compliant teachers made greater progress compared to those taught by non-compliant teachers (+0.22). School leadership support and research background and teachers' confidence and skills in evidence-use may all be relevant factors influencing their compliance. Interestingly, some comparison group teachers were also found using evidence-informed strategies, particularly peer tutoring and embedding formative assessment. This indicated that the experiment might risk contamination, or the evidence-led strategies were popular that teachers had already used without knowing their evidence basis.

The qualitative comparative analysis was conducted for possible paths for teachers to make a positive difference in students' Chinese scores. Experienced teachers tended to benefit from observing demonstration lessons, whereas early-career teachers required more diverse forms of support. Although no single determinant of success was identified, knowledge of research evidence was consistently crucial, particularly when teachers adapted new teaching strategies (e.g., Growth mindset, Writing about Values). Schools facing challenges in organising workshops and events should prioritise enhancing teachers' research knowledge to address the gaps created by limited support.

This chapter set the stage for the subsequent process evaluation in following three chapters. This progression allows a comprehensive understanding of key factors that facilitated teachers' evidence-use in daily practice.

Chapter 14 Process evaluation: what is the role of leadership in the use of research evidence?

Through decision-making and leadership-level planning, monitoring, and support, school leadership ensures the effective implementation of the use-of-research-evidence intervention. As reported in Chapter 13, teachers' exposure to the intervention relied on school-level arrangement. Inadequate arrangements, such as insufficient workshops or a lack of time allocation and attendance requirements for treatment group teachers, directly limited teachers' engagement with the intervention. This chapter focuses on the leadership perspective, exploring the value of using research evidence and how schools coordinate this use-of-research-evidence intervention, while considering future research initiatives. Interviews with four headteachers, and five school research leads from RCT schools were included.

14.1 Value of research evidence from school leaders' perspectives

Headteachers' decisions can affect schools' development. In this use-of-research-evidence experiment, they made the final decision on the selected strategies. When asked about the value of research evidence after the intervention, all headteachers agreed that educational research findings were closely linked to teaching quality and played a role in facilitating school decision-making. Two headteachers highlighted societal challenges, viewing research evidence as a potential means to address them.

14.1.1 Promotion of school teaching practice

Headteachers reported that they valued the use of research evidence, particularly for aligning new educational practices with existing school values and practical constraints. Understanding 'what works' and receiving reliable research evidence highlighted promising pathways to achieve their goals and helped them save time and effort.

> We previously explored deep learning, which was quite popular at the time. After reviewing the literature, we realised that while deep learning might be the goal, it needed a clear pathway. Research from both domestic and international teams led us to conclude that learning communities offered such a path, so we decided to adopt it. (Headteacher, School A)

> These examples of evidence-informed strategies were in line with the style of our school, and we are working on peer tutoring and embedding formative assessment. Since they have been

> well examined widely by research, why not directly 'borrowing as needed'. (Headteacher, School B)

> Without the support of educational research and theory, practice lacks quality. (Headteacher, School D)

14.1.2 School decision-making and management

Headteachers are responsible for a wide range of school decisions. While not all decisions directly affect students, they often involve coordinating across teachers, students, and administrative staff, and therefore require careful consideration. Headteachers consistently seek evidence to guide their decision-making and to monitor continuous progress in school development. Strong research evidence appeared convincing and was seen as a safeguard for decision-making.

> Through reading and research, we don't necessarily aim for innovation or breakthroughs at the start but prefer to build on reliable research. This allows us to move forward more steadily. Leadership decisions affect the entire school, so they must be cautious. (Headteacher, School A)

> We feel more assured using these evidence-informed resources. Numerous studies have shown their effect on student achievement, so I'm confident they will benefit our students too. (Headteacher, School B)

> Decisions must be grounded in solid foundations. The teaching strategies in this research are backed by empirical evidence and extensive practical experience. Having both theoretical support and proven results helps us make better choices. (Headteacher, School D)

The headteacher at School D also noted that the school expanded from 170 students and ten teachers in 2011 to over 6,000 students and 300 teachers by 2024. In such rapid expansion, the headteacher noted that their approach 'has always been to base our decisions on reality and evidence'. In addition to educational strategies, 'management strategies validated by research have also helped me manage the school effectively. This is what has kept the "ship" steady and allowed our school to move forward quickly.'

When schools make decisions that require teacher engagement, strong research evidence can serve as a powerful tool to support this process. For instance, School C was formed through the merger of several rural schools that were subsequently closed. Most teachers and students from those schools were transferred to the new school. Many experienced teachers preferred to maintain the status quo rather than embrace new approaches. The headteacher at School C used evidence from rigorous research to encourage and persuade teachers to change their existing practices.

> Once we know what is effective, it's a great way to convince teachers to get involved. I believe no teacher would resist seeing their students improve, and research evidence becomes a powerful tool for that. (Headteacher, School C)

14.1.3 Adaptation for societal changes and policy shifts

Schools in mainland China are under increasing pressure to adapt to society and policy changes and meet the expectations of the government, society, parents, and students. The introduction of the new Chinese National Curriculum Standards and the 'double reduction' policy has pushed schools to innovate and reform their teaching practices. Research findings serve as a valuable external resource, offering inspiration for innovation.

> …where do new ideas come from? On the one hand, from problems we encounter in our daily teaching. On the other hand, we also need external information from research, such as thematic unit teaching, which is now being promoted under the new curriculum. There's also the issue of homework design following the 'double reduction' policy. These changes push us to address both internal issues and learn from external sources. (Headteacher, School A)

Furthermore, the three-year disruption caused by the COVID-19 pandemic affected the school system in mainland China. After the pandemic, schools faced heightened expectations for teaching quality, creating a potential crisis for the teaching profession. According to the headteacher at School D, research evidence offered crucial support to help teachers meet these increased demands while also fostering professional development.

> …Teachers have struggled to meet these high expectations during this time. Professional development has become our lifeline. It's no longer a question of whether teachers are performing well but whether they can survive in this profession. Research evidence plays a crucial role in schools. (Headteacher, School D)

14.2 Barriers to the experiment management

The use-of-research-evidence intervention encouraged changes in teaching practices, demanding extra time and effort. Teachers typically followed their routines and styles and may resist change for various reasons or encounter practical difficulties at different stages of adaptation. Across all schools, a central challenge was motivating teachers to participate.

> Initially, our main concern with this project was how to manage the extra responsibilities. (Headteacher, School A)

> This project wasn't mandated by higher authorities nor applied for by the school itself, but conducted in partnership with researchers, it's hard to avoid a lack of intrinsic motivation from the teachers. (Headteacher, School B)

> Teachers were already busy with their existing workload, and some experienced teachers preferred to remain in their comfort zones. (Headteacher, School C)

> If teachers feel like they're doing this solely for the headteacher, they'll find ways to avoid it. (Headteacher, School D)

School research leads encountered a similar challenge. However, as a colleague, they avoided being overly critical of their peers and to provide feedback with careful moderation. While they did not explicitly blame their colleagues during interviews, hints of these challenges could be sensed through their comments.

> Teachers in our group were experienced and skilful… yeah… There are too many competent teachers. (Research lead, School B)

> We are just colleagues, so I have to be mindful of how I communicate. I try to be considerate of their feelings. When discussing things with the teachers, I focus more on encouragement. (Research lead, School C)

The experiment introduced additional demands that may heighten teachers' anxiety, particularly for those already feeling overwhelmed. Simply forcing teachers to participate in the project was unlikely

to motivate teachers. Instead, effective management and communication played a role in ensuring the feasibility of the use-of-research-evidence intervention.

In addition to the pressure from organising their peers, most research leads expressed concerns about their grasp of the research evidence and their ability to apply evidence-based teaching strategies. They felt uncertain about how to organise tasks and provide suitable guidance to their peers.

> Sometimes I feel like my understanding of the project is not deep enough, and I am not sure if I will be able to adequately address the teachers' questions, especially when it comes to the practical details. (Research lead, School A)

> When we first received the materials for peer tutoring, it felt unfamiliar to us. I didn't know where to start or how to guide other teachers in implementing it. (Research lead 1, School B)

> Although we often talk about formative assessment, our understanding remains quite surface-level. I'm not yet familiar enough with the strategies involved, and my communication with the teachers has also been somewhat superficial. (Research lead 2, School B)

14.3 Efforts to organise and coordinate the experiment

The experiment introduced additional demands that may heighten teachers' anxiety, particularly for those already feeling overwhelmed. Simply forcing teachers to participate in the project was unlikely to motivate teachers. Instead, effective management and communication played a role in ensuring the feasibility of the use-of-research-evidence intervention.

14.3.1 Workload management

Schools integrated the use-of-research-evidence activities considering teachers' existing workload. Schools A and B carefully chose time for use-of-research-evidence activities to minimise disruption to teachers' timetables. This approach ensured that teachers had sufficient time and opportunities to participate in the workshop without encroaching on their personal time.

> We messaged beforehand to find a time that suited everyone. (Headteacher, School A)

> We often scheduled the follow-ups during the third and fourth teaching sessions in the morning because most Chinese literacy lessons took place in the first two. Teachers did not need to adjust their existing timetables for the workshops. (Headteacher, School B)

To ensure dedicated time for the intervention, School C replaced the treatment group's regular weekly meetings with use-of-research-evidence workshops and demonstration lessons. This ensured that the intervention activities were integrated into teachers' timetables, replacing existing commitments rather than increasing their workload.

> In terms of time, we had to carve out dedicated time for these teachers to participate, exempting them from the usual weekly collective teaching activities so they could focus on the project. (Headteacher, School C)

At School D, the workshops were incorporated into the regular schedule, with both treatment and comparison group teachers attending. Treatment group teachers concentrated more on peer tutoring and formative assessment, two evidence-based strategies being selected by the school.

> By integrating this project into our regular research activities, rather than treating it as something separate, we avoided triggering teachers' defensive reactions to extra work. (Headteacher, School D)

School D's approach placed the least additional burden on teachers. However, this came with risks, such as the possibility of experiment contamination or teachers not using the provided materials as required, which could impact the study's outcomes. This was reflected in the findings from Chapter 10 on impact analysis and Chapter 13 on compliance analysis, where School D had the smallest effect size and most teachers were classified as non-compliers (mainly because of not attending enough intervention workshops).

Despite time allocation to the experiment, headteachers at Schools A and C noted that clear guidance from researchers and well-organised management of the experiment were equally important in managing teacher workload. When researchers provided detailed, actionable instructions and schools carefully planned activities, teachers felt less stressed.

> As researchers, you provided clear, actionable instructions, not vague directions, which helped teachers feel less overwhelmed. (Headteacher, School A)

> Our approach to this project wasn't just about adding extra tasks; we carefully planned the teachers' involvement, research topics, and overall management. (Headteacher, School C)

Through the pilot, School A recognised that the workload was manageable, which increased their confidence for the main experiment.

> Once we started, we realised that the workload was not as heavy as expected, and teachers were more receptive to it. Teachers often feel pressured by unfamiliar tasks, but after a semester, there was a collective sense of relief—following clear steps made it manageable. (Headteacher, School A)

14.3.2 Leadership engagement and monitoring

Headteachers indicated different forms of leadership engagement and monitoring. In the three researcher-randomised schools (A, B, and C), headteachers more frequently attended activities during the experiment semester, whereas the headteacher at School D monitored the experiment in a structured way without direct involvement. These differences in leadership engagement, as reported in 13.3, may help explain variations in teachers' compliance and impact of the use-of-research-evidence intervention.

At Schools B and C, the headteachers made a point of demonstrating a serious commitment to the experiment in front of all the teachers, underscoring the importance of participation. They actively participated in workshops and observed the teachers' practices to signal their strong commitment to the project, which motivated the teachers and provided them with the necessary support.

> If the school doesn't prioritise it, neither will the teachers, making it difficult to move forward. We had to explain the project's significance and emphasise the school's commitment. I attended every demonstration lesson and workshop to show the teachers that the school was taking this seriously. This motivated them to engage more actively. (Headteacher, School B)

> I tried to coordinate my schedule so that I could participate in all demonstration lessons and workshops. (Headteacher, School C)

The teacher 'review' system at School B also played a role in motivating teachers.

> At the end of the semester, for the project's reporting meeting, when asked if sharing their experiences could count towards their teacher review scores, I replied, 'Depends on how well you present'.

The headteacher at School C recommended offering more emotional support and using encouragement and positive feedback to motivate teachers, especially those who were middle-aged. They recognised the courage and additional time and effort that teachers made to adjust their existing teaching.

> This (change in teaching practice) can be difficult, so I focused on boosting their enthusiasm and providing positive feedback, which was more about offering emotional support. (Headteacher, School C)

School A's headteacher closely supported the participating teachers through classroom observations and provided practical feedback. For example, teachers encountered barriers when adapting Growth mindset. The headteacher attended all their demonstration lessons, offering continuous guidance and helping them revise their lesson plans. This close involvement helped ease the teachers' sense of burden and increase their willingness to participate.

The headteacher at School D mentioned their efforts to boost teachers' intrinsic motivation from their 'pain point' of professional development. They managed the project through overall management and role allocation rather than direct participation.

> The first step was forming a core research team and it's important to clarify roles and responsibilities. We emphasised instilling certain concepts through pre-project preparation, mid-process adjustments, and post-project review. (Headteacher, School D)

14.3.3 Effective communication

While direct communication between headteachers and teachers might be limited, research leads in this study took on the role of hearing teachers' need for clear and consistent guidance and communication. At School B and C, in addition to the main research lead, grade-level leads were

appointed to enhance communication and coordination within the treatment group. This may ensure that teachers can effectively collaborate with both their colleagues and the leadership.

Research leads' ability to build strong relationships with teachers helped effective communication to better adapt research evidence, fostering a more positive and active work environment. This, in turn, reduced their resistance to the additional workload. The headteacher at school A emphasised the school research lead's strong relationship with the teachers, serving as a key factor in driving the success of the use-of-research-evidence project.

> Our school research lead, Ms Z, was crucial in this. Her strong execution skills and great rapport with the teachers helped drive the project forward. Although the project increased the teachers' workload, they were willing to take it on because of their long-standing trust in the leadership. (Headteacher, School A)

Two research leads mentioned online group chat and informal in-person conversations that facilitated smooth communication with teachers. The group chat allowed participant teachers to easily access information about the intervention and related activities. Teachers voluntarily shared their teaching materials in the group chat, sparking active discussions and feedback from colleagues.

> I set up a chat group and posted all relevant information so that teachers can quickly understand what we were doing and individual responsibilities. Throughout the semester, I shared updates, materials, photos, and videos with the group. (Research lead, School A)

Informal discussion may inspire and facilitate teachers' use of research evidence in teaching practice. Even brief conversations lasting a few minutes allowed teachers to quickly address specific issues related to using evidence-informed strategies.

> I frequently gathered all our members to have a short discussion. We focused on specific issues we encountered while using peer tutoring in our recent lessons. (Research lead 1, School B)

Informal discussion between research leads and participant teachers may help to adjust the use of evidence-informed strategies when mistakes occur. By creating a comfortable environment for dialogue, teachers may feel more at ease discussing their challenges and seeking guidance. The

research lead at School C shared an example of how an informal but lengthy conversation helped motivate a disengaged teacher.

> …she didn't perform well during her first demonstration lesson. To encourage her, I told her that as a young teacher, the school is keen to provide her with better development opportunities to enhance her teaching skills. And that works! I frequently saw her seeking feedback from her peers to improve her practice. (Research lead, School C)

14.4 What may happen after the experiment?

Three headteachers outlined their plans for teachers to utilise research evidence after the intervention concluded. They either emphasised enhancing teachers' abilities to use research evidence and conduct their own studies or focused on broader implementation across all grades and subjects.

The headteacher at School C aims to cultivate a broader 'research mindset' among teachers, planning to develop their ability to apply research beyond this project. They want teachers to learn from their experiences, identify issues in their teaching, and use research evidence to address those challenges.

> Through their experiences in this project, the teachers gained insights into the scientific process of educational research. I hope they can identify research questions from their classroom experiences and find answers through research, thereby fostering a mindset of inquiry among teachers. (Headteacher, School C)

School A planned to summarise the project findings and develop its own school-level research. However, the headteacher acknowledged that teachers still tended to be passive participants, rarely considering themselves as researchers.

> I feel that while the teachers are involved in this project, they remain relatively passive, viewing it primarily from the perspective of their teaching roles. It's like a game of chess. They act as pieces on the board rather than as the players making strategic moves. (Headteacher, School A)

On the other hand, the headteacher at School B expressed a clear intention to apply the findings in a wider teaching context, encompassing all grades and subjects. They organised a presentation for all

teachers at the school, inviting the researcher to discuss the use of research evidence and for teachers to share their experiences.

> This project only involved a small group of teachers from our school. In the next semester, I hope to expand it to the entire grade level. Just like this semester, the teachers should continue their professional development activities. (Headteacher, School B)

Based on follow-ups after the intervention, School A demonstrated a stable commitment to continuing the project. They planned to solidify the use of peer tutoring as a school initiative in the upcoming semester and developed a feasible support plan. While headteachers at Schools B and C were ambitious about their school research, they were reassigned to different schools and lost oversight of the experiment. However, after the reassignment, the headteacher at School C took steps to promote evidence-informed practice at the new school by inviting the researcher to deliver a talk on the use of research evidence and establishing a school research network. The school research lead at School C asked for help from the researcher to analyse data for their school project. This outcome may further support findings in the pilot that the use of research evidence can have a lasting impact on school leadership.

Follow-ups with the school research leads revealed that only a few teachers continued to use peer tutoring and embed formative assessment practices, but their implementation was not as structured as in the experiment. The research leads at Schools B and D noted that they were no longer in charge of the project and therefore, could hardly make any important difference to teachers.

14.5 Conclusion

This chapter presented the perspectives of school leadership on the value of research evidence, particularly in guiding decision-making. All headteachers recognised the importance of research evidence in enhancing teaching quality, stabilising school management, and addressing societal and policy challenges. However, engaging teachers with research evidence and encouraging them to adapt evidence-informed strategies into daily teaching was challenging for schools with less engaged teachers.

Key factors for successfully advancing the experiment and facilitating teachers' evidence-use included managing teacher workloads, leadership involvement and oversight, and effective communication. Variations in these factors were observed across schools, which provided further evidence of the

school's low teacher compliance rate. The headteacher at the self-randomised school did not closely work with teachers on this project and placed greater emphasis on reducing teacher workload, where teachers did not meet the requirement of the intervention. Additionally, the research lead at School D acted more as a research assistant, supporting data collection for the researcher rather than actively participating and leading the study.

Three headteachers expressed intentions to continue related projects, either by expanding them school-wide or by focusing on building teachers' research skills. However, due to leadership changes and staff reassignments, only one school continued to implement follow-up activities to promote teachers' use of research evidence in a structured manner. Nonetheless, there is no direct evidence to suggest that teachers have entirely abandoned the evidence-informed teaching strategies they had been using. A longer-term follow-up would be necessary to assess the lasting impact of the intervention.

Chapter 15 Process evaluation: how is research evidence adapted into teaching?

Beyond estimating the impact of teachers' use of research evidence, it is equally important to understand how teachers investigate and incorporate research evidence and to capture variations in classroom practices. This chapter focuses on how teachers adapted a specific evidence-informed teaching strategy, emphasising the process of exploring and using research evidence at each school, as well as the quality of their implementation.

The data were collected via thirty-eight classroom observations at the four RCT schools and six observations at the quasi-experiment school (School E) during the experiment semester. An example classroom observation record can be found in Appendix P. After each observation, teachers had a short conversation with the researcher. Their thoughts and reflections on their implementation are included in this chapter.

Overall, the process of adapting research evidence into teaching involved initial implementation, challenge identification, solution proposal, practice refinement, and reflection, with collaboration and adjustment among teachers. This represents a collective, time-sequenced response observed across the teacher group. While not all teachers completed every stage, and some remained in earlier phases, the overall trajectory reflects how treatment group teachers engaged in an evolving process of incorporating research evidence to contextually appropriate implementation.

15.1 Embedding formative assessment (EFA)

The investigation of embedding formative assessment serves as an example of how teachers explore strategies they may have heard of but rarely applied, especially when such strategies require flexible integration into daily teaching. Schools B, D, and E chose to adapt formative assessment into teaching practice during the experiment semester. Fifteen classrooms were observed, involving seven teachers at School B, one teacher at School D, and six teachers at School E. In the demonstration lesson, all teachers designed at least one stage to implement formative assessment. This strategy was based on the book, *Embedded Formative Assessment* by Wiliam (2021). Five techniques were introduced, involving:

- clarifying, sharing, and understanding learning intentions and success criteria,
- engineering effective discussion, tasks, and activities that elicit evidence of learning,
- providing feedback that moves learning forward,

- activating students as learning resources for one another,
- and activating students as owners of their own learning.

15.1.1 Initial implementation: translating Wiliam's ideas into class-level goals

At the preparation and initial implementation stage, all schools began by engaging with *Embedded Formative Assessment* by Dylan Wiliam. Schools B and D provided treatment group teachers with copies of the practical book and invited them to read the full text. In contrast, teachers at School E were required to read relevant book reviews and search for related materials before the experimental semester. Teachers at Schools B and D explored all five techniques described in the book to identify those most suitable for their own teaching. They adapted the techniques flexibly according to their individual needs. In contrast, teachers at School E opted to focus on one or two of the five formative assessment techniques, selected based on the developmental characteristics of their students at each grade level.

15.1.2 Challenge identification: difference in teachers' interpretation of EFA

Formative assessment is a broad strategy that can be applied at any time during teaching and learning sessions, with five specific techniques introduced in the book. However, its broad use increased the challenges for teachers in implementation. The translation of the book was often difficult for teachers without academic background to understand, even with plenty of teaching examples.

> The book covers many strategies, but each one lacks detailed depth and specificity. (Teacher Sy, School B)

> This book is very theoretical. After reading it, I am still not sure about how to apply formative assessment in our Chinese literacy lessons. (Teacher Wc, School E)

> The language in the book is quite obscure. (Teacher Cxm, School E)

> I'm very confused about what aspects are worth assessing in teaching and how to approach them in the classroom. (Teacher Zwx, School E)

The broad guidance in the book may have contributed to the difficulties some teachers experienced in interpreting and applying embedded formative assessment, which could partly explain the superficial or inconsistent adaptation observed in some cases. Teachers flexibly selected formative

assessment activities suited to their classrooms, such as 'no-hands-up', signalling by clapping, open-ended discussions, peer assessment, and whole-class response systems. While teachers selected activities and followed the steps outlined in the book, most teachers did not provide sufficient feedback to students or make consistent adjustments to their teaching pace and methods based on evidence of students' learning progress. They tended to use small techniques to make their lessons more engaging rather than understanding better students' learning status and adjusting their teaching accordingly.

Three teachers used a 'no-hand-up' approach for answering questions and signalling completion or agreement by raising hands, which was intended to engineer effective discussion, tasks, and activities that elicit evidence of learning. While this approach engaged all students, including those seated at the back and in the corners, both teachers provided limited feedback to students who made mistakes. In the hand-raising activity, not all students participated, possibly because they were too shy, faced difficulties, or disagreed with the teachers' answer. For instance, in Teacher Wm's class, students were required to find out the key sentence of each paragraph. One student was observed to highlight different answers and did not raise their hand. However, the teacher did not provide feedback to those who did not raise their hands. In the informal discussion after the lesson, they did not mention how the activity informed adjustments to the lesson plan.

Another three teachers organised activities to promote self-directed learning and peer feedback. However, no formative feedback, reflection on student learning, or instructional adjustments based on student needs were observed during the lessons. These cases illustrated a superficial use of formative assessment, primarily energising the classroom rather than genuinely understanding student progress.

Meanwhile, some activities were not appropriately handled, which failed to achieve the intended purpose of the original embedding formative assessment activities. One teacher provided students with activity sheets with reading comprehension questions about the text. However, the questions were unable to accurately reflect students' understanding. Some students completed the tasks with ease, not necessarily due to their understanding, but by relying on supplementary materials that provided standard answers. This gave the teacher a misleading impression of student progress and the teacher could hardly make progressive feedback.

Two teachers invited students to rate their peers' performance and discuss their gains. They clarified the main objectives, setting standards, such as 'clearly state your viewpoint' and 'accurately find evidence in the text', encouraging pair discussions and peer evaluation. This format of feedback may not effectively support students well, as they were observed to be unsure about what was correct or incorrect in their work. Some students were hesitant to give constructive feedback, offering only positive comments about their peers' work. Although these teachers grasped the requirements of formative assessment, their implementation did not match the particular needs of their classes.

15.1.3 Solution proposal and practice refinement: emphasising feedback

Good innovative and creative lessons existed in classroom observation, where teachers actively attempted to understand students' learning progress, clarify learning goals, and encourage student effort. Teachers showed thoughtful use of formative assessment to inform their teaching decisions.

Such feedback could occur even in subtle moments of classroom interaction to clarify learning intentions and success criteria. Teachers may notice that some students require additional support and respond by offering quick, concise, and targeted explanations with a brief amount of time. For example, when teaching the classical Chinese text *Self-contradictory*, the teacher observed that some students were confusing results (e.g., 'the spear breaks') with conclusions (e.g., 'the shield is stronger'). This prompted the teacher to provide additional clarification and guide students toward correct answers using an activity sheet.

Similarly, Teacher Fy's follow-up questions effectively supported students struggling with the lesson's pace. The teacher noticed one group did not adapt link words as required, reminded them of the task requirements, and encouraged them to reorganise their responses with scaffolding sentences as follows.

> I think you're on the right track, but let's check the requirement
> What linking words should we use here? (Teacher Fy, School B)

Two teachers at School E provided better examples of using 'no-hand-up' strategy and signalling completion or agreement to quickly gather students' learning evidence. Unlike the more superficial applications of EFA, these teachers incorporated additional steps to reply to students who either responded incorrectly or did not raise their hands, demonstrating a deeper engagement with the formative purpose of these strategies.

Did your peer make any mistakes?
Did you correct their mistakes?
Great! Can you provide an example to your peer for the correct pronunciation of this character? (Teacher Zxj, School E)

Don't worry. Listen to my description of this story's moral once more. I believe you can answer it! (Teacher Nss, School E)

Teachers' prompt feedback would also be important when activating students as learning resources for themselves and their peers. At School E, Teacher Mm encouraged such peer-supported learning by providing students with a learning log to reflect on their progress. Students who encountered difficulties were first encouraged to seek help from their peers. If questions remained unresolved, they were invited to repeat their questions and potential responses aloud to the teacher, who then offered immediate feedback.

However, their teaching methods did not significantly reflect formative assessment techniques as outlined in the provided book. They largely maintained their original teaching style while paying extra attention to students. In the observation, two teachers at School B demonstrated formative assessment activities more distinctly and innovatively.

One teacher collected students' preparatory homework, which invited students to raise three most-caring questions about the text, *Brush Master Li*. The class started with a discussion with questions drawn directly from students' homework. The teacher asked them to read through the text to find answers, which ensured that the discussion addressed the core issues they encountered. Here are example questions raised by students.

What is Brush Master Li's surname?
Why is he called Brush Master Li?
What is his background? (Teacher Sy, School B)

The teacher also organised an activity to communicate quality by using at least two pieces of anonymous work so that students could clearly understand high-standard character descriptions. The

teacher illustrated two anonymous works and prompted students to discuss their strengths and possible improvements to help students grasp, as shown in Figure 15.1.



Figure 15.1 Discussion of quality by using two pieces of anonymous work

Teacher Sy reflected on the experience of embedding formative assessment, realising that they had shifted the focus from strictly following the textbook to paying greater attention to students' learning states.

> In the past, I was more concerned with my lesson planning and content. Now, I step back more often to observe the students' overall engagement. I've learned to not only think about what I will teach beforehand but also maintain an ongoing awareness of students' learning progress. (Teacher Sy, School B)

While in the traditional teaching environment, teachers found it difficult to know all students' learning status, Teacher Zl illustrated the possibility by integrating a computer-assisted all-student response system (Figure 15.2). The teacher used multiple-choice questions to check understanding in real time. Each student used a tablet to respond through the software, and the classroom screen displayed the number of responses and the distribution across options. For questions with a high error rate, the teacher encouraged students to consult their books and try again.

We frequently employ this approach to assist students with revision and preparation for the final examination. I feel that I have gained a better understanding of my students and can provide more effective support. (Teacher Zl, School B)



Figure 15.2 Example of computer-assisted all-student response system

15.1.4 Reflection: better integrating formative assessment into teaching

Teacher discussions provided valuable opportunities to share experiences and address challenges. In a workshop, teachers recommended activities that had worked well in their classrooms and could inspire others' adaptation, such as such as 'question shells' to move the question from what to why (Teacher Ll), student reporter (Teacher Cxr), minus-equals-plus (Teacher Wzp), test-item design (Teacher Wzp), learning portfolios (Teacher Nss), and no-grading but comments (Teacher Pcj). They also discussed how the same technique, such as the 'no-hand-up' strategy, could be implemented differently and lead to varied outcomes.

When I use the no-hand-raising random questioning strategy in my class, I've encountered that some students rely on luck, thinking they won't actually be chosen. (Teacher Wzp, School E)

I use popsicle sticks to randomly select students, but that's just from the students' perspective. Sometimes, I observe their expressions and 'randomly' pick the student I actually want to question. (Teacher Nss, School E)

15.1.5 Conclusion

Experiment teachers explored diverse techniques and activities based on the book by Dylan Wiliam to collect solid evidence of students' learning progress and provide appropriate support. However, teachers interpreted embedding formative assessment differently. Some teachers used these activities mainly to engage students, with limited focus on feedback, while others intended to give feedback but chose activities poorly suited for that purpose. During the investigation process, peer discussion may inspire teachers' incorporation of formative assessment. A few teachers demonstrated creative use of EFA by restructuring lesson formats or incorporating computer-assisted tools to meet students' needs, offering valuable examples for future practice.

15.2 Peer tutoring

Peer tutoring is widely adopted in Chinese teaching contexts, which is strongly recognised by teachers as an indicator of student-centred teaching. The process serves as a clear example of how teachers use research evidence to structure and enhance existing classroom practices with three cycles of investigation.

> With the new curriculum reforms, the teaching content has become more integrated, practical, contextual, and open, raising the demands for collaborative learning, and peer tutoring is an important form of such collaboration. (Teacher Wtr, School C)

Forty-four teachers chose to investigate the peer tutoring. The supporting document suggested that peer tutoring should be implemented at least once a week, lasting 10 minutes. If teachers expect a visible change in students' achievement, the frequency should be 3 times a week, at a minimum of 15 minutes for 10 weeks. Across all classroom observations for the treatment group, teachers spent an average of approximately eleven minutes on peer tutoring, which indicated that they met the baseline of peer tutoring implementation. However, only five demonstration lessons extended peer tutoring to 15 minutes or more.

The teacher at School A displayed three demonstration lessons and dedicated the longest average time, spending 14.7 minutes per session. Schools B and C showed similar averages of around 11 minutes each. In contrast, School D, which only provided one demonstration lesson from a treatment teacher, spent just 5 minutes on peer tutoring, a notably shorter duration. Additionally, a teacher in the comparison group at School D allocated about 9 minutes to peer tutoring.

15.2.1 First cycle: students' ability and rules of peer tutoring

Following guidelines of supplementary documents from the use-of-research-evidence intervention, teachers paired students, considering their prior Chinese literacy performance and emotional traits. Peer tutoring had mostly been used for demonstration lessons rather than daily teaching, and teachers were sceptical about students' skills to stay focused and support each other effectively. Some students frequently went off-task, chatting or behaving playfully instead.

> Some students disengage during group activities, believing that they can rely on their peers for support without putting in their own cognitive effort. (Teacher Yl, School B)

> Most teachers view group collaboration as merely a formality, time-consuming, inefficient, and difficult to implement. While we may accept the idea of peer tutoring in theory, we have not truly implemented it in practice. (Teacher Wtr, School C)

> We don't trust students' cognitive and comprehension levels enough. We stick to a model where the teacher talks and students listen; the teacher asks, and students answer. (Teacher Jsz, School C)

Teachers emphasised the rules of peer tutoring to cultivate students' habits and awareness of cooperative learning. While the template in the supporting document (D.2.7 in Appendix D) was a detailed and extensive form, teachers condensed it into a simpler format. Teachers usually provided students with a form of standard and required students to rate themselves and their peers for their performance in peer tutoring. Examples are outlined as follows.

> For peer tutoring tasks:
> Listen: Watch and listen with care, fix mistakes then and there.
> Speak: Speak up strong, don't mumble or prolong.
> Evaluate: Praise what's great, and help your peers elevate.
> For group discussion:
> Clarify tasks, think actively, and follow the order.
> Avoid interrupting, don't cut in, and contribute thoughtfully.
> Summarise ideas, represent the group, and express confidently. (Teacher Tc, School A)

> In discussions, my partner clearly expressed their ideas.

When our views differed, they offered helpful suggestions.
We made progress together and shared enjoyable moments. (Teacher Zc, School C)

Reads accurately, fluently, and with expression, at a moderate pace, conveying the softness of words like ‘slowly’, ‘gently’, and ‘softly’.
Reads accurately and fluently, generally able to convey the softness of words like ‘slowly’, ‘gently’, and ‘softly’.
Reads mostly accurately, with occasional added or missed words, slightly faster or with dragging tones, and some issues with phrasing and pauses. (Teacher Pxh, School C)

Teacher Tc printed out these guidelines and invited students to paste them on the first page of their Chinese textbooks (Figure 15.3). Most students demonstrated an awareness of peer tutoring, actively helping their peers.

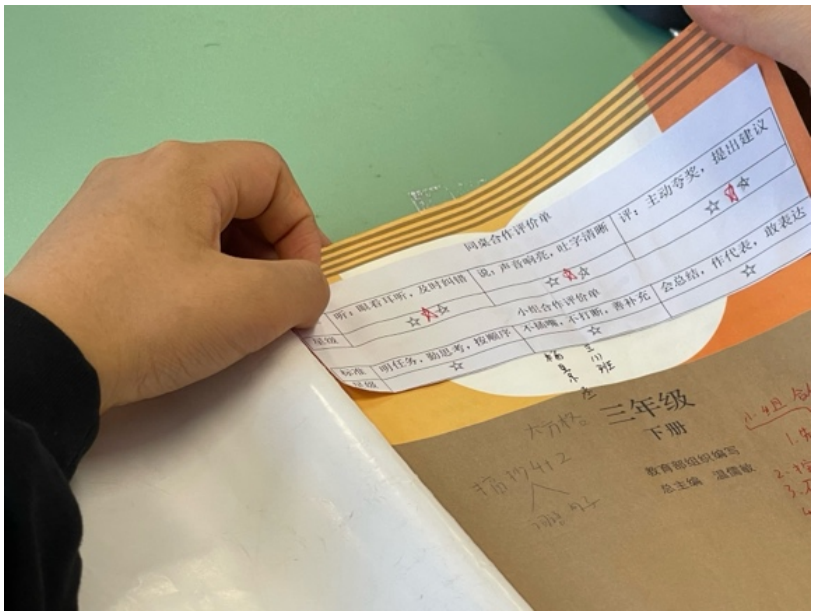


Figure 15.3 Amended rules for peer tutoring, School A

Other teachers used verbal orders and feedback as a way to establish guidelines, encouraging students to notice the peer tutoring activities. Examples are outlined as follows.

[Before peer tutoring] What order should our group discussion follow? And how loud should we speak?
[After peer tutoring] How is peer reading with your partner different from reading the text alone? (Teacher Zm, School B)

> Students sitting on the left read aloud first and students on the right followed. (Teacher Yjj, School C)

> When listening to your partner read, did you notice where they read well or made mistakes? (Teacher Yl, School B)

> Follow the sequence of 1-2-3-4. This time, I would like to invite the student No.3 in each group to present. (Teacher Cy, School C)

> If you think there's an issue with your partner's answers, calmly and reasonably persuade them. (Teacher Jd, School C)

> How would you evaluate your groupmates' performance? (Teacher Sh, School D)

To better allocate students to different roles during the group discussion, Teacher Jsz at School C placed a folded card on each group's table (Figure 15.4). The card listed specific roles, including recorder, leader, reporter, and observer, which rotated regularly. Teacher Jsz found that this approach fostered a stronger sense of unity among students during peer tutoring activities.

> Everyone has a task, with three or four students forming a team. I could see a spark in their eyes, as if they were completing a project together, coordinating and cooperating. (Teacher Jsz, School C)



Figure 15.4 Group role assignment card, School C

When teachers and students are accustomed to peer tutoring practices, the majority of students followed instructions and completed peer tutoring tasks effectively. They would naturally assign roles among themselves, with one student often taking the lead in organising and maintaining order, as illustrated in examples:

> There are 21 paragraphs in this text, and each of us read 7 paragraphs. (Class of Teacher Zm, School B)

> I'll assign tasks. [Name1], [Name2], you handle organising the third section, and [Name3] and I will organise the fourth. (Class of Teacher Xh, School B)

> [Name1], please hold on! [Name2] hasn't spoken yet. You should go after him. (Class of Teacher Yl, School B)

During this cycle, teachers realised that they had previously talked and guided too much in the classroom, rather than allowing students the opportunity to work and complete tasks with their peers. The exploration of peer tutoring gave them the insight to 'give back the class to students'.

> I found my students to be reliable and capable of completing tasks with their peers. I felt more relaxed during my teaching. (Teacher Wtr, School C)

15.2.2 Second cycle: looking for the most suitable peer tutoring activity

The use-of-research-evidence project listed eight example activities to support teachers in organising peer tutoring. The activities included partner reading, paragraph 'shrinking', prediction of the content/ advancing questions, activating background knowledge, reading comprehension strategies, group goals, peer feedback, and after-class paired reading. Teachers could choose any of them for their teaching sessions and even create their own peer tutoring activities to include more detailed instructions and ensure feasibility within the limited class time.

Partner reading was the most popular activities, with nearly all teachers incorporating them into their peer tutoring demonstration lessons. It appeared to be a warm-up activity for the participant teachers. The supporting document recommended that students read the same paragraph to each other, with the stronger student reading first and the weaker student following. In practice, teachers generally asked students to read through the entire text, which usually took five to ten minutes. Rather than reading the same paragraph in turn with their partners, students in most demonstration lessons tended to read by paragraph sections. Some teachers also highlighted key words and sentences on the slides as peer reading tasks.

Prediction of the following content would be another wide used activity, which would be particular suitable for the text was an excerpt from the original novel. For instance, when reading *The little mermaid*, Teacher Yl at School B asked students to predict what would happen after the mermaid reached the shore and discuss their thoughts with peers. In one observed group, a student described with simple phrases, 'golden sand and rocks along the coast', to which their peer added more detailed imagery, 'The setting sun casts a golden hue over the sand, and as the sun dips lower, the ocean surface turns orange. Waves crash against the rocks…'. This kind of interaction encouraged active participation from each student and allowed peers to build on each other's ideas, creating richer, more nuanced descriptions.

Similarly, Teacher Zm at School B used prediction techniques with *The secret of the magic gourd* since. The teacher encouraged students to anticipate the next part of the story by introducing a 'story ladder' (Figure 15.5). In their groups, students filled in the story background, including the protagonist's goal, the supernatural events in the story, and the reasons why the protagonist couldn't achieve their goal, before predicting the resolution and possible outcomes.

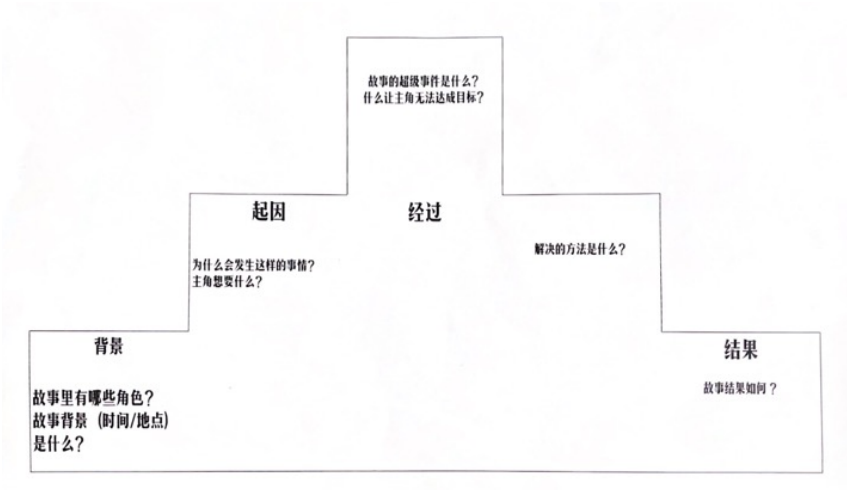


Figure 15.5 Story ladder

Predicting the next part of the story may not be applicable to non-fiction texts. Depending on the type of text, teachers asked comprehensive and thoughtful questions to prompt group discussions. Teachers generally set up peer discussion before they nominated students to present the answer that required deeper thinking and investigation. Examples included:

> What do you think of the student-representative rotation system? Should it be implemented? (Teacher Tc, School A, Communication and conversation)

> In groups, identify the flowers mentioned in the text, their blooming times, and postures, then represent these in drawings. (Teacher Tc, School A, Scientific essay: *Flower clock*)

> Based on the translation of this classical text, recount the story of Che Yin's studies by glowworm light. (Teacher Lcf, School B, Classical Chinese: Study by the light of glowworms)

> Discuss with peers: Which view do you favour more, the pond or the river? (Teacher Yjj, School C, Poem: *Pond and river*)

> Discuss your understanding of the phrase 'everything seems to be alive' based on your life experiences within your group, and come to a collective viewpoint. (Teacher Gyw, School C, Prose: *Our wonderful world*)

> Imagine yourselves as photographers. Highlight key scenery in paragraphs 3 to 6 and then describe what you would capture within your groups. (Teacher Smj, School C, Prose: Venice's Gondolas)

> Read the fifth paragraph silently, circle keywords, consider the features of the gaps, and discuss with your partner. (Teacher Sh, School D, Narrative essay: *Jinhua's Shuanglong Cave*)

Additionally, Teacher Lxl at School C introduced the reading comprehension strategies and organised group discussions for analysing the text, *A nation of ranches*. At the start of the lesson, the teacher introduced the process of 'reading', 'highlighting', 'thinking', 'understanding', and 're-reading' to explore how the author portrays the 'true Netherland'. After guiding students through the second paragraph, the teacher arranged group work for paragraphs 3 and 4, with two groups presenting their findings to the class.

> A clear example before peer tutoring activities made the process more straightforward and helped students better understand their roles and responsibilities. (Teacher Lxl, School C)

However, it is impossible to implement a peer tutoring activity that fits all classroom contexts. While teachers were encouraged to explore a variety of activities, some were poorly suited to their classes or not well organised. For instance, Teacher Llj at School C revised the teachers' observation form provided by the intervention (D.2.8 in Appendix D) to track group performance but ran out of time to observe all groups and forgot to use the form during the demonstration lesson. Afterwards, they simply ticked the form, indicating that all students had done well.

Inadequate adaptation of research evidence appeared especially common among early-career teachers, as high-quality use often depended on having strong foundational teaching skills. For example, Teacher Tc at School A and Teacher Wzq at School C, both with around three to four years of teaching experience, asked students to write a short essay in the last ten minutes of class, intending for them to exchange and give peer feedback. While the activity aimed to foster peer feedback, few students managed to complete their essays with limited time, let alone give feedback to their peers.

Additionally, some teachers treated peer tutoring merely as a step before announcing the correct answers, which decreased the necessity of peer tutoring activities. Students could simply rely on reference books to find the correct answers for peer tutoring rather than engaging in critical thinking

or exploration. Teacher Jd at School C sharply pointed out the potential risk of a new form of tokenism during the workshop.

> We need to integrate evidence-informed strategies into our teaching, rather than applying them as a compulsory task without thinking. (Teacher Jd, School C)

These suboptimal implementations were revealed during demonstration lessons and subsequently discussed in workshops by both the demonstrating teacher and observers, enabling all participants to learn from the experience. Teachers gained practical insights and reflected on how to better organise peer tutoring activities. For those that proved less suitable in Chinese literacy lessons, they also proposed improvement plans.

For example, Teacher Llj at School C reflected on the difficulty of using a classroom observation form designed for teachers. This challenge may be linked to the large class sizes in China. Even with four students per peer tutoring group, there could be up to ten groups in a single class, making it hard for teachers to monitor each group closely. One teacher added a potential solution by having a list of focus students. However, no other teacher further adopted this approach.

The early-career Teacher Tc at School A as mentioned above received direct help from the headteacher after the demonstration lesson. The headteacher noticed that the teacher did not pose questions accurately, which made students confused about the peer tutoring task. Here are questions posed by the teacher:

> What do you think are the most marvellous things in the sky?
> Why are they marvellous, and why are they full of life?

The teacher had intended for students to analyse descriptive words and phrases about the sky's features, while most students struggled to summarise answers systematically during peer tutoring and instead resorted to reading long passages from the text. The headteacher suggested making the question clearer, such as, 'What elements of the sky does the author describe, and what words are used to depict their characteristics?'

Teacher Wzq at School C received support from both peer teachers in the same grade and the school research lead. During the workshop, the teacher recognised that organising peer writing activities in

class was not appropriate in that context. Some teachers commented that, Teacher Wzq continued to pose questions expecting a 'preferred' answer, merely inserting a peer tutoring session before students responded. In response to this rather critical feedback, the research lead noticed that Teacher Wzq appeared demotivated and upset, and therefore offered words of encouragement after the session. Encouragingly, Teacher Wzq then voluntarily proposed to conduct another demonstration lesson. While certain issues remained, there was clear progress in delivering peer tutoring activities. The teacher also demonstrated a more positive attitude towards their teaching and professional responsibilities.

These cases indicated that engaging with research evidence, especially for early-career teachers, may contribute to the development of teachers' professional competence. In addition, teachers also learned from observing others' cases and reflecting on their experiences. For instance, Teachers Lhm and Cy at School C learned from the failure of peer writing activity and simplified the writing task to a fill-in-the-blanks activity. Students only needed to write a few sentences before moving on to peer feedback and discussion, which was more feasible within class time.

15.2.3 Third cycle: embedded peer tutoring

Most treatment teachers only completed the first two cycles of exploring peer tutoring, mainly focusing on finding suitable activities for their classrooms. However, a few experienced teachers showed how peer tutoring could be flexibly integrated into teaching and classroom management. They recognised that the core idea of peer tutoring was to return learning autonomy to the students. While they did not emphasise peer feedback or assign specific responsibilities for peer tutoring, the classroom was more student-driven, with students actively supporting each other to complete tasks, taking Teacher Yl at School B as an example. This teacher, with nine years of teaching experience, demonstrated well-developed skills through fluent delivery and well-organised classroom interactions.

Like many other teachers, this teacher initially organised peer tutoring activities using a set of rules and regulations, assigning clear roles and responsibilities to students. While the implementation was well-structured, the teacher later reflected that it felt 'forced' and kept teacher centred.

> …these activities still felt somewhat forced and remained teacher-centred, with students not truly constructing their own understanding. (Teacher Yl, School B)

Based on the teachers' reflections and continued exploration, the third demonstration lesson exhibited a more natural and spontaneous form of peer tutoring among students. In this lesson, the teacher did not emphasise peer feedback or assign specific responsibilities for peer tutoring. Instead, the classroom was more student-driven, with students actively supporting each other to complete tasks.

This teacher was not the only one who could flexibly integrate peer tutoring into classroom teaching. Their development was traceable because they voluntarily delivered three demonstration lessons during the intervention semester. Some teachers also embedded peer tutoring into extra-curriculum activities and classroom management with group goals. They promoted students' sense of group cooperation through group competitions and a points-based system, which helped sustain motivation and engagement in peer tutoring. For example:

> Organising group projects where peers collaborate to research Chinese traditional festivals, perform fables, and share stories while gathering information and creating posters. (Teacher Pxh, School C)

> Awarding a 'no-homework pass' to the highest-scoring peer tutoring group each week based on their group performance. (Teacher Yl, School B)

> Awarding points for homework completion: two points if all group members complete homework correctly, and one point if corrections are made promptly and rechecked. Groups with weaker students can earn double points for noticeable progress. (Teacher Zc, School C)

15.2.4 Conclusion

In the demonstration lessons, most teachers met the minimum implementation requirement of ten minutes, with some exceeding this duration. Three cycles of exploration were observed. Teachers encountered difficulties in establishing peer tutoring rules, identifying suitable activities, and embedding peer tutoring into classroom practice. Through peer discussions, they collaboratively reflected on these issues and develop practical solutions. The majority of teachers focused on the initial two, establishing rules for peer tutoring and identifying appropriate activities. A smaller number of teachers engaged in deeper exploration, extending the use of peer tutoring beyond Chinese literacy lessons to extracurricular activities and classroom management. While teachers initially followed the

supporting document, they found that not all activities were suitable for the Chinese literacy lessons, which required teachers to refine their approaches through practical experience.

The self-randomised school only organised one peer tutoring demonstration lesson, which could hardly represent a school's general implementation of peer tutoring. According to the observation, the teacher seemed to treat peer tutoring as a procedural step and did not meet the baseline time requirement for peer tutoring. If this case is representative of the school's implementation quality, it may partly explain the relatively smaller impact on students compared to other schools.

15.3 Growth mindset

Growth mindset was introduced as an entirely new strategy for teachers, serving as an example of adapting research evidence to enhance practice. Two teachers at School A employed the Growth mindset idea during the experiment semester. None of them had prior knowledge of growth mindset, let alone carried out related activities in the classroom. Although the participating teachers found it challenging to explore the use of the growth mindset, they did not express any resistance to this practice.

Teachers organised two Growth mindset workshops to introduce the basic idea of Growth mindset to students and help them apply it in real-life situations. Before the experiment started, participant teachers were invited to read the book, *Mindset: The New Psychology of Success* (Dweck, 2017). After each workshop, teachers gathered to reflect on students' responses, identify strengths, and discuss areas for improvement.

15.3.1 Initial implementation: direct use of materials

The first Growth mindset workshop was intended to raise the discussion of mindsets and introduce Dweck's Growth mindset theory. In the first demonstration lesson, the teacher organised following main activities:

> Warm-up discussion: What makes people intelligent? Students brainstorm and list characteristics or behaviours they associate with intelligent individuals.
> Discussion 1: Is a person's intelligence fixed or changeable?
> Reading activity: *《你可以变得更聪明》* (Yeager et al., 2016, Chinese version of You Can Grow Your Intelligence)
> Discussion 2: Reconsider your answer, Is intelligence fixed or not? Why?

> Discussion 3: Do you think you are clever? When do you feel clever? If a classmate who usually performs similarly to you scores higher on an exam, how would you feel and respond?
> Video activity: Carol Dweck – The Power of Believing That You Can Improve (https://www.ted.com/talks/carol_dweck_the_power_of_believing_that_you_can_improve?language=en)

15.3.2 Challenge identification: overload of content

The lesson was originally planned to last 50 minutes and aimed to cover all key materials related to the concept of a Growth mindset. However, it ended up extending to around 90 minutes and students lost interest, particularly in Dweck's TED talk. While the first Growth mindset workshop achieved its overall goal, the teacher expressed concerns that the concept of a Growth mindset may not have been effectively conveyed.

> I often find myself relying on existing materials, feeling constrained by the need to complete each step. Instead of approaching it from my understanding and guiding students to grasp the concept of a Growth mindset in a child-friendly manner, I focus too much on just finishing the activities. (Teacher Zjj, School A)

The teacher felt that the details of activity management could have been improved. For example, the teacher handed out the reading material before the lesson began to save time. Many students skimmed through it in advance and anticipated the 'correct' and 'preferable' answers for whether the intelligence could be developed, which undermined the purpose of the activity. Furthermore, some of the questions posed after the reading were repetitive, with both the teacher and students repeatedly addressing similar points, without engaging in a deeper exploration of the Growth mindset concept.

15.3.3 Solution proposal: selective use of materials

Both the teacher delivering the lesson, and the observing teacher realised that using the provided materials without adaptation did not fully meet their aims. They needed to understand Growth mindset first and select videos, images, and activities that were more suitable for their students to clearly grasp the Growth mindset. Additionally, attention needed to be given to the implementation details to prevent wasting time by repeatedly addressing the same question.

15.3.4 Practice refinement: refined growth mindset workshop

Based on the reflections of the first demonstration lesson, Teacher Hym made a quick adjustment of materials and activities to arrange the workshop. The warm-up sessions remained similar, but the reading material was handed out after the discussion on whether intelligence is malleable and what makes a person intelligent. After reading the article, the teacher took the opportunity to develop students' Growth mindset with following conversation.

> Teacher: When do you think you are clever?
> Student: …when we solve a problem
> Teacher: If you cannot solve that problem, are you still clever?
> [All students remained silent.] (Teacher Hym, School A)

At this point, the teacher played a cartoon showing a student struggling with chemistry but afraid to ask questions for fear of looking silly. A classmate reassured them that asking questions is helpful, not foolish, highlighting how a Growth mindset supports problem-solving in real classroom situations. Students learned from the video and became more willing to respond to questions that had previously left them silent.

> …just because you cannot do it now does not mean you will not be able to in the future.
> Do not easily dismiss yourself. (Class of Hym, School A)

The teacher then presented example statements reflecting both fixed and Growth mindsets and invited students to read them aloud. This 'saying-is-believing' activity, adapted from Growth mindset literature, aimed to help students recognise and reframe fixed mindset thinking. For instance, a statement like 'I cannot do better' was transformed into a growth-oriented version such as 'I will try another way, and I can do better'.

With the adjustment of Teacher Hym, the workshop was delivered smoothly within 50 minutes and students remained focused and engaged throughout. Students' reactions and teachers' reflections suggested the success of this workshop.

15.3.5 Reflection: skills in adjusting materials from research

The two demonstration lessons in the first growth mindset workshop highlighted the need to adapt research-based materials to classroom contexts. The adjustment process of the first workshop

indicated that teachers could benefit from others' experiences. They demonstrated the ability to gather information and resources from various channels to support their classroom activities, while also exercising sound judgment in avoiding misleading examples.

The second growth mindset workshop was to make growth mindset more self-relevant and help students practice the idea of growth mindset for daily issues. Building on the experience from the first workshop, teachers were generally more confident and flexible in the second session. They used the materials provided by the intervention and adapted the activities to better suit their students' specific needs. Teachers adapted the challenge board and hero's story into the second workshop from four regular 'saying-is-believing' exercises, including preparation for the next examination with growth mindset, a letter to encourage future students, a challenge board, and a hero's story (Appendix D.5).

Students were asked to describe a hero and explain how that person worked to achieve success. Students had a group discussion on the hero's efforts. Additionally, Teacher Zjj creatively designed a 'catching candy' game to remind students of the idea of Growth mindset. Then, the challenge board encouraged students to reflect on something they had always wanted to do but felt hesitant about because they were not confident in their abilities. Students were asked to create a plan to pursue that goal.

While students were highly engaged in the activities, teachers observed that some of their plans lacked the specificity needed to achieve the intended goals. For example, one student filled the challenge board:

> My challenge is not enjoying reading. The solution is to read every day. What I needed is starting with my favourite books and read more recommended books. The difficulties are not insisting reading and giving up. To overcome the difficulties, I will be patient. If the plan does not work, I will look for help from the teacher. My goal is enjoying reading at the end of the semester.

Although the student filled in all blanks, the answers were not specific or concrete, which was difficult to carry out actions. On the other hand, some students had a more specific answer. For instance:

> My challenge is that I run too slowly. The solution is running for 30 minutes a day after class. What I needed are a track, a stopwatch, strategies to improve my pace, and determination. The difficulties are hitting a plateau where, despite my efforts, there's little improvement. To

> overcome the difficulties, I will motivate myself to keep going. If the plan does not work, I will look for help from the P.E. teacher. My goal is to run 50 metres without feeling tired and to achieve a performance that's above the class average at the end of the semester.

In comparing the two examples, teachers were worried that students might not be applying the Growth mindset concept effectively and that their plans were not practical to implement. They held private conversations with students who struggled to set up their challenge boards and regularly followed up on the students' progress to ensure they were acting on their plans.

> Some students provided very vague descriptions of their challenges, which led to equally vague solutions. Their plans lacked structure, and in some cases, they didn't follow through with them at all. (Teacher Zjj, School A)

In response to this concern, teachers regularly checked on students' progress with their plans. They also integrated Growth mindset messages into their daily teaching and observed changes in students' attitudes toward learning.

> I asked them, 'Some of you made lots of mistakes in the exam. How do you think?'. Unlike before, when they were concerned about receiving a low mark and approached it with a passive mindset, most of them now said they could learn from their mistakes and avoid them in the actual examination. I felt that all the students were showing a positive and proactive attitude. (Teacher Hym, School A)

15.3.6 Conclusion

The investigation of Growth mindset highlighted that resources and guidelines derived directly from research require modification before and during practical application. Instead of strictly adhering to prescribed instructions, teachers are encouraged and given the flexibility to adapt materials based on their students' characteristics. Throughout this process, collaboration with colleagues proved crucial, allowing teachers to reflect on their practices and discuss lesson plans.

15.4 Writing about Values

Writing about Values was also a completely new strategy for teachers. Unlike the growth mindset activity, it required little adaptation, as it contained only two writing tasks and teachers were provided

with a checklist. However, despite its ease of implementation, Writing about Values still required careful organisation and monitoring by teachers to ensure students complete the tasks thoroughly.

At the initial implementation stage, teachers made considerable efforts to create a free-writing environment. They consistently emphasised, both through verbal reminders and printed notes on the writing sheets, that students should express their true thoughts without fear of judgment. To reduce pressure, some teachers asked students to place their completed work in an envelope and take it home immediately.

> In the classroom, we repeatedly emphasise writing down students' true thoughts, with the assurance that the teacher won't judge the content. We clearly stated this on the writing practice sheets. After completing their writing, students place their work in an envelope and take it home. This ensures that they don't feel burdened or pressured, allowing them to write freely. 'I write my heart.' (Research lead 3, School B)

> A few students, during the writing exercise, repeatedly asked, 'Will you mark it?' I told them, 'Don't worry, write freely, and write it for yourself.' (Teacher Cy, School B)

However, some students became less engaged in the second writing task. They put in less effort, writing carelessly or leaving tasks unfinished because they knew their work would not be collected. To ensure students completed the writing tasks with quality, teachers gave feedback during the activity to those who were less engaged and encouraged them to approach the task with a serious attitude. This also helped the teachers gain valuable experience for organising similar activities in the future.

> A few students, after hearing that the work wouldn't be collected, stopped being as serious as they were the first time. Even when there was enough time, some students became lazy. In future, it may be necessary to find ways to encourage consistent effort, possibly by introducing accountability measures that do not conflict with the goal of fostering creative and independent thinking. (Research lead 3, School B)

> I am worried if this waves the effectiveness of the two writing tasks. I am also thinking of how to strike a balance between free writing and adequate monitoring. (Teacher Cy, School B)

Since this strategy required minimal preparation, schools and teachers encountered few issues during its implementation. However, despite its ease of implementation, Writing about Values still required careful organisation and monitoring by teachers to ensure students complete the tasks thoroughly.

15.5 Conclusion

This chapter discussed the process of adapting evidence-informed strategies into teaching practice. These strategies included those unfamiliar to teachers and requiring long-term implementation, those previously used but lacking structure, those never used before with limited supporting materials, and those new to teachers but accompanied by detailed instructions. Teachers encountered different challenges when implementing these strategies. Some strategies (e.g., peer tutoring and Growth mindset) went through multiple rounds of adaptation and eventually took on a structured form in practice, while others may expect a longer period of adjustment before being effectively integrated.

It is feasible for teachers to incorporate research evidence into their teaching, though the quality of implementation may vary considerably. Teachers may interpret research evidence in various ways. They dedicated different amounts of time and energy to its adaptation, either a superficial borrowing of strategies or a deep alignment with their own teaching practices These differences can ultimately result in varied student learning outcomes. Overall, successful implementation required ongoing support through peer collaboration, especially when inappropriate implementation occurred. Teachers continually refined their evidence-informed practices, drawing on both their own experiences and insights gained from colleagues.

Chapter 16 Limitation, implication, and discussion

This chapter serves as the conclusion of the thesis. Following the concept of the research cycle (Gorard, 2013, p. 14), this study synthesised evidence on access to educational research, teachers' views on the use of research evidence, and the most promising approaches for disseminating research findings to teachers. Then, the study explored the feasibility of integrating research evidence into teaching practices and rigorously evaluated the impact of a use-of-research-evidence intervention on students' literacy outcomes in China.

While the study tried to deepen understanding of teachers' use of research evidence and its effectiveness, certain limitations may influence the findings' trustworthiness. This chapter begins by reflecting on these limitations and pulls together the findings. Training on basic research evidence and active follow-up support can help promote teachers' use of research evidence. Teachers' use of research evidence showed potential sustainable impacts on students' Chinese learning outcomes, teachers' skills, and teaching practices. However, maintaining and improving compliance and implementation quality will require sustained efforts from both schools and teachers. Therefore, the thesis concludes with recommendations for further research and evidence users.

16.1 Limitations

This study involved secondary data analysis for TALIS 2008-2018 and the NFER Survey of Teachers 2010, a structured review, and a randomised controlled trial with integrated process evaluation. Each sub-study aimed to answer different but complementary research questions. The limitations of each sub-studies appear in the following sections.

16.1.1 Secondary data analysis

Both TALIS and the 2010 Survey of Teachers were limited by the age of their data. The most recent available TALIS dataset was from 2018, and the Survey of Teachers data were collected in 2010. Even the most up-to-date dataset was already six years old. Over the past decade, the number of systematic reviews on the effectiveness of educational interventions and teaching strategies has grown. Teachers now have greater access to research evidence and support through initiatives such as the Research Schools Network and the EEF Teaching and Learning Toolkit. The movement of using research evidence may also bring pressure and opportunities to teachers to improve their practice. As a result, these old datasets may not fully capture the current state of teachers' access to educational research, views, confidence in their skills, and needs.

These two datasets, based on cross-sectional self-report surveys from teachers, may not offer an objective or comprehensive picture of their access to research evidence. In particular, the limited number of survey items restricts the potential for detailed analysis of teachers' perceptions, confidence, needs, and practice. as both datasets are cross-sectional, they only allow for the identification of correlations rather than causal relationships.

Sampling issues also raise concerns about the generalisation. In TALIS, some respondents represented specific regions rather than entire countries. For example, in China, TALIS survey included only teachers from Shanghai, which cannot adequately represent the situation across mainland China. In the Survey of Teachers 2010, the NFER contacted over 13,000 teachers, but only around 4,000 from Key Stages 1 to 4 were included in the final analysis.

Despite these limitations, these two datasets remain the most relevant sources available at the time of writing, offering valuable insights into teachers' access to research-related resources and their perspectives.

16.1.2 Structured review

As with any large-scale review, some relevant studies may have been missed. The structured review in this study was confined to publications in English and relied on a strict search criterion requiring terms related to 'evidence'. Given the large number of search results, the screening process was terminated if 100 consecutive entries were deemed irrelevant to the research topic. Though unlikely, this approach may have led to the omission of a few pertinent studies, but these less relevant studies would be unlikely to alter the main findings. Additionally, the review covered literature published up to January 2023, so more recent work may not have been included. These constraints inevitably narrowed the scope of the review.

Another possible limitation lies in the use of Gorard's two-dimension framework to categorise types of research evidence use. The distinctions between categories were not always clear-cut and required subjective judgement. In particular, the level of researcher engagement in the transfer process was not consistently defined across studies, which may have introduced variability in classification.

16.1.3 Randomised controlled trial

The main experiment involved two forms of randomisation. Three schools were randomised by the researcher, while one school conducted its own randomisation. These varying approaches to randomisation may weaken the validity of the results compared to full randomisation. Furthermore, teacher-level randomisation did not strictly prevent communication between participants. Although teachers in the treatment group and school leads were instructed not to share intervention materials, the possibility of contamination could not be entirely ruled out.

While this study had a reasonable attrition rate of 2% for the primary outcome of students' Chinese literacy, missing cases in the teacher survey and the student survey may also lead to some bias. Particularly in the self-randomised school, the response rate of both teacher and student surveys was much lower than in the researcher-randomised schools. Among 154 teachers, 106 teachers finished the teacher survey before and after the intervention. In the student survey, 3,216 students completed the two surveys, more than half of whom were in the treatment group. Some schools conducted the student survey during class, while some schools and grades found it difficult to arrange the survey during school hours due to the large sample size. Some students were asked to complete the survey at home. Teachers expressed concerns that parents might complete the survey on behalf of their children, potentially compromising the accuracy of the results.

To reduce practice effects from completing the Teacher Use-of-Research-Evidence Survey twice, this study used different questions in the pre- and post-tests. However, this approach may have introduced measurement differences due to varying question difficulty. A better design in future studies could involve creating two question banks and randomly assigning teachers to different sets. This would help minimise measurement error caused by difficulty variation while still preventing fatigue from answering identical questions.

Caution is needed when generalising the findings of the use-of-research-evidence intervention to broader contexts. The sample for both the pilot and main experiments was drawn from the same city in China, with participating schools being relatively prestigious and well-regarded within the city. These schools demonstrated a high level of awareness and enthusiasm, which likely motivated their willingness to participate in the experiment despite the additional workload it imposed on schools and teachers.

Furthermore, the intervention and the evaluation were delivered by the researcher. While the measurement for the primary outcome was standardised and independent, the classroom observation and interviews might have bias. The researchers' expectations and involvement in the intervention stage may lead to variations in the impact. Overall though, this was a large-scale and in-depth study with a wide range of data sources. It is robust with the limitations of any real-life research.

16.2 Findings summary

Despite these possible limitations, this study offers unique insights into the effectiveness of teachers' use of research evidence in the Chinese education context. A use-of-research-evidence teacher training intervention was developed, derived from a large-scale systematic review (Chapter 8) and refined through a pilot study (Chapter 9) to better support teachers' implementation. A randomised controlled trial was then conducted in Anhui, China, involving 6,904 students and 154 teachers from four primary schools, to evaluate the impact of the intervention on students' Chinese literacy outcomes, which represents the core contribution of this study.

A small but positive effect size of +0.07 was found on students' Chinese literacy attainment, with particularly pronounced benefits for low- to middle-achieving students, females, migrant students, and those from rural or low-income backgrounds (Chapter 10). Although the improvement was modest, the outcome measure was a local standardised exam, indicating that the intervention enabled tangible progress within the existing educational system. Beyond academic outcomes, a small reduction in time spent on Chinese literacy homework was observed (-0.06). The intervention also enhanced students' attitudes towards reading and their confidence in understanding texts (Chapter 11). These findings highlights that teachers' use of research evidence can contribute to enhanced learning efficiency and help reduce achievement gaps among students from different backgrounds.

The intervention influenced teachers' perceptions and skills of using research evidence (Chapter 12). While treatment teachers became more sceptical of most decision-making resources, they became more aware of the usefulness of research evidence (+0.26) and gained confidence in selecting, adapting, and applying research findings. They also reported fewer barriers related to access to research and workload. Teachers became better equipped to engage with research evidence. Improvements were observed in teachers' skills to accurately describe (+0.21) and interpret research evidence (+0.15), though effects on judging research quality and evidence-informed decision-making were mixed.

Teachers' adherence with the use-of-research-evidence intervention contributed to variation in its impact, with a Complier Average Causal Effect (CACE) of +0.22 (Chapter 13). Adherence was influenced by school-level factors, including research opportunities and leadership monitoring (also in Chapter 14), as well as individual teacher factors such as confidence, skills, and peer support, consistent with the secondary data analysis in Chapter 7. Successful evidence use depended on different combinations of these factors, as Qualitative Comparative Analysis indicated that no single condition reliably predicted outcomes. The process of adapting evidence-informed strategies could be summarised in five stages: initial implementation, challenge identification, solution proposal, practice refinement, and reflection (Chapter 15). These insights from process evaluation provide valuable guidance for designing future interventions and supporting teacher professional development.

16.3 Use of research evidence: a feasible movement for teaching

This study highlights the feasibility of teachers' evidence-use. Teachers have the flexibility and autonomy to adapt research evidence into teaching practice when offered in the form of a menu choice. This flexibility, combined with the intervention's design, made it cost-effective and ensured it did not disrupt schools' regular work pace, which makes it applicable for use with a larger sample size. The intervention can boost their awareness of the usefulness of research evidence and enhance their skills in evidence-use. Changes in their practice were also observed.

16.3.1 Teachers' views on the value of research evidence

Compared to other European nations, England has the fastest-growing use of research evidence in education, and it is also the first to have a legislative provision that specifically emphasises its use (Pellegrini & Vivanet, 2021). New research evidence on educational projects is being explored and disseminated by organisations like the Evidence-based Teachers Network and the Education Endowment Foundation.

While researchers and policymakers dedicate much effort to finding out and disseminating research evidence, existing studies on teachers' perceptions of its usefulness may present a biased picture. About a quarter of teachers who participated in the Survey of Teachers 2010 did not claim any experience of using research findings. The proportion of teachers holding a positive view of research evidence for teaching practice was notably smaller than suggested by earlier studies (e.g. Everton et al., 2002; Mills et al., 2021). This study found a preference among teachers in England for alternative approaches, such as reflecting on their own practices rather than relying on research evidence to enhance teaching.

England has advanced considerably in incorporating research evidence into educational practice, but its application in other contexts, such as China, remains relatively underexplored. Although there has not been a nationwide investigation into teachers' views on the value of research evidence in China, this study provides valuable insights from teachers participating in the use-of-research-evidence experiment. It might be challenging for teachers to take the first step in adjusting their existing teaching practices based on research evidence. Headteachers and school research leads in the experiment noted difficulties in engaging teachers, especially those who were middle-aged. However, teachers who received the intervention demonstrated notably more positive attitudes towards the usefulness of research evidence. When teachers gain more familiarity with research-related concepts and have opportunities to adapt research evidence, they may develop more positive views on the value of research evidence.

Interestingly, this increase in perceived usefulness of research evidence did not diminish the scepticism teachers expressed towards its trustworthiness. Teachers often remain sceptical of educational research, often rejecting it due to concerns about its lack of contextual relevance or scientific rigour (Sato & Loewen, 2022). The use-of-research-evidence intervention encouraged treatment group teachers to critically evaluate the quality and relevance of research, which made them more aware that not all educational studies are equally trustworthy. This critical engagement likely contributed to their cautious stance towards the trustworthiness of various decision-making resources, particularly towards academic researchers and educational articles.

In addition, Lysenko et al. (2014) suggested that teachers' attitudes and beliefs about research evidence may be a strong predictor of their willingness and ability to use it in daily teaching. Consistent with this, the present study found treatment teachers accessed social media more frequently, which could be the quickest and most cost-effective way to obtain the latest research findings. Teachers were actively seeking and preferred easily accessible research sources, whether to directly support teaching practice or to expand professional knowledge. Such openness and proactive engagement are crucial for fostering sustained interaction with research evidence. However, in this study, there was no direct evidence to suggest that greater awareness or more positive attitudes were associated with stronger compliance or higher-quality implementation.

16.3.2 Empowering teachers' use of research evidence through training

Aligned with prior research (e.g. McGoey et al., 2014; Thomm et al., 2021), key predictors of teachers' perceived usefulness of research findings include their access to relevant research evidence, time constraints, and confidence in their skills to understand research. These restrictions discourage teachers from engaging with research, leading them to conclude that educational evidence is irrelevant or impractical for their teaching.

Professional development activities offer a promising route to engage teachers with educational research and enhance their ability to interpret research findings. However, existing studies suggest that teachers express dissatisfaction with the quality of professional development opportunities for engaging with research evidence (e.g., Mills et al., 2021). This dissatisfaction may stem partly from an insufficient focus on practical access to research evidence and concerns about increased workload (Williams & Coles, 2007). As a result, they may continue relying on ineffective practices.

This study highlights that appropriate support can transform teachers' engagement with research evidence. The use-of-research-evidence intervention addressed three critical barriers. It provided sufficient access to research channels, offered guided practice in using resources effectively, and achieved this without increasing teachers' overall workload, which mitigated teachers' concerns of using research evidence. Teachers may require more concrete and accessible examples to guide their use of research evidence. Example lessons and teacher collaboration may provide a more concrete understanding of how to adapt research evidence effectively.

Findings from this study support the value of professional development (not limited to conferences but also including teacher practice and follow-up support). Teachers' attention to research evidence increased, leading them to avoid sifting through ambiguous or irrelevant information on general search engines. Beyond academic databases, social media may represent the most cost-effective way for teachers to quickly access research information.

Appropriate training on evidence-use can enhance teachers' confidence in their ability to integrate research into their teaching and develop a greater interest in conducting their research. However, treatment group teachers showed decreased confidence in understanding or judging the trustworthiness of research. This could be attributed to the intervention, which, by exposing teachers to more information about research evidence, may have led them to adopt a more cautious or self-critical perspective on their abilities, or to develop a more critical stance towards research in general.

Teachers are able to clearly describe the definition of research evidence, gain insights into the statistical aspects (e.g., 'effect size') of research findings, and they can show some awareness of the differences in research evidence quality. Through this process, they came to realise that developing these skills is more complex than it initially appears. These findings highlight the feasibility and impact of carefully designed professional development initiatives in fostering teachers' research engagement.

16.3.3 Can teachers adapt research findings into teaching practice?

Changing established teaching practices can be challenging (e.g., Cuban, 2013; Davis, 2003). In this study, some teachers showed reluctance to adopt evidence-informed strategies, often doing so in a superficial or half-hearted manner. At the initial investigation stage, many teachers raised concerns about how to implement specific evidence-informed strategies within their own teaching contexts. Some teachers confused the use of research evidence with the application of educational theories. However, this study demonstrates that teachers are capable of incorporating high-quality research findings into their teaching, whether by implementing new strategies or adjusting their existing practices.

Teachers can extract relevant insights from the evidence and adapt strategies to align with their specific teaching requirements, showcasing their ability to bridge research with practice. Even with a completely unfamiliar evidence-informed teaching strategy, such as the Growth mindset and Writing about Values, teachers demonstrated the ability to cooperatively explore its application. They thoughtfully incorporated their own interpretations and adjustments into the approach rather than simply relying on the intervention materials.

During the process of using research evidence, teachers demonstrated notable flexibility. While their activities and lesson plans reflected the origin of research evidence, they were creatively adapted to suit specific classroom contexts. These findings challenge concerns raised in some literature that evidence-use might undermine teachers' autonomy (Holloway & Larsen Hedegaard, 2023). In contrast, this study suggests that evidence does not dominate, as teachers actively interpret and adapt evidence-informed strategies in ways that best serves their students' needs.

In addition, monthly workshops and demonstration lessons might be good opportunities for teachers to reflect on their use of specific evidence-informed teaching strategies. These opportunities allow teachers to assess both their own and their peers' teaching practices, exchanging experience, identify

areas for improvement, and make necessary adjustment. Some teachers innovated their own teaching structures with the support of existing research evidence. Therefore, teachers' evidence-use is a feasible and developing process with teaching practice and reflection.

16.3.4 Cost-effectiveness and minimal disruption

Consistent with the structured review in Chapter 8, teachers' use of research evidence can be cost-effective (e.g., Lord et al., 2017b; Rose et al., 2017; Wiggins et al., 2019). This study adopted a broad approach that required no specific products or costly resources. Schools can easily get access to online resources, such as teaching plans, activity sheets, and example teaching videos. For schools seeking more structured guidance, relevant books can be purchased through school funding. Except for the existing resources, schools can accumulate their school teaching materials that can be usable for years and appropriate for their school-specific needs. The low cost of this approach makes it particularly attractive for schools operating on limited budgets and facilitates wider implementation.

Low disruption to schools' regular teaching schedules would be another key aspect that ensures the feasibility of using research evidence. The intervention respects each school's autonomy and preferences. Schools have the flexibility to choose from a range of evidence-informed strategies that align with their development goals. By focusing on improving teaching methods rather than altering the curriculum, the project ensured that most strategies could be incorporated into daily lessons with only minor adjustments.

For instance, the embedding formative assessment and peer tutoring are two regular strategies in the Chinese teaching context. Participant schools chose them to improve and structure existing teaching practice. Conversely, dialogic teaching was not selected in the main experiment as it did not meet the specific needs of these schools. Approaches like Growth mindset and Writing about Values, though involving specific activities, can be easily implemented during two regular Chinese literacy lessons. The schools selected these strategies to alleviate academic pressure among senior-grade students and foster a healthy learning and competitive atmosphere. They remain feasible, even for senior-grade students, despite limited time.

The use-of-research-evidence intervention involved a 90-minute presentation and monthly workshops. The presentation could be delivered during after-school meetings, requiring no additional work hours for teachers or extra space for schools. In China, schools regularly hold teaching workshops and meetings to maintain teaching quality and innovation (Sargent, 2015), which aligns perfectly with

the structure of the monthly workshops in this intervention. Even in schools with lower compliance, the intervention slightly improved students' Chinese learning outcomes.

In addition, teachers' perception of research evidence and their skills are improved with the use-of-research-evidence intervention, which may have an impact on their further teaching practice. Considering the long-term benefits, the time and financial investment for training are minimal compared to the potential long-term benefits they can bring to both teaching quality and student outcomes. The sustainability of evidence-use is discussed in detail in Section 16.5.

16.4 Implementation quality

Even well-supported programmes often fail to achieve their intended outcomes due to poor implementation. Implementation science focuses on identifying the critical elements needed to ensure the genuine adoption of research findings, thereby enhancing their overall effectiveness (Moir, 2018). As Lyon and Bruns (2019) suggested, 'there is no implementation without adaptation.' In this use-of-research-evidence project, teachers' compliance with the intervention was linked to its effectiveness on students' learning outcomes with a CACE of +0.22.

Existing frameworks about implementation quality involve the regional, school, and individual levels. Domitrovich et al. (2008) proposed a multilevel implementation quality framework that integrates the regional political context, school dynamics, and individual contributions. Strunk et al. (2016) emphasised that plan quality at the school level is positively related to the implementation quality in school improvement initiatives. The implementation process from existing research emphasis on building a supportive environment, sustaining implementation efforts, and iteratively improving practices to ensure long-term effectiveness (Fixsen, 2009; Meyers et al., 2012). Additionally, the de-implementation and implementation strategies outlined by Wilcox et al. (2023) suggest that the adjustment of teaching practice required administrative support to determine how to reinforce the adoption of alternative teaching approaches.

Inspired by existing implementation frameworks and grounded in the process evaluation conducted in this study (Chapters 13–15), this study proposes a diagram illustrating five core steps for adapting research evidence, highlighting the role of teachers' individual efforts and collaboration, within a supportive school environment (Figure 16.1). Factors at multiple levels shape teachers' compliance with the intervention, which in turn underpins high-quality implementation. At the school administrative level, effective planning and close monitoring are largely relevant to teachers' exposure

to the intervention and efforts to transfer research findings into teaching. The fluent and frequent interaction between leads and teachers may also alleviate the concerns of existing workload. At the individual teacher level, their skills and confidence in evidence-use are relevant to their compliance with the use-of-research-evidence intervention and their implementation quality. In addition, teacher peer interaction is an essential factor during the process, which has been ignored or rarely noticed by prior frameworks. Teachers are mirrors to their peers, where their interaction could support each other to discover ineffective teaching practices and adjust based on research findings.

With the support of individual teacher efforts, peer collaboration, and a conducive school environment, the process of adapting research evidence began with initial implementation (phase 1), triggered by the use-of-research-evidence intervention. As the process unfolded, teachers often identified elements of the research-based materials that were either unsuitable or incorrectly applied in their context (phase 2). They responded by proposing solutions (phase 3), abandoning ineffective practices, refining their teaching (phase 4), and engaging in reflection (phase 5). This implementation cycle reflects a collective trajectory observed across the teacher sample and is further analysed in Section 16.4.4, where classroom observations are used to discuss how different types of evidence-informed strategies were adapted in practice.

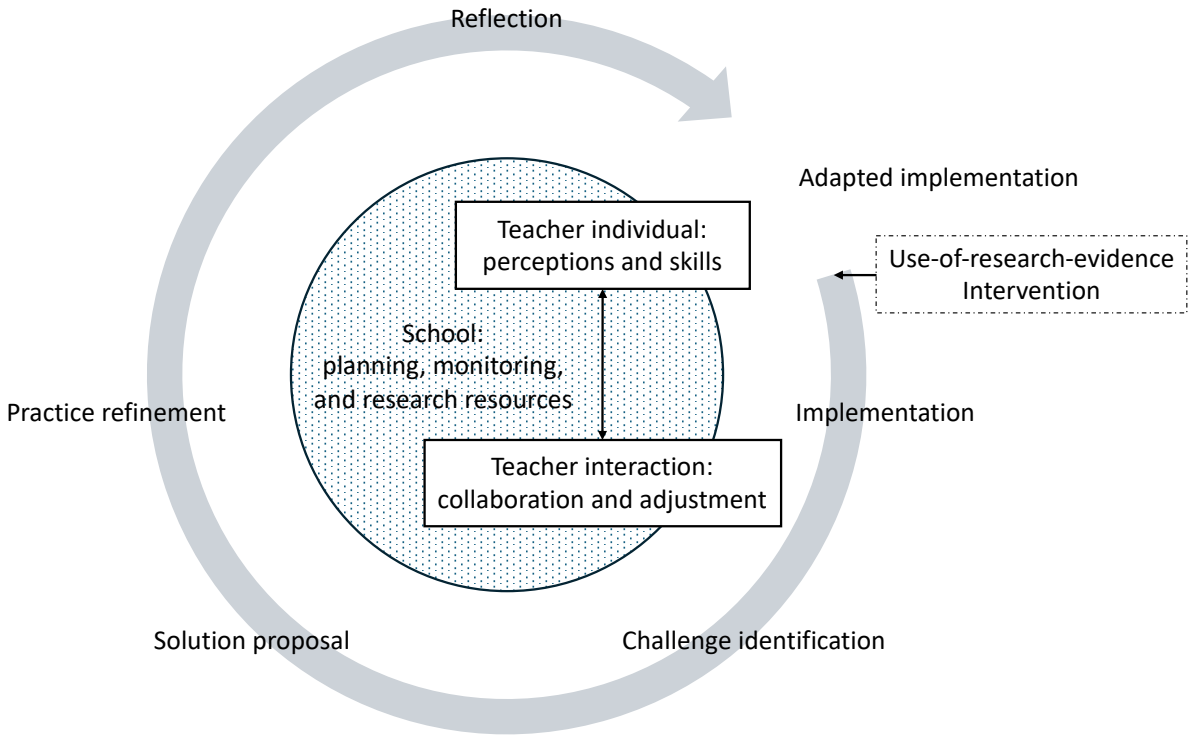


Figure 16.1 Diagram of implementation at school and teacher levels

16.4.1 School-level factors: planning, monitoring, and research tradition

It is crucial for schools to select strategies that align with their unique contexts. In the use-of-research-evidence experiment, schools typically tailored the proffered interventions to their needs, discarding those entirely unfeasible for their conditions. However, selecting a feasible evidence-informed practice is only one basis for using research evidence. School-level planning, regular monitoring, and the allocation of research resources shaped the context in which teachers engaged with the use-of-research-evidence activities and thus related to the success of their exploratory practices.

Effective planning is essential to ensure sufficient activities, such as presentations or professional development conferences for knowledge about using research evidence and workshops to discuss and reflect on teachers' evidence-informed practice. Without adequate exposure to these activities, teachers' ability to implement evidence-informed strategies may be compromised. For example, in this study, the self-randomised school failed to schedule monthly follow-up sessions for all its teachers. As a result, none of the teachers consistently participated in intervention activities. These teachers expressed a need for more specific training and guidance from researchers. Unsurprisingly perhaps, this school demonstrated smaller effect sizes in its use of research evidence compared to others that scheduled regular monthly workshops.

Teachers' existing workload may always be a concern when teachers are encouraged to use research evidence. Since evidence-use is not yet a standard expectation in most schools, incorporating it to improve teaching practice is often perceived as an additional responsibility for teachers. Reducing the frequency of these activities to accommodate teachers' workload may not only diminish the effectiveness of the intervention but could also inadvertently increase teachers' sense of burden. Without sufficient support through regular activities, teachers might feel overwhelmed by the lack of guidance, perceiving evidence-use as more complex and unmanageable.

Instead, striking a balance between planning adequate activities and supporting teachers' workload is crucial for successful evidence-use implementation. Schools can integrate these activities into the existing timetable with minimal disruption to teachers' schedules and consider replacing regular teaching workshops with use-of-research-evidence workshops.

The level of planning required may vary depending on a school's size and operational model. In the main experiment, large-scale schools developed detailed plans at the start of the intervention semester, including schedules for presentations, workshops, and school visits. They also planned

specific dates for activities like Writing about Values. Such planning was essential given the complexity of coordinating tasks in larger schools, where teachers' demanding workloads required advanced scheduling to avoid conflicts with other responsibilities.

Conversely, a smaller school, which had roughly one-third the number of classes compared to the larger schools, adopted a more flexible approach. Their workshop and school visit schedules were adjusted based on the school's regular schedule, supported by a collaborative environment that allowed teachers to communicate easily and work closely together. This level of cooperation reduced the need for tightly coordinated planning while maintaining clear objectives and designated leads for the intervention.

Beyond effective planning, adequate monitoring and fostering a collaborative and supportive environment may also facilitate the successful adaptation of research evidence into teaching. Teachers tend to believe in the value of research findings and to incorporate them into their teaching when they work within a comfortable and supportive environment which praises their efforts to understand and apply research (Joram et al., 2020). According to this study, leadership's participation is relevant to teachers' compliance and potentially encourages teachers to explore the use of research findings.

Headteachers and school research leads may need to be actively involved in teachers' exploration of research evidence, regularly engaging with them to understand their challenges and maintaining close communication with researchers. Headteachers' instructional leadership could directly work on teachers' instructional quality (Bellibaş et al., 2021). Their involvement helps surface teachers' difficulties more quickly, allowing timely support and guidance. At the same time, when teachers see leadership taking an active role, they may be more likely to view the use of research evidence as a genuine effort to improve teaching rather than a procedural task. This sense of shared purpose could encourage greater commitment. Leadership involvement also functions as a form of management, ensuring accountability and sustained engagement with evidence-use.

The collaboration may exist more frequently among teachers, while schools act as enablers who allocate time, opportunities, and foster a culture that supports teacher collaboration. Teachers receive feedback from their peers and find inspiration from demonstration lessons by their peers, incorporating effective techniques into their own practices. However, without school-level

organisation of such activities, teachers may lack sufficient opportunities to showcase their teaching and learn from peer feedback and experience.

Additionally, schools with strong traditions of teacher research have ample awareness and experience in organising and supporting teachers' engagement in research. They understand how to monitor teachers to maintain implementation fidelity. For example, one school in this study incorporated teacher research into annual performance review while another combined the current project with another school-level research project. On average, teachers in such schools have more opportunities to get involved in school research. They tend to be more research-oriented and demonstrate higher levels of engagement with educational initiatives, further enhancing the effectiveness of interventions.

16.4.2 Teacher individual level factor: perceptions and skills

Previous research has shown that in-service coaching can foster ongoing professional growth and enhance teachers' precision in applying evidence-based practices (e.g., Kretlow et al., 2012; Reinke et al., 2014). However, the effectiveness of such training and ongoing support also depends on individual teachers' engagement and willingness to participate. Teachers ultimately make the decision on whether and to what extent to apply high-quality research findings in their daily teaching.

Findings of this study suggest that enhancing teachers' confidence and skills in using research evidence may be an effective way to promoting their active engagement with the use-of-research-evidence intervention. Among these factors, skills appear to play a more critical role than confidence. While compliant teachers showed stronger skills in using research evidence, their confidence levels were largely similar to those of non-compliant teachers, except for being more confident in selecting appropriate research for their teaching. They also tended to be more sceptical about the overall usefulness of research evidence.

This could be partly explained by the possibility that compliant teachers already held high regard for the value of research. However, it also implies that valuing research in itself does not necessarily lead to use of evidence. What enables teachers to meaningfully integrate research into practice lies more in their skills, including interpreting research findings, judging research quality, and making evidence-informed decisions. These skills empowered teachers to adjust their teaching practices based on a sound understanding of high-quality research evidence and informed decision-making. The use of research evidence process requires individual inquiry, through which teachers recognise the gap

between research contexts and real classrooms and eventually develop evidence-informed practices suited to their own classes.

16.4.3 Teacher interactive level: collaboration and adjustment

Teachers' incorporation of research evidence into their practice is neither an isolated effort nor simply a response to school mandates. Instead, collaboration with peers plays a role in supporting this process. Teachers who are exposed to stronger peer support, which includes more opportunities to observe others' practices and to receive feedback through delivering demonstration lessons, are more likely to comply with the use-of-research-evidence intervention. Compared to working alone, collaborative approaches may be more feasible, as shared challenges can be discussed, and strong practices can inspire peers through demonstration lessons.

Peer observation and feedback is a promising practice for instructional improvement (Ridge & Lavigne, 2020). This process helps teachers feel that they are not alone in adapting research evidence but are part of a collective effort. Delivering demonstration lessons and receiving feedback from peers enables teachers to identify shortcomings in their own adaptations and collaboratively explore solutions. Observing peer classrooms also offers fresh perspectives and insights into alternative, evidence-informed practices. Teachers can learn from others' successes and their missteps, reflecting on which adaptations are inadequate and working toward more effective approaches. Without peer support, teachers may repeatedly struggle with similar 'pitfalls' in evidence adaptation. Given their limited time and effort, such repeated challenges could lead to inefficient or ineffective use of research evidence.

Beyond the practice of using research evidence itself, the process of exploration may offer teachers, particularly those early in their careers, valuable opportunities to refine their general teaching skills. If teachers struggle with 'what to teach,' the question of 'how to teach' might be merely a superficial constraint. In this use-of-research-evidence project, some early-career teachers' inadequate adaptations were not necessarily due to a lack of commitment but rather stemmed from broader difficulties in their teaching. Peer support often extended beyond the use of evidence-informed strategies, focusing more fundamentally on pedagogy and interpretation of curricular content. This initiative may serve to develop teachers' awareness and skills in using research evidence and support their overall professional growth.

16.4.4 Cycle of adapting research evidence

The implementation process is an iterative process involving the adjustment of research-informed strategies and modifications to existing practices to suit specific classroom dynamics, with a reflective process. The cycle of investigating evidence-informed teaching practice has been illustrated in Figure 16.1. This iterative cycle involves five phases and underscores the potential for teachers to evolve their practice through critical engagement with research evidence and reflection.

When applying a new evidence-informed strategy (e.g., Growth mindset, formative assessment, and Writing about Values in the current study), teacher may have no clear concept of how to construct lessons using these new strategies. They may dedicate substantial time and effort to interpreting both the research findings and supplementing instructional details. The adaptation may require teachers' skills in understanding, judging, and using research evidence. At the initial implementation stage (phase 1), they establish a logical connection between their students' needs and the rationale for how the new strategy addresses these needs. Many teachers initially imitate and follow the simplified instructions derived from the research.

However, research evidence is typically derived from experimental contexts that differ from classroom realities, and research fails to offer detailed implementation guidance. Teachers may interpret the same evidence differently, leading to variations in the quality of implementation. Teachers demonstrated instructions derived from research were insufficiently specific and inadequate for practical implementation (phase 2). For instance, the Growth Mindset content was too long, causing students to lose interest and undermining the workshop's objectives. In such moments, adhering rigidly to research-based instructions could result in teachers losing effective control over the lesson. Therefore, teachers propose solutions (phase 3) and adjust their teaching pace and materials (phase 4) by drawing on demonstration lessons, peer feedback, and student interactions. They update the lesson plan and pass it on to the next teacher scheduled to teach the lesson, summarising key points and reflecting on aspects that require special attention (phase 5).

The implementation process seemed more streamlined when teachers worked with strategies that were familiar but not yet systematically employed in their practice (e.g., peer tutoring). Teachers more readily understand the required actions, having either observed these strategies in use or applied them to some extent in the past. In the initial adaptation phrase, teachers can quickly embed suggested activities to their teaching practice (phase 1). However, not all activities were feasible in

real classrooms. For example, some teachers scheduled peer tutoring in the final ten minutes, but time constraints prevented students from completing the task (phase 2).

When such issues are observed, teachers quickly notice and document them. They raised these challenges during workshops, inspiring their peers and discussing potential solutions (phase 3). This unexpected difficulty with the writing task, rather than discouraging teachers from using peer-assisted writing, led them to refine the activity. They simplified it into a fill-in-the-blanks exercise, making it more manageable within the available time (phase 4). During the reflection and discussion phase, teachers agreed that this adjustment was both practical and worth continuing (phase 5).

16.5 Sustainability of using research evidence in teaching practice

While the experiment on teachers' use-of-research-evidence in this study lasted only one semester, it helped equip teachers with the skills to continue adapting research evidence into their daily practice over time. Some evidence from this study supports the potential long-term impacts. The interaction between teachers' changes in their teaching practice and students' learning status might create a positive feedback loop.

16.5.1 Potential long-term impact

Interventions that encourage teachers to actively explore evidence-informed strategies and build relevant skills may lead to sustainable changes in teaching practice. For instance, Bierman et al. (2013) found that even one year after the trial, teachers continued to implement evidence-informed strategies with high quality and demonstrated overall improvements in their teaching. During the short-term implementation, this study highlights the change in teachers' awareness, skills, and teaching practice, which could persist well beyond the intervention. Headteachers and school research leads all reported noticeable changes in teachers, particularly among those who had previously been less active in school teaching activities. The use-of-research-evidence intervention seems to provide teachers with the opportunity to make a shift in their teaching habits and perceptions.

Such improvement in teaching quality may subsequently enhance student learning outcomes. The current study found that treatment students initially lagged behind the comparison group on average, while after a semester of intervention, they caught up with the comparison group. If treatment group students continue to progress at this rate, they may surpass the control group by the end of the next semester. Additionally, with a longer period of implementation, teachers may develop a deeper

understanding of research evidence, and the impact could be more pronounced. The self-randomised school had approximately two weeks less exploration time and showed relatively weaker effects. Compared to the four-month total implementation period, even a two-week delay could make a difference.

Treatment group students expressed a stronger passion for reading and a greater belief in its value. They also reported higher confidence in their ability to understand texts. Teachers' observations further suggest that students engaged more actively in classroom interactions. These changes in attitudes, confidence, and daily performance are likely to influence students' learning efforts and contribute to better academic performance over time (Clark & De Zoysa, 2011; Petscher, 2010).

Moreover, schools may take long-term actions to pursue the use of research evidence. In China, schools are encouraged to conduct their own research as an indication of school evaluations. This use-of-research-evidence intervention provides a platform for school-level research and equips school leaders and teachers with research skills. These skills foster high-quality school-based research, and the outcomes can contribute to new evidence supporting further evidence-informed teaching practices.

Currently, two experimental schools have shared the study materials with their teachers across different subjects, reflecting the broader applicability of these strategies beyond Chinese teaching. At the start of the Autumn semester 2024, one experimental school reflected on their practices with the report of this evidence-use research and developed plans for the new semester. In March 2025, one headteacher established a school research network. These actions align with the idea of sustainability from McIntosh et al. (2009), where practices become ingrained as a natural part of institutional operations rather than being seen as temporary interventions.

These actions provide further evidence that use-of-research-evidence intervention can be a lasting influence on teaching methods, potentially influencing their approach to teaching over the long term. Many of the strategies introduced are not limited to Chinese literacy teaching instructions but can be applied across various subjects.

16.5.2 A positive loop

Evidence supports teachers' decision-making about their teaching strategies that have been rigorously evaluated for the impact on students, potentially reducing unexpected workload and ineffective

efforts. In the education context of China, for instance, updates to teaching materials, changes in curriculum standards, and the integration of digital tools into education have all transformed the teaching landscape. Headteachers in this study highlighted the increasing expectations from parents and society for higher teaching quality, especially in the post-pandemic era.

In such a demanding environment, research evidence serves as an alternative resource that differs from relying solely on theoretical books or drawing on teaching experiences. Incorporating rigorously evaluated research findings, evidence-use lowers the risk of ineffective strategies and saves teachers time and effort that might otherwise invest in low-quality, low-impact approaches. This also provides teachers with fresh perspectives and practical strategies that can complement their existing knowledge and support innovative teaching practices.

Students may receive benefits from teachers' change in practice in terms of their learning burden, attitudes, and literacy proficiency. The double reduction policy in China underscores the reduction of the quantity of homework as an indicator of students' burden, with one key emphasis being the reduction of homework burdens (Xue & Li, 2023). This study found that treatment group students spent slightly less time on Chinese homework compared to their peers in the control group, while showing equal or even greater improvements in reading attitudes and Chinese literacy outcomes. This creates a positive cycle of improvement.

Since students could achieve the same or even better learning outcomes with a reduced homework burden and increased enjoyment, teachers can ease their workloads by spending less time on correcting homework and addressing reluctant students. This allows them to dedicate more time to integrating high-quality research evidence into their daily practice. This positive loop, where evidence-use leads to better teaching efficiency and student outcomes, potentially encourages and inspires teachers to keep exploring research evidence for promoting teaching practice. This loop supports the sustainability of evidence-use, contributing to a more balanced and effective teaching environment. The use-of-research-evidence intervention acted as a gentle push, setting the interaction between using research evidence and teaching effectiveness in motion. Once teachers experience the benefits of a lighter workload and improved results, they are more motivated to engage further with research evidence.

16.6 Implications for researchers and policymakers

16.6.1 Recommendations for further research

Given the limitations and implications of the current study, future research could focus on the development and the evaluation sides. The development side encourages future research to focus on accessible, high-quality evidence with clear and practical guidance and adequate training for teachers to facilitate their use of research evidence. On the evaluation side, priorities include understanding teachers' perspectives on evidence-use, examining its impact across diverse contexts, and enhancing implementation quality.

At the developing level, access to high-quality research evidence is foundational for teachers' evidence-use. However, many studies lack robust designs, such as the studies without comparison groups, making it challenging to draw reliable conclusions about effectiveness. Establishing quality standards to identify promising research evidence and providing clear, practical implementation instructions are crucial. Schools and teachers also need accessible resources, whether commercially available products, books, or open-access materials, to facilitate their use of high-quality research findings.

Schools and teachers often face challenges in understanding research findings and selecting appropriate research evidence for adaptation to improve teaching practice. While fostering such skills may initially heighten teachers' scepticism towards research, researchers can mitigate this by offering detailed explanations to strengthen trust in high-quality evidence. Training and follow-up activities are expected to help build the skills needed for effective evidence-use and enhance the possibility for long-term use. However, such materials are often inaccessible or poorly advertised. Educators may not even know where they can get this support.

Meanwhile, future research should pay particular attention to supporting teachers' capacity to judge the quality of research evidence, an area that proved especially resistant to improvement in this study. With the increasing availability of online resources that present research evidence in accessible language, teachers now have more avenues to engage with research findings and are becoming increasingly reliant on these sources of information. This comes with a large risk of encountering overwhelming and potentially inaccurate content (e.g., Test et al., 2015), requiring teachers to exercise sound judgment and carefully select the most relevant and reliable evidence for their classrooms.

Providing access alone is not sufficient. Developers aiming to support teachers' use of research evidence must also carefully consider appropriate dissemination approaches. Existing systematic reviews (e.g., Erkan et al., 2021; Gorard et al., 2020) and the structured review conducted in this study emphasised that passive dissemination of research evidence provided limited support for teachers' use of research evidence. In contrast, interactive activities offer teachers opportunities to engage in meaningful exchanges, fostering a collaborative platform for learning. Researchers may also explore creative strategies for re-packaging research evidence in accessible formats and disseminating it through diverse channels that better match teachers' needs.

While teachers have the flexibility and autonomy to adapt research findings into teaching practice with their students' learning status, they are also expected to implement evidence-informed teaching strategies aligned with suggestions and instructions from existing high-quality research findings. Researchers and future evidence users could work more on how to support teachers to maintain high-quality implementation of using research evidence, which requires careful organisation and oversight at the institutional level. Researchers are recommended to work closely and cooperatively with schools to guarantee sufficient exposure to evidence-use activities.

At the evaluation level, insights into teachers' perceptions of research evidence and its effects remain underdeveloped. The most recent large-scale cross-sectional dataset about teachers' views and confidence in using research findings is Survey of Teachers 2010, which could not accurately illustrate the current situation of teachers' use of research evidence. Access to research evidence and attitudes towards its use likely vary with teaching experience, country, and current educational contexts. It is necessary to hear the facilitators and barriers of using research evidence from teachers' perspectives. Future research may also consider exploring these variations through longitudinal or repeated cross-sectional studies to track trends and shifts in teachers' perspectives.

Future studies can evaluate the impact of teachers' use of research evidence on students' outcomes, extending the focus from cognitive domains such as language and literacy, mathematics, and science to non-cognitive skills. Such studies should employ larger-scale, longer-term designs across diverse countries and age groups to generate robust evidence on the effectiveness of evidence-use. Longitudinal studies are particularly needed to assess whether short-term improvements in attitudes and efficiency, as observed in this study, translate into sustained gains in learning and teaching quality. Comparative studies across different educational systems could also clarify whether the facilitators and barriers identified in China are context-specific or more universal.

16.6.2 Recommendations for policymakers

This study demonstrates that evidence-informed teaching has the potential to enhance educational quality and equity at relatively little cost, with implications for long-term learning benefits. However, the adoption and sustainability of such practices cannot rely on individual teachers or schools alone. Government leadership and systemic support are essential to embed evidence-informed teaching into mainstream practice.

In the Chinese context, an urgent priority is to strengthen the production of rigorous research evidence in education, particularly those addressing the question of 'what works' in Chinese classroom teaching. At present, education research in China has largely relied on conceptual approaches and action research (Yao, 2021), which can hardly establish robust causal claims about effectiveness. The selection of evidence-informed teaching strategies in this study was also constrained by the limited availability of local research. As a result, the included strategies had restrictions in terms of generalisation and practical implementation, lacking context-specific guidance tailored to Chinese classrooms. Most of the books we used were international publications translated into Chinese, which not only introduced substantial delays in accessibility but also reduced the timeliness and relevance of the research. By the time these resources became available, student needs and educational contexts in China might already have shifted, further limiting their practical value.

This highlights the need for locally generated, timely research evidence that can respond directly to the evolving challenges of Chinese classrooms, reinforcing the urgency for policymakers to invest in building a national infrastructure for large-scale trials and systematic evaluations, alongside databases that collate and disseminate reliable evidence of effective practices. Such initiatives require targeted funding and coordinated efforts between universities, schools, and government agencies, but they would lay the foundation for a more evidence-informed education system.

Looking at the global evidence-use movement, leading countries such as the UK and the US have established institutions like the Education Endowment Foundation and the What Works Clearinghouse to conduct large-scale RCTs and synthesise the effectiveness of educational products and teaching strategies. However, research shows that schools do not necessarily consider research evidence when making decisions (e.g., Brown & Greany, 2018; Pegram et al., 2022).

Beyond generating robust evidence, it is equally important to foster an evidence-informed decision-making environment that helps schools understand why evidence should be used. This could be supported through public lectures, short videos, policy incentives, or accountability frameworks. Schools need first to be exposed to the rationale for evidence-use ('input') and subsequently be subject to regular monitoring of how evidence is actually applied in practice. For example, in England, schools are required to publish a Pupil Premium strategy that explicitly links their spending decisions to research evidence. While there is little evidence that such monitoring substantially improves schools' use of research evidence, these measures may nonetheless help cultivate a gradual habit of referencing research in decision-making.

At the same time, stronger collaborations between universities, government agencies, and schools are needed to address the practical challenge of how teachers can effectively use research evidence in their classrooms. Evidence-use modules could be integrated into school-based practicum experiences for initial teacher training, allowing pre-service teachers to develop an understanding of evidence-informed practice and apply it in real classroom settings before they begin their careers. For in-service teachers, continuous professional development (CPD) programmes should provide ongoing opportunities to translate research evidence into practice, including co-designed workshops, school–university partnerships, and regional networks where teachers can share experiences, receive feedback, and refine implementation strategies. These approaches aim to build sustainable capacity for evidence-informed teaching, ensuring that both new and experienced teachers are equipped to use research effectively in their classrooms.

16.7 Conclusion

To sum up, incorporating research evidence into teaching practices is a feasible and effective approach to enhancing teaching quality and improving students' learning outcomes within the Chinese teaching context. Adequate training potentially fosters teachers' positive views on using research evidence while building their fundamental skills of evidence-use, particularly understanding research findings. The use-of-research-evidence intervention is cost-effective and brings minimal disruption to school routines. The implementation quality could be understood at three levels involving school, teacher individual, and teacher interactive, highlighting the iterative process teachers undergo when adapting research findings to practice. Finally, the use of research evidence has the potential for long-term impact, creating a positive feedback loop that encourages the ongoing adoption of rigorous research evidence in teaching practices.

References

Abramczyk, A., & Jurkowski, S. (2020). Cooperative learning as an evidence-based teaching strategy: What teachers know, believe, and how they use it. *Journal of Education for Teaching*, *46*(3), 296–308.

Abu-Odah, H., Said, N. B., Nair, S. C., Allsop, M. J., Currow, D. C., Salah, M. S., Abu Hamad, B., Elessi, K., Alkhatib, A., ElMokhallalati, Y., Bayuo, J., & AlKhaldi, M. (2022). Identifying barriers and facilitators of translating research evidence into clinical practice: A systematic review of reviews. *Health & Social Care in the Community*, *30*(6), e3265-e3276.

Allen, R. (2013). *Evidence-based practice: why number-crunching tells only part of the story.* IOE London Blog. ioelondonblog.wordpress.com/2013/03/14/evidence-based-practice-why-number-crunching-tells-only-part-of-thestory.

Anders, J., Foliano, F., Bursnall, M., Dorsett, R., Hudson, N., Runge, J., & Speckesser, S. (2022). The Effect of Embedding Formative Assessment on Pupil Attainment. *Journal of Research on Educational Effectiveness*, *15*(4), 748–779. https://doi.org/10.1080/19345747.2021.2018746

Andreassen, R., & Bråten, I. (2011). Implementation and effects of explicit reading comprehension instruction in fifth-grade classrooms. *Learning and Instruction*, *21*(4), 520–537. https://doi.org/10.1016/j.learninstruc.2010.08.003

Arnold Ventures. (n.d.). *Social Programs That Work.* Evidence Based Programs. http://www.evidencebasedprograms.org

Bao, Y. (2024). 如何有效开展小学语文大单元教学 [How to effectively carry out large-unit Chinese teaching in primary schools]. *读写算 [Reading, Writing, and Calculation]*, *35*, 34–36.

Barnett, A. G., Van Der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: what it is and how to deal with it. *International journal of epidemiology*, *34*(1), 215–220. https://doi.org/10.1093/ije/dyh299

Bartholo, T., & Koslinski, M. C. (2020). First two years at school: evidence-based policy for early childhood education in Brazil. In S. Gorard (Ed.), *Getting Evidence into Education* (pp. 111–118). Routledge.

Benner, G. J., Filderman, M. J., Barnard-Brak, L., Pennefather, J., Smith, J. L. M., & Strycker, L. A. (2023). Evidence of efficacy of the Integrated Literacy Study Group professional learning program to enhance reading instruction for students with emotional and behavioral disorders. *Psychology in the Schools*, *60*(1), 182-198. https://doi.org/10.1002/pits.22773

Bellibaş, M. Ş., Gümüş, S., & Liu, Y. (2021). Does school leadership matter for teachers' classroom practice? The influence of instructional leadership and distributed leadership on instructional quality. *School Effectiveness and School Improvement*, *32*(3), 387–412.

BERA-RSA. 2014. The Role of Research in Teacher Education: Reviewing the Evidence. Interim Report of the BERA-RSA Inquiry. London: British Educational Research Association.

Bergeron, P. J., & Rivard, L. (2017). How to engage in pseudoscience with real data: A criticism of John Hattie's arguments in visible learning from the perspective of a statistician. *McGill Journal of Education*, *52*(1), 237–246. https://doi.org/10.7202/1040816ar

Berk, R., & Freedman, D. (2001). Statistical assumptions as empirical commitments. University of California, Berkeley. http://www.stat.berkeley.edu/~census/berk2.pdf

Bero, L. A., Grilli, R., Grimshaw, J. M., Harvey, E., Oxman, A. D., & Thomson, M. A. (1998). Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. *BMJ*, *317*(7156), 465–468. https://doi.org/10.1136/bmj.317.7156.465

Beycioglu, K., Ozer, N., & Ugurlu, C. T. (2010). Teachers' views on educational research. *Teaching and teacher education*, *26*(4), 1088–1093. https://doi.org/10.1016/j.tate.2009.11.004

Beyer, S., Ahmad, C., Mayer, T., Sander, P., & Kerres, M. (2025, March). Improving Pre-Service Teachers' Use of Research to Support Technology-Enhanced Practices. In *Society for Information Technology & Teacher Education International Conference* (pp. 2668–2673). Association for the Advancement of Computing in Education (AACE).

Bierman, K. L., Sanford DeRousie, R. M., Heinrichs, B., Domitrovich, C. E., Greenberg, M. T., & Gill, S. (2013). Sustaining high-quality teaching and evidence-based curricula: Follow-up assessment of teachers in the REDI project. *Early Education & Development*, *24*(8), 1194–1213.

Blömeke, S., Olsen, R. V., & Suhl, U. (2016). Relation of student achievement to the quality of their teachers and instructional quality. In T. Nilsen & J. E. Gustafsson (Eds.), *Teacher quality, instructional quality and student outcomes: Relationships across countries, cohorts and time* (pp. 21–50). Springer Nature.

Booher, L., Nadelson, L. S., & Nadelson, S. G. (2020). What about research and evidence? Teachers' perceptions and uses of education research to inform STEM teaching. *The Journal of Educational Research*, *113*(3), 213–225. https://doi.org/10.1080/00220671.2020.1782811

Borman, G. D., Choi, Y., & Hall, G. J. (2021). The impacts of a brief middle-school self-affirmation intervention help propel African American and Latino students through high school. *Journal of Educational Psychology*, *113*(3), 605.

Bowen, S., & Zwi, A. B. (2005). Pathways to "evidence-informed" policy and practice: a framework for action. *PLoS medicine*, *2*(7), e166. https://doi.org/10.1371/journal.pmed.0020166

Brown, C., & Greany, T. (2018). The evidence-informed school system in England: Where should school leaders be focusing their efforts?. *Leadership and Policy in Schools*, *17*(1), 115–137. https://doi.org/10.1080/15700763.2016.1270330

Brown, C., & Zhang, D. (2017). How can school leaders establish evidence-informed schools: An analysis of the effectiveness of potential school policy levers. *Educational Management Administration & Leadership*, *45*(3), 382-401. https://doi.org/10.1177/1741143215617946

Brown, C., Schildkamp, K., & Hubers, M. D. (2017). Combining the best of two worlds: A conceptual proposal for evidence-informed school improvement. *Educational research*, *59*(2), 154–172. https://doi.org/10.1080/00131881.2017.1304327

Burn, K., Conway, R., Edwards, A., & Harries, E. (2021). The role of school-based research champions in a school–university partnership. *British Educational Research Journal*, *47*(3), 616–633. https://doi.org/10.1002/berj.3675

Cabinet Office & HM Treasury. (2018, January 29). *The What Works Network: Five years on.* https://www.gov.uk/government/publications/the-what-works-network-five-years-on

Cahill, S. M., Egan, B. E., Wallingford, M., Huber-Lee, C., & Dess-McGuire, M. (2015). Results of a School-Based Evidence-Based Practice Initiative. *The American Journal of Occupational Therapy*, *69*(2), Article 6902220010. https://doi.org/10.5014/ajot.2015.014597

Cain, T. (2015a). Teachers' engagement with published research: Addressing the knowledge problem. *The Curriculum Journal*, *26*(3), 488–509. https://doi.org/10.1080/09585176.2015.1020820

Cain, T. (2015b). Teachers' engagement with research texts: Beyond instrumental, conceptual or strategic use. *Journal of Education for Teaching*, *41*(5), 478–492. https://doi.org/10.1080/02607476.2015.1105536

Campbell, C., & Levin, B. (2012). Developing knowledge mobilisation to challenge educational disadvantage and inform effective practices in England (Discussion Paper). *London: Educational Endowment Foundation*.

Carrier, N. (2017). How educational ideas catch on: the promotion of popular education innovations and the role of evidence. *Educational Research*, *59*(2), 228–240. https://doi.org/10.1080/00131881.2017.1310418

Clark, C., & De Zoysa, S. (2011). Mapping the Interrelationships of Reading Enjoyment, Attitudes, Behaviour and Attainment: An Exploratory Investigation. *National Literacy Trust*. https://eric.ed.gov/?id=ED541404

Clarke, B., Smolkowski, K., Baker, S. K., Fien, H., Doabler, C. T., & Chard, D. J. (2011). The impact of a comprehensive Tier I core kindergarten program on the achievement of students at risk in mathematics. *The Elementary School Journal*, *111*(4), 561–584. https://doi.org/10.1086/659033

Coburn, C. E., & Talbert, J. E. (2006). Conceptions of evidence use in school districts: Mapping the terrain. *American Journal of Education*, *112*(4), 469–495. https://doi.org/10.1086/505056

Cohen, J. (2013). Statistical power analysis for the behavioral sciences (2nd edition). Routledge.

Coldwell, M., Greany, T., Higgins, S., Brown, C., Maxwell, B., Stiell, B., Stoll, L., Willis, B., & Burns, H. (2017). *Evidence-informed teaching: an evaluation of progress in England. Research Report*. Department for Education.

CONSORT Group. (n.d.). *CONSORT flow diagram.* CONSORT Statement. Retrieved January 2023, from http://www.consort-statement.org/consort-statement/flow-diagram

Cooper, A., Levin, B., & Campbell, C. (2009). The growing (but still limited) importance of evidence in education policy and practice. *Journal of educational change*, *10*(2), 159–171. https://doi.org/10.1007/s10833-009-9107-0

Cuban, L. (2013). Why so many structural changes in schools and so little reform in teaching practice?. *Journal of Educational Administration*, *51*(2), 109-–125. https://doi.org/10.1108/09578231311304661

Cui, Y., Lei, H., & Zhou, W. (2018). Changes in school curriculum administration in China. *ECNU Review of Education*, *1*(1), 34–57. https://doi.org/10.30926/ecnuroe2018010103

Davies, P. (1999). What is Evidence-based Education? *British Journal of Educational Studies*, *47*(2), 108–121. https://doi.org/10.1111/1467-8527.00106

Davis, K. S. (2003). "Change is hard": What science teachers are telling us about reform and teacher learning of innovative practices. *Science education*, *87*(1), 3–30. https://doi.org/10.1002/sce.10037

DfE. (2018). Get school-led support to use evidence to improve teaching practice. GOV.UK. https://www.gov.uk/guidance/get-school-led-support-to-use-evidence-to-improve-teaching-practice

Diery, A., Knogler, M., & Seidel, T. (2021). Supporting evidence-based practice through teacher education: A profile analysis of teacher educators' perceived challenges and possible

solutions. *International Journal of Educational Research Open*, *2*, 100056. https://doi.org/10.1016/j.ijedro.2021.100056

Doabler, C. T., Nelson, N. J., Kosty, D. B., Fien, H., Baker, S. K., Smolkowski, K., & Clarke, B. (2014). Examining teachers' use of evidence-based practices during core mathematics instruction. *Assessment for Effective Intervention*, *39*(2), 99–111. https://doi.org/10.1177/1534508413511848

Domitrovich, C. E., Bradshaw, C. P., Poduska, J. M., Hoagwood, K., Buckley, J. A., Olin, S., Romanelli, L. H., Leaf, P. J., Greenberg, M. T., & Ialongo, N. S. (2008). Maximizing the implementation quality of evidence-based preventive interventions in schools: A conceptual framework. *Advances in school mental health promotion*, *1*(3), 6–28. https://doi.org/10.1080/1754730X.2008.9715730

Drabick, D. A., & Goldfried, M. R. (2000). Training the scientist–practitioner for the 21st century: Putting the bloom back on the rose. *Journal of Clinical Psychology*, *56*(3), 327–340. https://doi.org/10.1002/(SICI)1097-4679(200003)56:3<327::AID-JCLP9>3.0.CO;2-Y

Drake, E. K., Aos, S., & Miller, M. G. (2009). Evidence-based public policy options to reduce crime and criminal justice costs: Implications in Washington State. *Victims and offenders*, *4*(2), 170–196. https://doi.org/10.1080/15564880802612615

Dweck, C. (2017). Mindset: The new psychology of success (Y. Chu, Trans.). Jiangxi People's Publishing House. (Original work published 2006)

Dweck, C. (2014). *The power of believing that you can improve* [Video]. TED Conferences. https://www.ted.com/talks/carol_dweck_the_power_of_believing_that_you_can_improve?language=en

Edmonds, M. S. (2007). *Utilizing implementation data to explain outcomes within a theory-driven evaluation model.* The University of Texas at Austin.

Education Endowment Foundation. (2023, February 15). *A new evidence review from the Education Endowment Foundation explores effective professional development for teachers*... [Social media post]. X [formerly Twitter]. https://x.com/educendowfoundn/status/1625819639635759106

Education Endowment Foundation. (n.d.). *Projects and evaluation.* https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects

Education Endowment Foundation. (2021, December 15). *EEF announces new work to facilitate and support evidence-based practice.* https://educationendowmentfoundation.org.uk/news/eef-announces-new-work-to-facilitate-and-support-evidence-based-practice

Erkan, C. (2021). *A Systematic Review of Evidence on the Best Ways to Disseminate Research Evidence to Teachers.* In S. Riddle & P. Bhatia (Eds.), *Imagining Better Education 2020* (pp. 54–74). Durham, England.

Erkan, C. (2023). *Investigation into how best to disseminate research evidence to teachers* (Doctoral dissertation). Durham University

Evans, D. (2003). Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *Journal of Clinical Nursing*, *12*(1), 77–84. https://doi.org/10.1046/j.1365-2702.2003.00662.x

Everton, T., Galton, M., & Pell, T. (2002). Educational research and the teacher. *Research Papers in Education Policy and Practice*, *17*(4), 373–401. https://doi.org/10.1080/0267152022000031388

Fair, F., Haas, L. E., Gardosik, C., Johnson, D. D., Price, D. P., & Leipnik, O. (2015). Socrates in the schools from Scotland to Texas: Replicating a study on the effects of a Philosophy for Children program. *Journal of Philosophy in Schools*, *2*(1), 18–37.

Farley-Ripple, E. N. (2012). Research use in school district central office decision making: A case study. *Educational Management Administration & Leadership*, *40*(6), 786–806. https://doi.org/10.1177/1741143212456912

Farley-Ripple, E., May, H., Karpyn, A., Tilley, K., & McDonough, K. (2018). Rethinking connections between research and practice in education: A conceptual framework. *Educational Researcher*, *47*(4), 235–245. https://doi.org/10.3102/0013189X18761042

Farrington, D. P., MacKenzie, D. L., Sherman, L. W., & Welsh, B. C. (Eds.). (2003). *Evidence-based crime prevention*. Routledge.

Fixsen, D. L., Blase, K. A., Naoom, S. F., & Wallace, F. (2009). Core implementation components. *Research on social work practice*, *19*(5), 531–540. https://doi.org/10.1177/1049731509335549

Flynn, N. (2019). Facilitating evidence-informed practice. *Teacher Development*, *23*(1), 64–82. https://doi.org/10.1080/13664530.2018.1505649

Foliano, F., Rolfe, H., Buzzeo, J., Runge, J., & Wilkinson, D. (2019). *Changing mindsets: Effectiveness trial –Evaluation report.* Education Endowment Foundation. https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/changing-mindsets

Forman, S. G., Olin, S. S., Hoagwood, K. E., Crowe, M., & Saka, N. (2009). Evidence-based interventions in schools: Developers' views of implementation barriers and facilitators. *School Mental Health*, *1*, 26–36. https://doi.org/10.1007/s12310-008-9002-5

Frey, N., Fisher, D., & Everlove, S. (2009). *Productive group work: How to engage students, build teamwork, and promote understanding.* ASCD.

General Teaching Council for England et al. (2012). Surveys of Teachers, 2004-2010. [data collection]. UK Data Service. SN: 6890, DOI: http://doi.org/10.5255/UKDA-SN-6890-1

Gersten, R. (2001). Sorting out the roles of research in the improvement of practice. *Learning Disabilities Research & Practice*, *16*(1), 45–50. https://doi.org/10.1111/0938-8982.00005

Gleeson, J., Cutler, B., Rickinson, M., Walsh, L., Ehrich, J., Cirkony, C., & Salisbury, M. (2023). School educators' engagement with research: an Australian Rasch validation study. *Educational Assessment, Evaluation and Accountability*, *35*(2), 281–307. https://doi.org/10.1007/s11092-023-09404-7

Gorard, S. (2013). *Research design: Creating robust approaches for the social sciences.* SAGE Publications.

Gorard, S. (2020). Handling missing data in numeric analyses. *International journal of social research methodology*, *23*(6), 651–660. https://doi.org/10.1080/13645579.2020.1729974

Gorard, S. (2021). *How to make sense of statistics.* SAGE Publications Ltd.

Gorard, S. (2023). A sensitivity test does everything that a significance test does, and better. *Journal of Research and Method in Education, 13*(2), 50–56. https://doi.org/10.9790/7388-1302045056

Gorard, S. (2024). Judging the relative trustworthiness of research results: How to do it and why it matters. *Review of Education*, *12*(1), e3448. https://doi.org/10.1002/rev3.3448

Gorard, S., & Chen, W. (2025). What is the evidence on research-informed education? In D. Wyse, V. Baumfield, N. Mockler, & M. Reardon (Eds.), *The BERA–SAGE handbook of research-informed education practice and policy.* SAGE Ltd.

Gorard, S., & Gorard, J. (2016). What to do instead of significance testing? Calculating the 'number of counterfactual cases needed to disturb a finding'. *International Journal of Social Research Methodology*, *19*(4), 481–490. https://doi.org/10.1080/13645579.2015.1091235

Gorard, S., See, B. H., & Siddiqui, N. (2020). What is the evidence on the best way to get evidence into use in education? *Review of Education*, *8(2)*, 570–610. https://doi.org/10.1002/rev3.3200

Gorard, S., Siddiqui, N., & See, B. H. (2015). Philosophy for Children: Evaluation Report and Executive Summary. *Education Endowment Foundation*. https://eric.ed.gov/?id=ED581147

Gough, D., Maidment, C., & Sharples, J. (2022). Enabling knowledge brokerage intermediaries to be evidence-informed. *Evidence & Policy*, *18*(4), 746–760. https://doi.org/10.1332/174426421X16353477842207

Goyer, J. P., Garcia, J., Purdie-Vaughns, V., Binning, K. R., Cook, J. E., Reeves, S. L., Apfel, N., Taborsky-Barba, S., Sherman, D. K., & Cohen, G. L. (2017). Self-affirmation facilitates minority middle schoolers' progress along college trajectories. *Proceedings of the National Academy of Sciences*, *114*(29), 7594–7599.

Graves, S., & Moore, A. (2018). How do you know what works, works for you? An investigation into the attitudes of senior leaders to using research evidence to inform teaching and learning in schools. *School Leadership & Management*, *38*(3), 259–277.

Greenwood, C. R., Tapia, Y., Abbott, M., & Walton, C. (2003). A building-based case study of evidence-based literacy practices: implementation, reading behavior, and growth in reading fluency, K—4. *The Journal of Special Education*, *37*(2), 95–110. https://doi.org/10.1177/00224669030370020401

Groß Ophoff, J., Brown, C., & Helm, C. (2023). Do pupils at research-informed schools actually perform better? Findings from a study at English schools. *Frontiers in Education*, *7*, Article 1011241. https://doi.org/10.3389/feduc.2022.1011241

Gu, L., & Wang, J. (2006). School-based research and professional learning: An innovative model to promote teacher professional development in China. *Teaching Education*, *17*(1), 59–73. https://doi.org/10.1080/10476210500528079

Guyatt, G. H., Oxman, A. D., Kunz, R., Vist, G. E., Falck-Ytter, Y., & Schünemann, H. J. (2008). What is "quality of evidence" and why is it important to clinicians?. *BMJ*, *336*(7651), 995–998. https://doi.org/10.1136/bmj.39490.551019.BE

Han, W., Chen, J., Yang, F., & Xia, L. (2024). 苏洵小学"心理戏剧"校本课程构建与实践 [The construction and practice of a school-based "psychodrama" curriculum at Suxun Primary School]. *教育科学论坛 [Forum on Educational Science]*, *31*, 66–68.

Hargreaves, D. H. (1996). Teaching as a research-based profession: possibilities and prospects. *Teacher training agency* annual *lecture*.

Hattie, J. (1992). Measuring the effects of schooling. *Australian Journal of Education*, *36*(1), 5–13. https://doi.org/10.1177/000494419203600102

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hedges, H. (2012). Teachers' funds of knowledge: A challenge to evidence-based practice. *Teachers and Teaching, 18*(1), 7–24. https://doi.org/10.1080/13540602.2011.622548

Hemsley-Brown, J., & Sharp, C. (2003). The use of research to improve professional practice: A systematic review of the literature. *Oxford Review of Education*, *29*(4), 449–471. https://doi.org/10.1080/0305498032000153025

Higgins, S. (2020). The development and worldwide impact of the Teaching and Learning Toolkit. In S. Gorard (Ed.), *Getting evidence into education* (pp. 69–83). Routledge.

Higgins, S., Katsipataki, M., Villanueva-Aguilera, A. B., Coleman, R., Henderson, P., Major, L. E., Coe, R., & Mason, D. (2016). *The Sutton Trust-Education Endowment Foundation Teaching and Learning Toolkit.* Durham University. https://durham-repository.worktribe.com/output/1606338

Hinton, C., & Fischer, K. W. (2008). Research schools: Grounding research in educational practice. *Mind, Brain, and Education*, *2*(4), 157–160. https://doi.org/10.1111/j.1751-228X.2008.00048.x

Hollands, F., Pan, Y., & Escueta, M. (2019). What is the potential for applying cost-utility analysis to facilitate evidence-based decision making in schools?. *Educational Researcher*, *48*(5), 287–295. https://doi.org/10.3102/0013189X19852101

Holloway, J., & Larsen Hedegaard, M. L. (2023). Democracy and teachers: The im/possibilities for pluralisation in evidence-based practice. *Journal of Education Policy*, *38*(3), 432–451. https://doi.org/10.1080/02680939.2021.2014571

Hood, P. (2003). *Scientific research and evidence-based practice* (pp. 1–51). WestEd.

Ilhan, N. (2021). The Effect of Research Evidence-Based Teaching Practices in Science Classrooms on Student Teachers' Attitudes towards Educational Research. *Journal of Science Learning*, *4*(4), 316–326.

Ilic, D., & Maloney, S. (2014). Methods of teaching medical trainees evidence-based medicine: a systematic review. *Medical Education*, *48*(2), 124–135. https://doi.org/10.1111/medu.12288

Innvaer, S., Vist, G., Trommald, M., & Oxman, A. (2002). Health policy-makers' perceptions of their use of evidence: a systematic review. *Journal of health services research & policy*, *7*(4), 239–244.

Institute of Education Sciences. (n.d.). *What Works Clearinghouse.* U.S. Department of Education. Retrieved October 2022, from https://ies.ed.gov/ncee/wwc/

Ion, G., Kowalczuk-Walędziak, M., & Brown, C. (2024). Unlocking the potential of research-informed practice: Insights into benefits, challenges, and significance among teachers in Catalonia, Poland, and England. *European Journal of Education*, *59*(3), e12639. https://doi.org/10.1111/ejed.12639

Ion, G., Suárez, C. I., & Díaz-Vicario, A. (2021). Evidence-informed educational practice in Catalan education: From public agenda to teachers' practice. *Center for Educational Policy Studies Journal*, *11*(2), 37–57. https://doi.org/10.26529/cepsj.1062

Jacob, B. (2017). When evidence is not enough: Findings from a randomized evaluation of Evidence-Based Literacy Instruction (EBLI). *Labour Economics*, *45*, 5–16. https://doi.org/10.1016/j.labeco.2016.09.006

James, M. (2013). New (or not new) directions in evidence-based practice in education. *British Educational Research Association*.

Jay, T., Willis, B., Thomas, P., Taylor, R., Moore, N., Burnett, C., Merchant, G., & Stevens, A. (2017). Dialogic teaching: Evaluation report and executive summary.

Jiang, Y., Zhang, J., & Xin, T. (2019). Toward education quality improvement in China: A brief overview of the national assessment of education quality. *Journal of Educational and Behavioral Statistics*, *44*(6), 733–751. https://doi.org/10.3102/1076998618809677

Joram, E., Gabriele, A. J., & Walton, K. (2020). What influences teachers'"buy-in" of research? Teachers' beliefs about the applicability of educational research to their practice. *Teaching and Teacher Education*, *88*(102980), 1–12.

Ke, Z., Yin, H., & Huang, S. (2019). Teacher participation in school-based professional development in China: does it matter for teacher efficacy and teaching strategies?. *Teachers and Teaching*, *25*(7), 821–836. https://doi.org/10.1080/13540602.2019.1662777

Kennedy, M. J., Hirsch, S. E., Rodgers, W. J., Bruce, A., & Lloyd, J. W. (2017). Supporting high school teachers' implementation of evidence-based classroom management practices. *Teaching and Teacher Education*, *63*, 47–57. https://doi.org/10.1016/j.tate.2016.12.009

Kien, C., Grillich, L., Nussbaumer-Streit, B., & Schoberberger, R. (2018). Pathways leading to success and non-success: a process evaluation of a cluster randomized physical activity health promotion program applying fuzzy-set qualitative comparative analysis. *BMC Public Health*, *18*, 1–14. https://doi.org/10.1186/s12889-018-6284-x

Kim, J. S., Burkhauser, M. A., Quinn, D. M., Guryan, J., Kingston, H. C., & Aleman, K. (2017). Effectiveness of structured teacher adaptations to an evidence-based summer literacy program. *Reading Research Quarterly*, *52*(4), 443–467. https://doi.org/10.1002/rrq.178

Konstantinou, I. (2022). Establishing a centre for evidence-informed practice within a school: Lessons from the Research and Evidence Centre. *Review of Education*, *10*(1), e3324. https://doi.org/10.1002/rev3.3324

Kostoulas, A., Babić, S., Glettler, C., Karner, A., Mercer, S., & Seidl, E. (2019). Lost in research: educators' attitudes towards research and professional development. *Teacher Development*, *23*(3), 307–324. https://doi.org/10.1080/13664530.2019.1614655

Kretlow, A. G., & Bartholomew, C. C. (2010). Using coaching to improve the fidelity of evidence-based practices: A review of studies. *Teacher Education and Special Education, 33*(4), 279–299. https://doi.org/10.1177/0888406410371643

Kretlow, A. G., Cooke, N. L., & Wood, C. L. (2012). Using in-service and coaching to increase teachers' accurate use of research-based strategies. *Remedial and Special Education*, *33*(6), 348–361. https://doi.org/10.1177/0741932510395397

Kung, S. H., & Teo, T. (2022, under review). Effects of peer-assisted learning strategies on reading literacy of fourth-grade Chinese-speaking students. *Asia-Pacific Journal of Teacher Education*.

Langer, L., Tripney, J. S., & Gough, D. (2016). The science of using science: Researching the use of research evidence in decision-making.

Lee, H., Chung, H. Q., Zhang, Y., Abedi, J., & Warschauer, M. (2020). The effectiveness and features of formative assessment in US K-12 education: A systematic review. *Applied Measurement in Education*, *33*(2), 124–140. https://doi.org/10.1080/08957347.2020.1732383

Lewin, C., Morris, S., Ainsworth, S., Gellen, S., & Wicker, K. (2024). *Peer assisted learning strategies–UK (PALS-UK): A whole class reading approach–evaluation report.* Education Endowment Foundation.

Lingard, B. (2013). The impact of research on education policy in an era of evidence-based policy. *Critical Studies in Education*, *54*(2), 113–131. https://doi.org/10.1080/17508487.2013.781515

Lord, P., Dirie, A., Kettlewell, K., & Styles, B. (2021). *Evaluation of philosophy for children: An effectiveness trial.* Education Endowment Foundation.

Lord, P., Rabiasz, A., & Styles, B. (2017a). *'Literacy Octopus' Dissemination Trial: Evaluation Report and Executive Summary.* Education Endowment Foundation.

Lord, P., Rabiasz, A., Roy, P., Harland, J., Styles, B., & Fowler, K. (2017b*). Evidence-Based Literacy Support: The 'Literacy Octopus' Trial.* Education Endowment Foundation. https://eric.ed.gov/?id=ED581248

Lord, P., Sims, D., White, R., & Roy, P. (2017c). *Evidence for the Frontline: Evaluation Report and Executive Summary.* Education Endowment Foundation. https://eric.ed.gov/?id=ED581162

Lyon, A. R., & Bruns, E. J. (2019). From evidence to impact: Joining our best school mental health practices with our best implementation strategies. *School mental health*, *11*, 106–114. https://doi.org/10.1007/s12310-018-09306-w

Lysenko, L. V., Abrami, P. C., Bernard, R. M., Dagenais, C., & Janosz, M. (2014). Educational research in educational practice: Predictors of use. *Canadian Journal of Education/Revue canadienne de l'éducation*, *37*(2), 1–26.

Mayer, D. (2021). The connections and disconnections between teacher education policy and research: Reframing evidence. *Oxford Review of Education*, *47*(1), 120–134.

McGoey, K. E., Rispoli, K. M., Venesky, L. G., Schaffner, K. F., McGuirk, L., & Marshall, S. (2014). A preliminary investigation into teacher perceptions of the barriers to behavior intervention implementation. *Journal of Applied School Psychology*, *30*(4), 375–390. https://doi.org/10.1080/15377903.2014.950441

McIntosh, K., Horner, R. H., & Sugai, G. (2009). Sustainability of Systems-Level Evidence-Based Practices in Schools: Current Knowledge and Future Directions. In W. Sailor, G. Dunlap, G. Sugai, & R. Horner (Eds.), *Handbook of Positive Behavior Support* (pp. 327–352). Springer US. https://doi.org/10.1007/978-0-387-09632-2_14

Meyers, D. C., Durlak, J. A., & Wandersman, A. (2012). The quality implementation framework: a synthesis of critical steps in the implementation process. *American Journal of Community Psychology*, *50*, 462–480. https://doi.org/10.1007/s10464-012-9522-x

Mills, M., Mockler, N., Stacey, M., & Taylor, B. (2021). Teachers' orientations to educational research and data in England and Australia: Implications for teacher professionalism. *Teaching Education*, *32*(1), 77–98. https://doi.org/10.1080/10476210.2020.1843617

Mincu, M. E. (2015). Teacher quality and school improvement: what is the role of research? *Oxford Review of Education*, *41*(2), 253–269. https://doi.org/10.1080/03054985.2015.1023013

Ministry of Education of the People's Republic of China. (2019, November 7). *教育部关于加强新时代教育科学研究工作的意见[Opinions on strengthening educational research in the new era].* Retrieved November 16, 2024, from http://www.moe.gov.cn/srcsite/A02/s7049/201911/t20191107_407332.html

Ministry of Education of the People's Republic of China. (2022). *义务教育语文课程标准（2022 年版）[Chinese curriculum standards for compulsory education: Chinese literacy (2022 edition)].* Beijing: Beijing Normal University Press.

Moir, T. (2018). Why Is Implementation Science Important for Intervention Design and Evaluation Within Educational Settings? *Frontiers in Education*, 3, Article 61. https://doi.org/10.3389/feduc.2018.00061

Mou, X. (2024). Research on Teaching Strategies of Group Reading in Primary Chinese based on Learning Task Group. *Journal of Education and Educational Research*, *8*(2), 310–314.

Nelson, J., & Campbell, C. (2017). Evidence-informed practice in education: Meanings and applications. *Educational Research*, *59*(2), 127–135. https://doi.org/10.1080/00131881.2017.1314115

OECD. (2008). *TALIS 2008 Database*. https://www.oecd.org/en/data/datasets/talis-2008-database.html

OECD. (2013). *TALIS 2013 Database*. https://www.oecd.org/en/data/datasets/talis-2013-database.html

OECD. (2018). *TALIS 2018 Database*. https://www.oecd.org/en/data/datasets/talis-2018-database.html

OECD. (n.d.). Teaching and Learning International Survey (TALIS). https://www.oecd.org/education/talis/

Owen, K. L., Watkins, R. C., & Hughes, J. C. (2022). From evidence-informed to evidence-based: An evidence building framework for education. *Review of Education*, *10*(1), e3342. https://doi.org/10.1002/rev3.3342

Oxford University. (n.d.). *International Database of Education Systematic Reviews (IDESR).* https://idesr.org

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, *372*.

Palmer, J. A. (1999). Research matters: A call for the application of empirical evidence to the task of improving the quality and impact of environmental education. *Cambridge Journal of Education*, *29*(3), 379-395. https://doi.org/10.1080/0305764990290308

Pappas, I. O., & Woodside, A. G. (2021). Fuzzy-set Qualitative Comparative Analysis (fsQCA): Guidelines for research practice in Information Systems and marketing. *International journal of information management*, *58*, 102310. https://doi.org/10.1016/j.ijinfomgt.2021.102310

Pawson, R., Boaz, A., Grayson, L., Long, A., & Barnes, C. (2003). *Types and quality of knowledge in social care.* Social Care Institute for Excellence.

Pegram, J., Watkins, R. C., Hoerger, M., & Hughes, J. C. (2022). Assessing the range and evidence-base of interventions in a cluster of schools. *Review of Education*, *10*(1), e3336. https://doi.org/10.1002/rev3.3336

Pellegrini, M., & Vivanet, G. (2021). Evidence-based policies in education: Initiatives and challenges in Europe. *ECNU Review of Education*, *4*(1), 25–45. https://doi.org/10.1177/2096531120924670

Petscher, Y. (2010). A meta-analysis of the relationship between student attitudes towards reading and achievement in reading. *Journal of research in reading*, *33*(4), 335–355. https://doi.org/10.1111/j.1467-9817.2009.01418.x

Poet, H., Rudd, P., & Kelly, J. (2010). Survey of teachers 2010: Support to improve teaching practice.

Procter, R. (2013). Teachers and reseach: What they value and what they do. https://uobrep.openrepository.com/handle/10547/335954

Procter, R. (2015). Teachers and school research practices: The gaps between the values and practices of teachers. *Journal of Education for Teaching*, *41*(5), 464–477. https://doi.org/10.1080/02607476.2015.1105535

Promising Practices Network. (n.d.). *Promising Practices Network.* RAND Corporation. http://www.promisingpractices.net

Qian, H., Walker, A., & Chen, S. (2024). The 'Double-Reduction'education policy in China: Three prevailing narratives. *Journal of Education Policy*, *39*(4), 602–621. https://doi.org/10.1080/02680939.2023.2222381

Ragin, C. C. (2009). *Redesigning social inquiry: Fuzzy sets and beyond*. University of Chicago Press.

Reinke, W. M., Stormont, M., Herman, K. C., & Newcomer, L. (2014). Using coaching to support teacher implementation of classroom-based interventions. *Journal of Behavioral Education*, *23*, 150–167. https://doi.org/10.1007/s10864-013-9186-0

Ren, P., & Li, X. (2021). Evidence-based Education: Origin, Dilemma, Framework and Implement Suggestions. *China Educational Technology*, *12*, 33–39.

Research Excellence Framework. (n.d.). *Research Excellence Framework 2029.* UK Research and Innovation. https://2029.ref.ac.uk

Rickinson, M., Cirkony, C., Walsh, L., Gleeson, J., Cutler, B., & Salisbury, M. (2022). A framework for understanding the quality of evidence use in education. *Educational Research*, *64*(2), 133–158. https://doi.org/10.1080/00131881.2022.2054452

Rickinson, M., Cirkony, C., Walsh, L., Gleeson, J., Salisbury, M., & Boaz, A. (2021). Insights from a cross-sector review on how to conceptualise the quality of use of research evidence. *Humanities and Social Sciences Communications*, *8*(1), 1–12. https://doi.org/10.1057/s41599-021-00821-x

Ridge, B. L., & Lavigne, A. L. (2020). Improving instructional practice through peer observation and feedback: A review of the literature. *Education Policy Analysis Archives*, *28*, 61–61.

Rienzo, C., Rolfe, H., & Wilkinson, D. (2015). Changing Mindsets: Evaluation Report and Executive Summary. *Education Endowment Foundation*. https://eric.ed.gov/?id=ED581132

Rogers, E. M., Singhal, A., & Quinlan, M. M. (2014). Diffusion of innovations. In An integrated approach to communication theory and research (pp. 432–448). Routledge.

Rojas-Andrade, R., & Bahamondes, L. L. (2019). Is Implementation Fidelity Important? A Systematic Review on School-Based Mental Health Programs. Contemporary School Psychology, 23(4), 339–350. https://doi.org/10.1007/s40688-018-0175-0

Rose, J., Thomas, S., Zhang, L., Edwards, A., Augero, A., & Roney, P. (2017). *Research Learning Communities: Evaluation Report and Executive Summary.* Education Endowment Foundation. https://eric.ed.gov/?id=ED581267

Rycroft-Malone, J., Seers, K., Titchen, A., Harvey, G., Kitson, A., & McCormack, B. (2004). What counts as evidence in evidence-based practice?. *Journal of advanced nursing*, *47*(1), 81–90. https://doi.org/10.1111/j.1365-2648.2004.03068.x

Sackett, D. L., Rosenberg, W. M., Gray, J. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. *BMJ*, *312*(7023), 71–72. https://doi.org/10.1136/bmj.312.7023.71

Sargent, T. C. (2015). Professional learning communities and the diffusion of pedagogical innovation in the Chinese education system. *Comparative Education Review*, *59*(1), 102–132.

Sato, M., & Loewen, S. (2019). Do teachers care about research? The research–pedagogy dialogue. *ELT Journal*, *73*(1), 1-10. https://doi.org/10.1093/elt/ccy048

Sato, M., & Loewen, S. (2022). The research–practice dialogue in second language learning and teaching: Past, present, and future. *The Modern Language Journal*, *106*(3), 509–527. https://doi.org/10.1111/modl.12791

Scheeler, M. C., Budin, S., & Markelz, A. (2016). The role of teacher preparation in promoting evidence-based practice in schools. *Learning Disabilities: A Contemporary Journal*, *14*(2), 171–187.

Schmidt, K., Edelsbrunner, P. A., Rosman, T., Cramer, C., & Merk, S. (2023). When perceived informativity is not enough. How teachers perceive and interpret statistical results of educational research. *Teaching and Teacher Education*, *130*, 104134. https://doi.org/10.1016/j.tate.2023.104134

Research Schools Network. (n.d.). Research Schools Network. https://researchschool.org.uk

Scott, S., & McNeish, D. (2013). *School leadership evidence review: Using research evidence to support school improvement.* National Centre for Social Research for CUBeC. https://www.dmss.co.uk/pdfs/evidencereview3.pdf

Sedgwick, P. (2013). What is intention to treat analysis?. *BMJ*, *346*. https://doi.org/10.1136/bmj.f3662

See, B. H. (2020). Why is it difficult to get evidence into use? In S. Gorard (Ed.), *Getting evidence into education* (pp. 84–99). Routledge.

See, B. H., & Gorard, S. (2020). Effective classroom instructions for primary literacy: A critical review of the causal evidence. *International Journal of Educational Research*, *102*, 101577. https://doi.org/10.1016/j.ijer.2020.101577

See, B. H., Gorard, S., & Siddiqui, N. (2016). Teachers' use of research evidence in practice: A pilot study of feedback to enhance learning. *Educational Research*, *58*(1), 56–72. https://doi.org/10.1080/00131881.2015.1117798

See, B. H., Gorard, S., Lu, B., Dong, L., & Siddiqui, N. (2022). Is technology always helpful?: A critical review of the impact on learning outcomes of education technology in supporting formative assessment in schools. *Research Papers in Education*, *37*(6), 1064–1096. https://doi.org/10.1080/02671522.2021.1907778

See, B. H., Morris, R., Gorard, S., Siddiqui, N., Easterbrook, M. J., Nieuwenhuis, M., Fox, K., Harris, P. R., & Banerjee, R. (2022). A conceptual replication study of a self-affirmation intervention to improve the academic achievement of low-income pupils in England. *Educational Research and Evaluation*, *27*(1–2), 83–116. https://doi.org/10.1080/13803611.2021.2022317

Sengalrayan, B., & Harvey, B. (2025). Engaging knowledge users in Canadian knowledge mobilisation research: a scoping review of research in education. *Evidence & Policy*, *21*(1), 129–151.

Siegel, S. T. (2024, April 19). Don't shoot the education mythbuster – Evidence-informed sustainability education (EvSusEd). Hypotheses.org. https://doi.org/10.58079/12m7v

Sing, S., Adamson, L., Konstantinou, I., Graham, S., Stenhouse, R., Wicking, A., Wadey, P., Rowden-Knowles, S., Taylor, E., & Hemming, H. (2018). Practical approaches to embedding research in schools: Key learning and reflections for the Research in Schools Learning Community. *Impact: Journal of the Chartered College of Teaching*, (4), 77–80.

Slater, H., Davies, N. M., & Burgess, S. (2012). Do teachers matter? Measuring the variation in teacher effectiveness in England. *Oxford Bulletin of Economics and Statistics*, *74*(5), 629–645. https://doi.org/10.1111/j.1468-0084.2011.00666.x

Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, *31*(7), 15–21. https://doi.org/10.3102/0013189X031007015

Slavin, R. E. (2008). Perspectives on evidence-based research in education—What works? Issues in synthesizing educational program evaluations. *Educational Researcher*, *37*(1), 5–14. https://doi.org/10.3102/0013189X08314117

Song, H., Xu, S., & Wen, N. (2023). 循证教学实践模型[Model of evidence-based teaching practice]. *湖北教育(教育教学)[Hubei Education (Teaching and Learning)]*, (02), 29–31.

Spencer, T. D., Detrich, R., & Slocum, T. A. (2012). Evidence-based practice: A framework for making effective decisions. *Education and treatment of children*, *35*(2), 127–151.

Van Staden, A. (2011). Put reading first: Positive effects of direct instruction and scaffolding for ESL learners struggling with reading. *Perspectives in Education*, *29*(4), 10–21.

Steiner-Khamsi, G., Faul, M. V., Baek, C., Hopkins, A. N., & Iwabuchi, K. (2024). Improving the use of evidence for education policy, planning and implementation: Strategic review.

Strunk, K. O., Marsh, J. A., Bush-Mecenas, S. C., & Duque, M. R. (2016). The best laid plans: An examination of school plan quality and implementation in a school improvement initiative. *Educational Administration Quarterly*, *52*(2), 259–309. https://doi.org/10.1177/0013161X15616864

Sutton Trust. (2019, January 9). Two thirds of school heads have cut teachers to save money. Sutton Trust. Retrieved January 2025 from https://www.suttontrust.com/newsarchive/two-thirds-of-school-heads-have-cut-teachers-to-save-money/

Taylor, B. M., Pearson, P. D., Peterson, D. S., & Rodriguez, M. C. (2005). The CIERA school change framework: An evidence-based approach to professional development and school reading improvement. *Reading research quarterly*, *40*(1), 40–69. https://doi.org/10.1598/RRQ.40.1.3

Taylor, C. (2002). *The RCBN consultation exercise: Stakeholder report.* University College Cardiff, School of Social Sciences.

Test, D. W., Kemp-Inman, A., Diegelmann, K., Hitt, S. B., & Bethune, L. (2015). Are online sources for identifying evidence-based practices trustworthy? An evaluation. *Exceptional Children*, *82*(1), 58–80. https://doi.org/10.1177/0014402915585477

Thomas, G. (2021). Experiment's persistent failure in education inquiry, and why it keeps failing. *British Educational Research Journal*, *47*(3), 501–519. https://doi.org/10.1002/berj.3660

Thomm, E., Sälzer, C., Prenzel, M., & Bauer, J. (2021). Predictors of teachers' appreciation of evidence-based practice and educational research findings. *Zeitschrift für Pädagogische Psychologie*. https://doi.org/10.1024/1010-0652/a000301

Toomey, A. H. (2016). What happens at the gap between knowledge and practice? Spaces of encounter and misencounter between environmental scientists and local people. *Ecology and Society*, *21*(2). https://www.jstor.org/stable/26270380

Topping, K., Duran, D., & Van Keer, H. (2015). *Using peer tutoring to improve reading skills: a practical guide for teachers.* Routledge.

Topping, K. J., Thurston, A., McGavock, K., & Conlin, N. (2012). Outcomes and process in reading tutoring. *Educational Research*, *54*(3), 239–258. https://doi.org/10.1080/00131881.2012.710086

Torgerson, C., Bell, K., Coleman, E., Elliott, L., Fairhurst, C., Gascoine, L., Hewitt, C., & Torgerson, D. (2018). *Tutor Trust: Affordable primary tuition. Evaluation report and executive summary.* [No known commissioning body]. https://durham-repository.worktribe.com/output/1605472

Torgerson, C., Brooks, G., Gascoine, L., & Higgins, S. (2019). Phonics: Reading policy and the evidence of effectiveness from a systematic 'tertiary' review. *Research Papers in Education*, *34*(2), 208–238. https://doi.org/10.1080/02671522.2017.1420816

Tortorelli, L. S., & Bruner, L. (2022). The Word Nerds project: Findings from a research–practice partnership focused on spelling instruction. *Journal of Research in Reading*, *45*(3), 385–405. https://doi.org/10.1111/1467-9817.12394

Trinder, L. (2000). A critical appraisal of evidence-based practice. *Evidence-based practice: A critical appraisal*, 212–241.

UNESCO. (2017). Unpacking sustainable development goal 4: Education 2030. https://unesdoc.unesco.org/ark:/48223/pf0000246300

United States. Department of Education. Office of the Under Secretary, United States. Office of Elementary, & Secondary Education. (2002). *No Child Left Behind: A Desktop Reference*. US Department of Education, Office of the Under Secretary, Office of Elementary and Secondary Education.

van Schaik, P., Volman, M., Admiraal, W., & Schenke, W. (2018). Barriers and conditions for teachers' utilisation of academic knowledge. *International Journal of Educational Research*, *90*, 50–63. https://doi.org/10.1016/j.ijer.2018.05.003

Vanderlinde, R., & van Braak, J. (2010). The gap between educational research and practice: Views of teachers, school leaders, intermediaries and researchers. *British educational research journal*, *36*(2), 299–316.

Vardy, E. J., Al Otaiba, S., Breadmore, H. L., Kung, S. H., Pétursdóttir, A. L., Zaru, M. W., & McMaster, K. L. (2022). Teacher–researcher partnership in the translation and implementing of PALS (Peer-Assisted Learning Strategies): An international perspective. *Journal of Research in Reading*, *45*(3), 517–526. https://doi.org/10.1111/1467-9817.12404

Wadhwa, M., Zheng, J., & Cook, T. D. (2024). How consistent are meanings of "evidence-based"? A comparative review of 12 clearinghouses that rate the effectiveness of educational programs. *Review of Educational Research*, *94*(1), 3–32. https://doi.org/10.3102/00346543231152262

Walker, M., Nelson, J., & Bradshaw, S. (2019). Teachers' Engagement with Research: What Do We Know? A Research Briefing. *Education Endowment Foundation*. https://eric.ed.gov/?id=ED620325

Wang, J., & Jin, X. (2011). Translating evidence into policy in China: opportunities and challenges. *Frontiers of Medicine*, *5*, 315–320. https://doi.org/10.1007/s11684-011-0146-6

Wang, J., Herman, J. L., Epstein, S., Leon, S., La Torre, D., & Bozeman, V. (2020*). Literacy Design Collaborative 2018–2019 evaluation report* (CRESST Report 867, Revised). National Center for Research on Evaluation, Standards, and Student Testing (CRESST). https://eric.ed.gov/?id=ED606356

Wang, K. (2021). 平等对话,主动探究——小学语文对话式教学模式分析[Equal dialogue and active inquiry: An analysis of dialogic teaching in primary Chinese literacy classes]. *新课程教学(电子版)[New Curriculum Teaching (Electronic Edition)]*, *24*, 47–48.

Wang, S., & Ji, P. (2024). 基于循证教育的中小学教师教育科研素养的培育路径[Cultivating teachers' educational research literacy in primary and secondary schools based on evidence-informed education]. *教育科学论坛[Forum on Educational Science]*, *17*, 47–49.

Warren, E., Melendez-Torres, G. J., & Bonell, C. (2022). Using fuzzy-set qualitative comparative analysis to explore causal pathways to reduced bullying in a whole-school intervention in a randomized controlled trial. *Journal of school violence*, *21*(4), 381–396. https://doi.org/10.1080/15388220.2022.2105856

Weimer, M. (2008). Positioning scholarly work on teaching and learning. *International Journal for the Scholarship of Teaching and Learning*, *2*(1), 4. https://doi.org/10.20429/ijsotl.2008.020104

Wen, R. (2024, January 5). Speech at the 20th-anniversary celebration of the Institute of Chinese Language Education and the symposium on 'Basic Issues in Chinese Language Education'. Peking University. Retrieved Nov 17, 2024, from http://www.jyb.cn/rmtzcg/xwy/wzxw/202401/t20240108_2111141624.html

White, H. (2019). The twenty-first century experimenting society: the four waves of the evidence revolution. *Palgrave Communications*, *5*(1), 1–7. https://doi.org/10.1057/s41599-019-0253-6

White, H. (2020). The global evidence architecture in health and education: A comparative scorecard. In S. Gorard (Ed.), *Getting evidence into education* (pp. 20–33). Routledge.

Wiggins, M., Jerrim, J., Tripney, J., Khatwa, M., & Gough, D. (2019). *The RISE project: Evidence-informed school improvement. Evaluation report.* Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/RISE_Report_final.pdf

Wilcox, G., Chatlos, S. B., McClure, E., Flowers, J., & Makarenko, E. (2023). De-implementation: A missing piece in bridging the research to practice gap in school psychology. *Contemporary School Psychology*, *27*(3), 400–410. https://doi.org/10.1007/s40688-021-00399-w

Wiliam, D. (2013). The importance of teaching. *EXCELLENCE AND EQUITY*, 50.

Wiliam, D. (2021). *Embedded formative assessment* (2nd ed., S. Wang, Trans.). Jiangsu Phoenix Science and Technology Publishing House. (Original work published 2011)

Williams, D., & Coles, L. (2007). Teachers' approaches to finding and using research evidence: An information literacy perspective. *Educational research*, *49*(2), 185–206. https://doi.org/10.1080/00131880701369719

Willingham, D. T., & Daniel, D. (2021). Making education research relevant: How researchers can give teachers more choices. Education Next, 21(2), 28–33. https://www. educationnext. org/making-education-research-relevant-how-researchers-can-give-teachers-more-choices/, last accessed, 6(06).

Wongdaeng, M., & Higgins, S. (2023). Effectiveness of metacognitive interventions in tertiary EFL contexts: Evidence from a systematic review and meta-analysis. *Innovation in Language Learning and Teaching*, *17*(4), 795–811. https://doi.org/10.1080/17501229.2022.2146122

Woodard, R., & Kline, S. (2015). Moving beyond compliance: Promoting research-based professional discretion in the implementation of the common core state standards in English language arts. Mid-Western Educational Researcher, 27(3), 4. https://scholarworks.bgsu.edu/mwer/vol27/iss3/4

Wu, Z., Spreckelsen, T. F., & Cohen, G. L. (2021). A meta-analysis of the effect of values affirmation on academic achievement. *Journal of Social Issues*, *77*(3), 702–750.

Xu, Y. (2009). School-based Teacher Development through a School–University Collaborative project: a case study of a recent initiative in China. *Journal of Curriculum Studies*, *41*(1), 49–66. https://doi.org/10.1080/00220270802546740

Xuan, Q., Cheung, A., & Sun, D. (2022). The effectiveness of formative assessment for enhancing reading achievement in K-12 classrooms: A meta-analysis. *Frontiers in psychology*, *13*, 990196.

Xue, E., & Li, J. (2023). What is the value essence of "double reduction"(Shuang Jian) policy in China? A policy narrative perspective. *Educational Philosophy and Theory*, *55*(7), 787–796. https://doi.org/10.1080/00131857.2022.2040481

Yao, J. (2021). Professor Slavin and the Development of Evidence-Based Education in China. *Best Evidence in Chinese Education*, *8*(1), 1021–1028. https://papers.ssrn.com/abstract=3878530

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., Tipton, E., Schneider, B., Hulleman, C. S., Hinojosa, C. P., Paunesku, D., Romero, C., Flint, K., Roberts, A., Trott, J., Iachan, R., Buontempo, J., Yang, S. M., Carvalho, C. M., … Dweck, C. S. (2019). A national experiment

reveals where a Growth mindset improves achievement. *Nature*, *573*(7774), 364–369. https://doi.org/10.1038/s41586-019-1466-y

Yeager, D. S., Romero, C., Paunesku, D., Hulleman, C. S., Schneider, B., Hinojosa, C., Lee, H. Y., O'Brien, J., Flint, K., Roberts, A., Trott, J., Greene, D., Walton, G. M., & Dweck, C. S. (2016). Using design thinking to improve psychological interventions: The case of the Growth mindset during the transition to high school. *Journal of educational psychology*, *108*(3), 374.https://doi.org/10.1037/edu0000098

Zhang, F. (2024). 基于大单元的小学语文整体性教学策略[Teaching strategies for primary Chinese language education based on the large unit approach]. *读写算[Reading, Writing, and Calculation]*, *34*, 64–66.

Zhao, L., & Chen, X. (2024, March). Facilitating Evidence-Based Practice in Classroom Teaching: Framework Development and a Demonstrative Case Study. In *Society for Information Technology & Teacher Education International Conference* (pp. 1663–1669). Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/primary/p/224189/

Zhou, & Jing, B. (2024). 循证教育视域下校本培训模式构建的依据与路径[The basis and path of constructing a school-based training model under the perspective of evidence-based education]. *辽宁教育[Liaoning Education]*, *08*, 10–15.

Zhou, S. (2021). The Role of China's Teaching Research System in Promoting Evidence-Based Reform in Education: A Case Study of Jiangsu Province. *Best Evidence in Chinese Education*, *9*(2), 1227–1241.

Appendices

Appendix A. Search terms for the structured review

ProQuest	abstract('research-bas*' OR 'research bas*' OR 'evidence-bas*' OR 'evidence-led*' OR finding* OR 'evidence-inform*' OR 'evidence led*' OR 'evidence bas* OR (('research knowledge' OR 'research evidence' OR 'finding*') (use OR usage OR utilis* OR appl* OR translat* OR judg* OR evaluat* OR practi* OR 'learn from' OR disseminat* OR transf* OR access*)) OR ('evidence into practice' OR 'research into practice')) AND abstract(school OR classroom OR teach* OR educat* OR student* OR pupil* OR test OR exam OR 'school outcome*' OR attainment OR proficiency OR assessment OR literacy OR qualification OR achievement OR 'school success*' OR 'performance acquisition' OR progress OR cognitive OR non-cognitive OR attitude OR 'self reliance' OR 'long-term development' OR autonomy OR 'self regulat*' OR 'self efficacy' OR motivat* OR enjoy* OR joy* OR happ* OR independen* OR expectation OR aspiration OR attendance) AND (interact* OR behav* OR relat* OR communicat* OR stimul*) AND (((judg* OR evaluat* OR assess*) ('research quality' OR 'study quality' OR 'research evidence')) OR rigo* OR robust* OR vigo* OR forceful OR ((high-quality) ('research evidence' OR finding* OR study OR evidence))) AND abstract(primary OR elementary OR secondary OR middle OR 'compulsory education' OR 'Key stag*' OR Grad*) AND abstract(literacy OR read* OR grammar OR spell* OR phonetic OR fluency OR comprehension OR vocabulary) AND (evaluat* OR interven* OR trial OR experiment OR cause* OR effect* OR determinant OR 'regression discontinuity' OR 'randomi* control' OR 'controlled trial' OR impact) NOT subject(medic* OR clinic* OR health* OR special education OR therapy) NOT subject(sixth form OR high school OR higher education OR university OR college OR post compulsory OR post-compulsory OR Key Stage 5 OR faculty OR postgraduate OR undergraduate OR Master) NOT subject(evidence-based learning OR research-based learning)
Scopus	TITLE-ABS((education OR school* OR college* OR classroom* OR teach* OR learn* OR educator* OR student* OR children OR pupil* OR achiev* OR attainment OR exam* OR attendance) AND (((‘Research knowledge’ OR ‘Research findings’ OR ‘Research evidence’) W/2 (base OR based OR inform OR informed OR use OR used OR using OR utili* OR utiliz* OR uptak* OR transf* OR translat* OR modif* OR engag* OR summar* OR access* OR disseminat* OR mobilis* OR mobiliz* OR implement* OR present* OR bring* OR push* OR shar*)) OR (research W/1 (base OR based OR inform OR informed OR use OR used OR using OR utilis* OR utili* OR transf* OR translat* OR disseminat* OR advocat* OR mobilis* OR mobili*)) OR (‘evidence into practice’ OR ‘research into practice’)) AND (facilitat* OR improv* OR promot* OR increas* OR develop* OR support* OR effective* OR better OR best OR strateg* OR pathway* OR intervention) AND (literacy OR read* OR grammar OR spell* OR phonetic OR fluency OR comprehension OR vocabulary) AND (impact OR evaluat* OR interven* OR trial OR experiment OR cause* OR effect* OR determinant) AND NOT (health* OR dent* OR medic* OR nurses OR nursing OR clinic*))

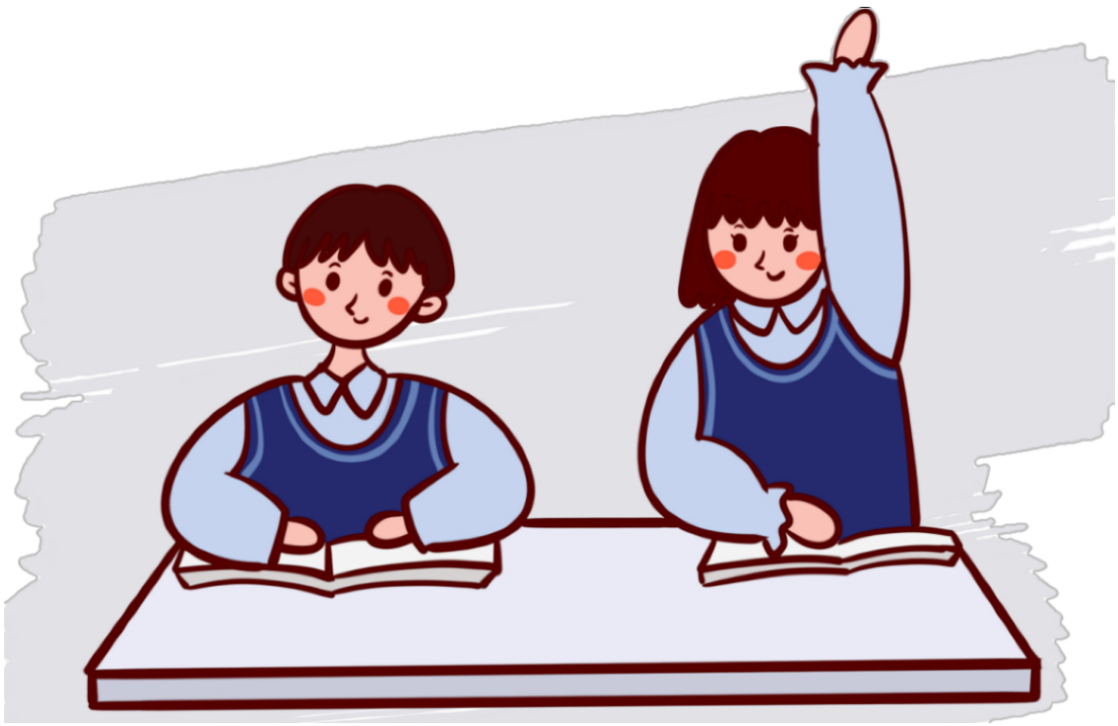
Web of Science	evidence base* or 'evidence led*' or 'evidence inform*' or 'research base*' or 'research led*' or 'research inform*' or 'evidence into practice' or 'evidence into use' or 'research into practice' or 'research into use' (Abstract) and use OR usage OR utilis* OR appl* OR translat* OR judg* OR evaluat* OR practi* OR 'learn from' OR disseminat* OR transf* OR access* (All Fields) and school OR classroom OR teach* OR educat* OR student* OR pupil* (Topic) and literacy OR read* OR grammar OR spell* OR phonetic OR fluency OR comprehension OR vocabulary (Abstract) and test OR exam OR 'school outcome*' OR attainment OR proficiency OR assessment OR literacy OR qualification OR achievement OR 'school success*' OR 'performance acquisition' OR progress OR cognitive OR non-cognitive OR attitude OR 'self reliance' OR autonomy OR 'self regulat*' OR 'self efficacy' OR motivat* OR enjoy* OR joy* OR happ* OR independen* OR expectation OR aspiration OR attendance (All Fields) and evaluat* OR interven* OR trial OR experiment OR cause* OR effect* OR determinant OR 'regression discontinuity' OR 'randomi* control' OR 'controlled trial' OR impact (All Fields) not medic* OR clinic* OR health* OR special education OR therapy OR nurs* (Topic)
British Education Index (EBSCOhost)	AB((('Research knowledge' OR 'research findings' OR 'research evidence') N2 (use OR used OR using OR utilis* OR utiliz* OR uptak* OR transf* OR translat* OR modif* OR engag* OR summar* OR access* OR disseminat* OR mobilis* OR mobiliz* OR implement* OR present* OR bring* OR push* OR shar* OR mobilis*)) OR (research N/1 (use OR used OR using OR utilis* OR utiliz* OR transf* OR translat* OR disseminat* OR mobilis*)) OR ('evidence into practice' OR 'research into practice')) AND (school OR classroom OR teach* OR educat* OR student* OR pupil) AND (literacy OR read* OR grammar OR spell* OR phonetic OR fluency OR comprehension OR vocabulary
ERIC	(('Research knowledge' OR 'research findings' OR 'research evidence') N/2 (base or based or inform or informed or use OR used OR using OR utilis* OR utiliz* OR uptak* OR transf* OR translat* OR modif* OR engag* OR summar* OR access* OR disseminat* OR mobilis* OR mobiliz* OR implement* OR present* OR bring* OR push* OR shar*)) OR (research N/1 (base or based or inform or informed or use OR used OR using OR utilis* OR utiliz* OR transf* OR translat* OR disseminat* or advocat* Or mobilis* OR mobiliz*)) OR ('evidence into practice' OR 'research into practice') AND (school OR classroom OR teach* OR educat* OR student* OR pupil*) AND (literacy OR read* OR grammar OR spell* OR phonetic OR fluency OR comprehension OR vocabulary)AND (facilitat* OR improv* OR promot* OR increas* OR develop* OR support* OR effective* OR better OR best OR strateg* OR pathway* OR intervention) AND (impact OR evaluat* OR interven* OR trial OR experiment OR cause* OR effect* OR determinant) AND NOT (health* OR dent* OR medic* OR nurses OR nursing OR clinic*)

Appendix B. Use-of-research-evidence intervention booklet

Here is the English version of the *Use-of-Research-Evidence* intervention booklet.

Durham University
Evidence Centre for Education

Research evidence-led strategies to improve literacy attainment in the primary age



How does research evidence work in assisting teaching?



Research evidence in education usually refers to findings from research with a formal research question, and a systematic design, conducted by research experts in education. Such research often evaluates the effectiveness of certain programmes or approaches in education.



The idea of transforming research evidence for use in practice originated in the field of medicine and extended to education, around the world. For example, the What Works Clearinghouse (WWC) from the U.S. and the EEF Teaching Toolkit from the UK call for the use of standards to evaluate research quality, and integrate research evidence into education practice. They are looking for promising, high-quality findings that provide strong evidence for educational decision-making.



When bridging research evidence and teaching practice, the ideal would be to select programmes that have the greatest impact, especially for lower attainers, backed by the strongest evidence, at the lowest cost. In reality, it is still up to schools and teachers to decide what programmes to use and what to achieve based on student learning. What we want in this project is that schools choose one or two of the programs provided in this booklet, depending on their school situation, and use them in their daily teaching in the coming semester.

Examples promising evidence for teaching reading literacy in the primary stage

The table below presents a summary of five possible approaches to improving overall attainment in literacy in primary school. These have been selected because they have been tested, are feasible, and show promise of enhancing pupil outcomes, especially for low-achieving or low-income pupils. Although there are many other promising teaching approaches, the following approaches are relatively easy for teachers to implement correctly with the guidance in this menu.

Strategies	Suggested grades	Estimated impact	Evidence strength
Embedding formative assessment	1-6	0.20	3
Peer tutoring	3-6	0.30	3
Dialogic teaching	4-6	0.10	3
Self-affirmation	6	0.07	2
Growth mindset	5-6	0.10	2

The table lists the following attributes of each approach:

Suggested grades. Teaching approaches are more appropriate for some year groups. This menu suggests year groups for teachers to take into consideration when they choose the approaches.
Estimate impact. The likely impact or benefit is presented as an estimated "effect" size based on the strongest prior research evaluations. The larger the "effect" size is, the more observable difference the intervention makes on reading attainment.

Evidence strength. The padlock rating from 0 to 4 gives an indication of how strong the prior evidence is for each approach. A high rating would indicate the evidence is as strong as could be imagined in real life. This rating is independent of the likely impact of any intervention.

Implementation Guidance

The following sections introduce each strategy in detail, involving:

- Description of the teaching strategy
- How the strategy works
- Supplementary materials for implementation
- Prior research evidence



Embedded formative assessment

Embedding Formative Assessment (EFA) aims to embed the use of effective formative assessment strategies in daily teaching. Promising evidence about the effectiveness of EFA on reading achievement (ES 0.1 to 0.3).

What is embedded formative assessment?

Formative assessment is defined as: to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited. Embedding formative assessment is a teaching strategy within the ongoing teaching and learning activities which serves teachers' decision-making for the next stage instruction. The goal is to provide immediate feedback to help students understand what they already know, what they need to know and what they need to do to go further.

The focus of EFA is on decisions. Rather than a focus on data-driven decision making, the emphasis is on decision-driven data collection. This is important because a focus on data-driven decision making emphasizes the collection of data first without any view about the claims they might support, so the claims are therefore accorded secondary importance. By starting with the decisions that need to be made, only data that support the inferences that are sought need be collected.

How does it work?

Five key strategies were recommended:

- Clarifying, sharing, and understanding learning intentions and success criteria—deals with the joint responsibility of teachers, the learners themselves, and their peers to break this down into a number of criteria for success.
- Finding out where learners are in their learning, once he is clear about the learning intentions (this sequence is deliberate—until you know what you want your students to learn, you do not know what evidence to collect).
- Providing feedback to the students that tells them not only where they are but also what steps they need to take to move their learning forward.
- Peer assessment can play in supporting student learning and also makes clear that the purpose of peer assessment within a formative assessment framework is not to judge the work of a peer so much as to improve it.
- Ultimate goal is always to produce independent learners.

<table>
<tr><th></th><th>Where the learner is going</th><th>Where the learner is now</th><th>How to get there</th></tr>
<tr><td>Teacher</td><td rowspan="2">Clarifying, sharing, and understanding learning intentions and success criteria</td><td>Engineering effective discussions, tasks, and activities that elicit evidence of learning</td><td>Providing feedback that moves learning forward</td></tr>
<tr><td>Peer</td><td colspan="2">Activating students as learning resources for one another</td></tr>
</table>

Learner		Activating students as owners of their own learning

Helpful resources

Embedded formative assessment supplementary materials

Wiliam, D. (2011). *Embedded formative assessment*. Solution tree press.

Prior evidence

A recent structured reviews combining 48 studies suggests that the average effect size of formative assessment on reading achievement is +0.18 (Xuan, 2022). Another review suggests larger effect size +0.33 on reading in the U.S. In 2018, a large-scale randomised control trial illustrates that using Dylan Wiliam's embedded formative assessment could benefit students' GCSE achievement with an effect size of +0.11.

Report	Year group	Duration	Evidence strength	Estimated impact
Speckesser, S., et al. (2018)	Year 10, U.K.	2 years	4	ES = +0.11

Lee, H., Chung, H. Q., Zhang, Y., Abedi, J., & Warschauer, M. (2020). The effectiveness features of formative assessment in US K-12 education: A systematic review. *Applied Measurement in Education*, *33*(2), 124-140.

Speckesser, S., Runge, J., Foliano, F., Bursnall, M., Hudson-Sharp, N., Rolfe, H., & Anders, J. (2018). Embedding Formative Assessment: Evaluation report and executive summary.

Xuan, Q., Cheung, A., & Sun, D. (2022). The effectiveness of formative assessment for. enhancing reading achievement in K-12 classrooms: A meta-analysis. *Frontiers in Psychology*, *13*, 990196.

Peer tutoring

Students are paired or clustered into groups based on previous reading attainment. Students switch the role of peer tutor or peer tutee so that they could get opportunities to teach each other. Some evidence of positive impacts (ES 0.1 to 0.4), especially for low-achievement pupils.

What is peer tutoring?

Peer tutoring is a teaching method that promotes collaboration among students during lessons. Students in the same class are divided into groups and help each other to learn by teaching and learning together. There are two types of same-age peer tutoring:

- fixed peer tutoring: the same student is always the tutor
- reciprocal peer tutoring: students take turns as the tutor and the tutee

Apart from same-age peer tutoring, cross-age peer tutoring also has promising effects for reading proficiency. However, cross-age peer tutoring involves reorganise classes, which is less feasible.

How does it work?

Same-age peer tutoring usually takes place in the same class group, so it does not require any special organisational actions. Teachers assign students into groups according to their judgement and implement the strategy based on the course content, typically 1-4 times a week, with each session lasting 20-40 minutes. Research evidence suggests using it three times a week to make observable improvement.

In fixed peer tutoring, students maintain their roles as either the tutor or the tutee for the entire programme. Pairs are matched on the basis of previous reading or mathematics attainment. The class is ordered from highest to lowest attainment in reading. Students above the mid-point become tutors, and those below become tutees. The highest-achieving tutor is matched with the tutee just a bit below average, while the weakest tutee is helped by an average tutor.

In reciprocal tutoring, students may also be paired according to their reading ability or personal characteristics. However, the roles of tutor and tutee are not permanent. Students switch between these roles in the midway through each session and offer each other coaching and feedback during problem-solving activities. The teacher's decision to assign pairs and skills is based on their understanding of students' needs and abilities, and these pairs are regularly reassigned.

The first stage of peer tutoring focuses on comprehension of continuous prose. At the second stage, there is even more emphasis on comprehension, focusing on longer and harder books and a wider range of reading materials beyond books.

Helpful resources

Peer tutoring supporting document 2.0

Prior evidence

Many studies have evaluated the effectiveness of peer tutoring, however, the majority of the research has been limited to case studies or lacks balanced comparison groups. The following two studies were randomised control trials.

More than three thousand pupils from 125 primary schools in Scotland participated in the first study which implemented both cross-age and same-age peer tutoring over two years. The second study directly evaluates the effects of peer-assisted learning, randomising 118 students from Grade 4. Both studies have promising results in improving reading achievement of low-achieving students.

Report	Year group	Duration	Evidence strength	Estimated impact
Topping et al. (2012)	Year 4 and Year 6	30-minute sessions, 1 or 3 times per week, for a duration of 15 weeks	3	+0.31 First Year of Implementation +0.34 Second Year of Implementation
Kung & Teo (2022)	Year 4 (Taiwan)	40-minute sessions, twice a week, for a duration of 15 weeks	2	+0.07 Vocabulary +0.11 Reading Comprehension

References

Kung, S.-H., & Teo, T. (2022 - under review). Effects of peer-assisted learning strategies on reading literacy of fourth-grade Chinese-speaking students. Asia-Pacific Journal of Teacher Education.

Topping, K. J., Thurston, A., McGavock, K., & Conlin, N. (2012). Outcomes and process in reading tutoring. Educational Research, 54(3), 239-258.

Dialogic teaching

Dialogic teaching encourages discussion and provides support for argumentation/critical thinking (one or more lessons per week, one term). Some evidence of positive impacts especially for weaker students (ES 0 to 0.2).

What is dialogic teaching?

Classroom talk and discussions can provide useful opportunities for learning and the social-emotional development of children. There are several approaches developed for improving the quality of classroom interaction with a specific focus on the quality of dialogue between pupils. Dialogic teaching harnesses the power of talk to engage interest, stimulate thinking, advance understanding, expand ideas, and build and evaluate arguments, empowering students for lifelong learning and critical thinking.

How does it work?

Philosophy for Children (P4C) is one of the successful models of dialogic teaching. P4C is a whole-class intervention which aims to stimulate classroom dialogue in response to children's own questions about shared stories, films, and other stimuli. Teachers do not need to be familiar with philosophical literature in order to facilitate philosophical inquiry. What is more important is an attitude of genuine (philosophical) curiosity, open-mindedness, and an ability to inquire with students. For observable changes in children's classroom talk and behaviour, the programme recommends regular implementation of one-hour sessions per week for a duration of 3 months.

P4C sessions involve a step-by-step discussion approach in a classroom setting. This requires both the teacher and the children to sit in a circle and agree on specific principles of turn-taking and communication. The teacher introduces a topic through a stimulus such as a picture, quotation, news item, short story, or poem, and the children are encouraged to generate questions based on it. These questions are necessarily not factual but can lead to reflective and conceptual discussions. The children are prompted to use dialogic language and argumentation to justify their opinions, ask questions, critique, and elaborate on their responses and perceptions.

Once the children have developed their questions, a voting process is used to select the most popular question, which becomes the focus of subsequent discussion. During the discussion, the children take turns speaking, deploying simple argumentation terms and structures that they have been taught beforehand to support their points. The teacher only intervenes if necessary to ensure that the dialogue leads to a useful and productive discussion. To be aware, teachers should avoid leading to the 'only' 'correct' conclusion. Teachers may use multi-choice questions to open students' views. After the discussion, the children are asked to reflect on their final thoughts and whether they have changed their views because of the discussion. The session is then reviewed by all participants in the circle of inquiry to evaluate the quality of the discussion.

Helpful resources

http://www.p4c.org.cn/cn/index.aspx
P4C supporting documents

Prior evidence

This implementation guidance selected four randomised control trials which evaluated the effectiveness of dialogic teaching. The first three evaluated the impact of P4C while the last one evaluated another dialogic teaching model. The research findings suggest that dialogic teaching has a positive impact.

Report	Year group	Duration	Evidence strength	Estimated impact
Gorard et al. (2018)	5, 6	One year	3 🔒	Reading +0.12 Writing +0.03
Lord et al. (2021)	6	Two years	3 🔒	Reading +0.01
Fair et al. (2015)	7, 8 (U.S.)	24 weeks	2 🔒	Reading +0.21
Jay et al. (2017)	5	20 weeks	3 🔒	CAT4 Reading +0.12 KS2 reading +0.03

References

Fair, F., Haas, L. E., Gardosik, C., Johnson, D. D., Price, D. P., & Leipnik, O. (2015). Socrates in the schools from Scotland to Texas: Replicating a study on the effects of a Philosophy for Children program. Journal of Philosophy in Schools, 2(1).

Gorard, S., Siddiqui, N., & See, B. H. (2018). Philosophy for Children: Evaluation report and executive summary. EEF: London UK

Lord, P., Dirie, A., Kettlewell, K., & Styles, B. (2021). Evaluation of philosophy for children: an effectiveness trial. EEF: London, UK.

Jay, T., Willis, B., Thomas, P., Taylor, R., Moore, N., Burnett, C., ... & Stevens, A. (2017). Dialogic teaching: Evaluation report and executive summary.

Writing about values

Pupils facing a high-stakes test write two or three brief essays about their values. Reasonable evidence of impact for disadvantaged pupils (ES 0.05 to 0.1).

What is writing about value?

Writing about values is an approach to overcome negative stereotypes experienced by some disadvantaged students. Some of them show low academic expectation to defend and protect themselves from stereotypes and being devalued. The intervention can broaden one's perspective beyond feeling like a negatively stereotyped student in school. Writing positive statements about the values when stereotype threats are most salient can give individuals a positive sense of value, negating harmful feelings, and fostering academic learning. The most notable benefit of this intervention is its cost-effectiveness and simplicity in implementation.

How does it work?

The intervention involves two or three writing activities, each lasting 10 to 15 minutes. The writing exercises are presented in booklets, placed in named envelopes and distributed to pupils individually. All completed booklets are placed in the named envelopes. They are collected and kept by the teachers.

Before the writing task, teachers give a short, structured introduction to the task explaining that the task is a writing exercise focusing on pupils' own thoughts and ideas and that there are no right or wrong answers to the task. Teachers explain that it is the process of doing the activity that matters. Their writing will not be read or marked, and no feedback will be provided. This is to encourage free and open reflection.

An important element of this intervention is that neither the teachers nor the pupils should know precisely what the intervention is, as there is some evidence that knowledge about the aim of the intervention can reduce its efficacy. Therefore, all efforts should be made to conceal the primary intention of the writing exercises.

The first writing exercise is supposed to be the most impactful because if a student performs/behaves better as a result of the first activity, their self-confidence may improve, and their teacher may have higher expectations of them, for example. This could lead to better performance, and the process perpetuates itself. It is therefore important that students complete the first writing exercise. The second (and third) exercises are meant to provide a boost to this process.

To mask the real intention of the exercise, these writing tasks can be delivered during Chinese lessons as part of the regular Chinese class or during PSHE lesson as a whole-class activity.

Helpful resources

WAV Writing Tasks
WAV Check List

Prior evidence

Self-affirmation interventions have been widely evaluated, mostly in the U.S. The evidence suggests that self-affirmation activities can have positive and long-term results improving academic achievement, especially for ethnic minority groups (e.g. Wu et al., 2021). Some

studies have also suggested that these effects persist through secondary school (Borman et al. 2021) and right up to university (Goyer et al., 2017).

Two studies conducted in England have shown that the self-affirmation approach is also effective in raising the attainment of low-SES pupils (See et al., 2022). See et al.'s (2022) involving c.11,000 pupils aged 14 to 16 in England also showed positive effects on pupils' Attainment 8 scores (ES = +0.05) for those eligible for free school meals.

.

Report	Age Group	Duration	Evidence strength	Estimated impact
See et al. (2022)	14-16	One year	4	+0.05

References

Borman, G. D., Choi, Y., & Hall, G. J. (2021). The impacts of a brief middle-school self-affirmation intervention help propel African American and Latino students through high school. Journal of Educational Psychology, 113(3), 605–620. https://doi.org/10.1037/edu0000570

Goyer, J. P., Garcia, J., Purdie-Vaughns, V., Binning, K. R., Cook, J. E., Reeves, S. L., Apfel, N.,Taborsky-Barba, S., Sherman, D. K., & Cohen, G. L. (2017). Self-affirmation facilitates minority middle schoolers' progress along college trajectories. Proceedings of the National Academy of Sciences, 114(29), 7594–7599. https://doi.org/10.1073/pnas.1617923114

See, B.H., Morris, M., Gorard, S., Siddiqui, N., Easterbrook, M.J., Nieuwenhuis, M., Fox, K., Harris, P.R. & Banerjee, R. (2022) A conceptual replication study of a self-affirmation intervention to improve the academic achievement of low-income pupils in England, *Educational Research and Evaluation*, DOI:10.1080/13803611.2021.2022317

Wu, Z., Spreckelsen, T. F., & Cohen, G. L. (2021). A meta-analysis of the effect of values affirmation on academic achievement. Journal of Social Issues, 77(3), 702–750. https://doi.org/10.1111/josi.12415

Growth mindset
Students participate in one reading workshop about the idea that the brain is malleable like a muscle to get basic information about growth mindset. Then they are invited to discuss how they could implement growth mindset in their daily life. Teachers are also encouraged to embed growth mindset in daily teaching and conduct relevant activities. Some evidence of positive impacts especially for low-achieving students (ES 0 to 0.2).

What is Growth mindset?
A *growth mindset* is the belief that human capacities are not fixed but can be developed over time, and mindset research examines the power of such beliefs to influence human behaviour.

How does it work?
Growth Mindset interventions aim to improve students' academic achievement and college persistence by helping students view intellectual ability as something that can grow over time with effort and practice rather than as an innate, fixed quantity. These interventions also aim to change students' mindsets to regard academic challenges as learning opportunities rather than permanent impediments.

Growth Mindset interventions are delivered to students individually or in groups, through online modules or in classroom settings. The interventions typically occur once, either at the start of a student's first year of college enrolment or around the time that students complete their first or midterm course exams, and are administered in one or more sessions, each lasting about 30 minutes.

Growth Mindset interventions include an exposure component, in which students are exposed to messages that intelligence is malleable and can grow with effort, and an application component, in which students describe in their own words how to apply these growth mindset messages. Students are exposed to information that intellectual abilities grow when an individual exerts effort toward new or challenging problems. This information may include summaries of research on brain plasticity, the benefits of working hard to solve difficult problems, or testimonials from other students who overcame academic challenges through exerting effort, using effective strategies for learning, or asking for help. Students are prompted to describe how they would apply growth mindset concepts to support their own or another student's persistence in the face of academic challenges. For example, students might explain how they will use these concepts to prepare for their next exam or write a letter encouraging future first-year students that they can improve their academic abilities.

Embedded growth mindset in classroom practice is also encouraged, which requires teachers to believe that humans' intelligence and capacity are malleable, set up high goals for all students, and help students reach the goals.

Helpful resources
Dweck, C. S. (2006). *Mindset: The new psychology of success*. Random house.

Brock, A., & Hundley, H. (2016). The growth mindset coach: a teacher's month-by-month handbook for empowering students to achieve. Simon and Schuster.

Gershon, M. (2016). How to develop growth mindsets in the classroom: The complete guide. CreateSpace Independent Publishing Platform.

Growth Mindset intervention document

Prior evidence

Foliano (2019) and Yeager (2019) conducted two large-scale randomised control trials to estimate the effectiveness of growth mindset. Yeager's work only consisted of two 25-minute sessions while Foliano's work provided more workshops to students and trained teachers to use growth mindset in daily practice. Rienzo (2015) also conducted a randomised control trial but the sample size was small and the attrition rate was high. These studies provide evidence that growth mindset is a promising approach to improve students reading outcomes. However, as Yeager is a member of Dweck's research team, the evidence is not considered as strong.

Report	Age Group	Duration	Evidence strength	Estimated impact
Foliano, F., et al. (2019)	Year 6, U.K.	6 workshops for students; 2 half-day training for teachers	3	ES = +0.01
Yeager, D. S., et al. (2019)	Grade 9, U.S.	2 sessions, 25 min each	3	Overall ES = +0.11, Low achieving ES = +0.33
Rienzo, C., et al. (2015)	Year 5, U.K.	6 workshops and 4 further sessions for students about malleability of intelligence	1	Overall ES = +0.18, Low-achieving ES = +0.34

References

Foliano, F., Rolfe, H., Buzzeo, J., Runge, J., & Wilkinson, D. (2019). Changing mindsets: Effectiveness trial.

Rienzo, C., Rolfe, H., & Wilkinson, D. (2015). Changing Mindsets: Evaluation Report and. Executive Summary. Education Endowment Foundation.

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., ... &. Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. Nature, 573(7774), 364-369.

Ending

Education is highly valued worldwidely. An increasing amount of research evaluates the effectiveness of teaching approaches and investigates the most promising approach. However, there is limited knowledge about using research evidence in teaching practice. Interventions during trials are like throwing a stone in a pool, which makes a small splash and then everything return to normal.

Therefore, you are in a pioneering study that may make a difference to the future. After this workshop, you may have a better understanding of research evidence which is not far from teaching practice at all. Now it is your turn to practice the above evidence in your classroom practice!

Research Team:

Durham University Evidence Centre for Education

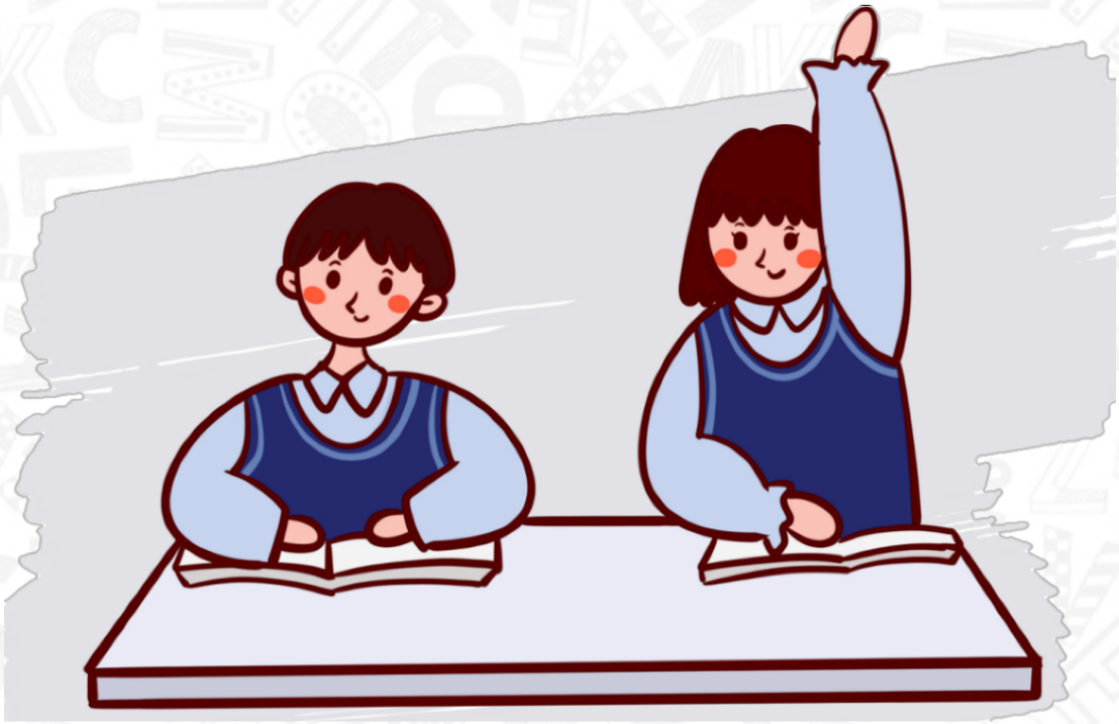


Here is the Chinese version of the Use-of-Research-Evidence intervention booklet, which was also used in the experiment. It did not directly translate the English version but adjusted the wording to enhance readability.

Durham University
Evidence Centre for Education

基于研究证据的小学语文教学策略

Research evidence-led strategies to improve literacy attainment in the primary age



研究证据如何辅助教学？

How does research evidence work in assisting teaching?



教育研究证据是指从一项有正式研究问题、系统性设计、由研究者操作的研究中得到的研究成果，这样的研究目的在于测试某些项目或教学方法的有效性。

将研究证据用于日常实践的想法来自于医学，并逐步扩展到教育学。近年来，美国的 What Works Clearinghouse (WWC) 和英国的 Education Endowment Foundation（EEF）Teaching Toolkit 正在总结归纳最可靠的高质量研究证据，辅助学校、教师优化教学决策。



搭建研究证据和教育实践之间的桥梁，最理想的方法是选择效果最好、证据最强且花费最少的项目。而在现实中，使用什么项目，达到什么目的还需要学校和教师根据学生学情决定。本次课题以学校为单位，需要学校根据实际情况，选择一到两项由本手册提供的项目，并在新的学期投入日常教学。



基于研究证据的高效教学策略示例
Examples promising evidence for teaching reading literacy

下表总结了五个提高小学生语文水平的教学策略示例，这些策略被广泛测试、有着良好的可实施性，且可以有效提高学生的（尤其是学困生和贫困生）学习成绩。

策略名称	建议年级	预估效果	证据可信度
融于课堂的形成性评价	1-6	0.20	🔒🔒🔒
同伴辅导	3-6	0.30	🔒🔒🔒
对话式教学	4-6	0.10	🔒🔒🔒
书写价值感	6	0.07	🔒🔒
成长型思维	5-6	0.10	🔒🔒

这个表格包含了每个策略的：

建议年级。某些策略只对特定年级的学生有效，本课题提供了适用于各个年龄段的建议教学方法，辅助教师合理决策。

预估效果。为了预估不同方法对学生学习成绩的效果，研究者基于先前研究的效应量预估每种教学方法的潜在影响。预估效果的数值越大，就意味着这一策略对学生阅读能力的潜在影响越大。

证据可信度。研究者利用 Gorard SIEVE 对每种策略的可信度从 0 到 4 评级。评级越高，表示证据越强，可信度越高。要注意的是，这些评级与任何干预措施的可能影响无关。

操作指南

Implementation Guidance

以下部分详细介绍了各个策略的细节，包括：
The following sections introduce each strategy in detail, involving:

- 教学策略的描述 Description of the teaching strategy
- 教学策略的运作机理 How the strategy works
- 补充材料 Supplementary materials for implementation
- 过往研究证据 Prior research evidence



融于课堂的形成性评价

Embedded formative assessment

融于课堂的形成性评价目的在于将多种高效的形成性评价策略融入日常教学中。有可靠的研究证据证明，这项策略对于学生的阅读成就有正向效果（效应量 0.1 至 0.3）。

Embedding Formative Assessment (EFA) aims to embed the use of effective formative assessment strategies in daily teaching. Promising evidence about the effectiveness of EFA on reading achievement (ES 0.1 to 0.3).

什么是融于课堂的形成性评价？

What is embedded formative assessment?

形成性评价的定义是：如果由教师、学习者或其同伴引出、解释和运用关于学生成绩的证据，可以帮助教师做出下一步教学的决策，而这种决策比缺乏证据时做出的决策可能更好或更可靠，那么这种评价就在形成性地运行。融于课堂的形成性评价包含了一系列持续的教学活动，为教师的下一步教学决策提供支持。他的目的在于提供及时反馈，帮助学生意识到他们已经明白什么、他们需要知道什么、以及他们需要怎么做才能达到目标。

需要注意的是，融于课堂的形成性评价关注决策。这个策略强调决策驱动的数据收集，而不是数据驱动的决策制定。支持数据驱动决策的人会关注数据，他们收集数据，期望能在将来某个时候派上用场。而那些专注于决策驱动数据收集的人在收集数据之前就已确定要用数据来做什么，且只收集他们需要的数据。

融于课堂的形成性评价是如何运作的？

How does it work?

五个关键技巧如下：

- 明晰、分享和理解学习目标与成功指标——涉及教师、学习者本身及其同伴的共同责任，将其分解为若干个实施阶段和标准。
- 在明确学习意图后，了解学习者当前的学习状况（这个顺序是经过深思熟虑的，除非教师知道自己想让学生学到什么，否则收集什么样的证据就是不明确的）。
- 向学生提供反馈，不仅要告知他们目前的学习状况，还要提示他们需要采取哪些措施才能改善现状。
- 同伴评价对学生的学习发挥着重要作用，同时要明确，基于形成性评价的同伴评价的目的不是为了批判，而是为了改进。
- 最终目标是培养独立的学习者。

	学习者要去哪里	学习者当前在哪里	如何去那里

教师	明晰、分享和理解学习目标与成功指标	设计有效的讨论、任务和活动，引出学习的证据	提供促进学习的反馈
同伴		激发学习者成为彼此的教学资源	
学习者		激发学习者成为自己学习的主人	

补充材料

Helpful resources

融于课堂的形成性评价教师自查清单

Wiliam, D. (2011). 《*融于课堂的形成性评价*》。 Solution tree press.

过往研究证据

Prior evidence

一篇近期发表的系统性文献综述通过总结归纳 48 篇论文，发现形成性评价对学生阅读能力的平均效应量为+0.18 (Xuan, 2022)。另一篇基于美国近年研究的文献综述发现了更大的效应量，为+0.33 (Lee et al., 2020)。2018 年一个大规模随机控制实验显示，Dylan Wiliam 的融于课堂的形成性评价对学生在 GCSE 标化考试成绩的效应量为+0.11。

研究	年级	持续时间	证据可靠度	效应量
Speckesser, S., et al. (2018)	10	2 年	4	+0.11

参考文献

References

Lee, H., Chung, H. Q., Zhang, Y., Abedi, J., & Warschauer, M. (2020). The effectiveness features of formative assessment in US K-12 education: A systematic review. *Applied Measurement in Education*, *33*(2), 124-140.

Speckesser, S., Runge, J., Foliano, F., Bursnall, M., Hudson-Sharp, N., Rolfe, H., & Anders, J. (2018). Embedding Formative Assessment: Evaluation report and executive summary.

Xuan, Q., Cheung, A., & Sun, D. (2022). The effectiveness of formative assessment for. enhancing reading achievement in K-12 classrooms: A meta-analysis. *Frontiers in Psychology*, *13*, 990196.

同伴辅导
Peer tutoring

同伴辅导要求教师根据学生以往的阅读水平，将他们配对或分组。学生的身份在辅导者和被辅导者之间转换，这样他们就有机会互教互学。有证据表明，这种方法尤其对成绩较差的学生有积极影响（效应量 0.1 至 0.4）。

Students are paired or clustered into groups based on previous reading attainment. Students switch the role of peer tutor or peer tutee so that they could get opportunities to teach each other. Some evidence of positive impacts (ES 0.1 to 0.4), especially for low-achievement pupils.

什么是同伴辅导
What is peer tutoring?

同伴辅导是一种促进学生合作学习的教学方法。学生被分成若干小组，通过互教互学来达到学习目标。同伴辅导有两种常见类型：

- 固定同伴辅导（fixed peer tutoring）：一个学业成绩较好的学生对另一个较弱的学生进行学习上的支持和辅导；
- 互惠同伴辅导（reciprocal peer tutoring）：学生轮流担任小老师，对自己的同伴进行学业指导。

同伴辅导又根据同伴的年龄段分为同龄同伴辅导和混龄同伴辅导，但混龄同伴辅导需要打乱原有的班级分配，其可行性不如同龄同伴辅导。

同伴辅导是如何运作的？
How does it work?

同龄同伴辅导既不需要打乱原有的分班，也不需要调整教学进度，只需要教师根据自己的判断（通常是学生的成绩）将学生匹配成学习小组，在课堂中适当地插入同伴辅导任务。研究显示，每周实施 1 至 4 次，每次 20-40 分钟，可以产生积极的影响。同时研究证据建议，为了更好地观察到学生的改变，最好能保证每周 3 次的实施频率。

在固定的同伴辅导中，学生们保持着他们作为整个课程的辅导者或被辅导者的角色。教师将学生语文成绩从高到低排列，前 50%的学生成为辅导者，后 50%的学生成为被辅导者。成绩最好的学生辅导成绩中等偏下的学生，而语文水平最弱的学生由成绩中等偏上的学生辅导。

在互惠同伴辅导中，教师基于他们对学生的需求和能力的理解决定分组，并定期进行重新分配。与固定同伴辅导不同的是，辅导者和被辅导者的角色不是固定不变的。学生互助的中途会对换角色，相互提供辅导并纠正同伴的问题。

同伴辅导的第一阶段关注对记叙文的理解和同伴互助规则的制定。在第二阶段，同伴辅导关注学生对长篇、阅读难度较高的阅读材料的阅读能力。

补充材料

Helpful resources

同伴辅导活动清单

同伴辅导教师、学生自查表

同伴辅导课例

Frey, N., Fisher, D., & Everlove, S. (2009). *《教师如何提高学生小组合作学习效率》* Ascd.

过往研究证据

Prior evidence

过往很多研究都评估了同伴辅导的效果，但大多数研究都局限于个案分析或缺乏对照组，以下的两个实验都基于随机对照试验。

第一个研究包含了 125 所苏格兰小学的 3000 名四年级和六年级的学生，在两年间随机参加了同龄同伴辅导和混龄同伴辅导。在第二个研究中，118 位四年级学生被随机分配到同伴互助阅读和对照组。结果显示，同伴辅导对辅导者和被辅导者均存在积极影响。

研究	年级	持续时间	证据可靠度	效应量
Topping et al. (2012)	4，6	30 分钟，每周 1 次或 3 次，持续 15 周，实施两年	3	+0.31（第一年） +0.34（第二年）
Kung & Teo (2022)	4	40 分钟，每周两次，持续 15 周	2	+0.07（词汇） +0.11（阅读理解）

参考文献

References

Kung, S.-H., & Teo, T. (2022 - under review). Effects of peer-assisted learning strategies on reading literacy of fourth-grade Chinese-speaking students. Asia-Pacific Journal of Teacher Education.

Topping, K. J., Thurston, A., McGavock, K., & Conlin, N. (2012). Outcomes and process in reading tutoring. Educational Research, 54(3), 239-258.

对话式教学
Dialogic teaching

对话式教学鼓励学生围绕某一主题展开讨论，为学生论证某一观点、发展批判性思维提供支持（每周至少一节课，持续一个学期）。研究证据表明，该方法尤其对阅读能力较弱的学生有积极影响（效应量 0 至 0.2）。

Dialogic teaching encourages discussion and provides support for argumentation/critical thinking (one or more lessons per week, one term). Some evidence of positive impacts especially for weaker students (ES 0 to 0.2).

什么是对话式教学？
What is dialogic teaching?

课堂谈话和讨论可以为儿童的学习和社会情感发展提供有益的机会，提高课堂师生、生生对话质量，对提高课堂互动质量尤为重要。对话式教学利用谈话的力量来培养兴趣，引发思考，促进对文本材料的理解，拓宽思想，构建和评价论点，使学生能够终身学习和培养批判性思维。

对话式教学是如何运作的？
How does it work?

儿童哲学（Philosophy for Children, P4C）是对话式教学的成功模式之一。P4C 是一种覆盖全班的干预措施，旨在激发引导课堂对话，来回答学生自己对阅读的故事、观看的电影或其他启发材料的问题。教师并不需要熟悉哲学作品，相比之下更重要的是一种真正的哲学好奇心、开放的思维和与学生一同探讨的能力。该方案建议每周定期实施一小时的 P4C 课程，至少持续 3 个月。

P4C 课程包括在课堂上采取逐步讨论的方法。老师和学生们围坐一圈，轮流发言。教师先出示一个启发材料，如图片、引文、新闻、短篇故事或诗歌来介绍一个话题，并鼓励儿童在此基础上产生开放式问题。这些问题不一定是真实事件，但可以引发反思性和概念性的讨论，同时这些问题不需要有唯一正确的答案。学生使用对话式语言来论证他们的观点，提出问题，进行批判，并阐述他们的反应和看法。

在学生们提出问题后，教师用投票的方式来选择最受欢迎的问题，成为讨论的焦点。讨论时，学生轮流发言，运用事先教给他们的简单论证术语和结构来支持他们的观点。教师只在必要时进行干预，以促成有用而富有成效的讨论。需要注意的是，教师应该避免自己的言论导致学生追求单一的、“正确的”答案。对此，在必要的时候，教师可以用多项选择题开拓学生的观点。讨论结束后，要求学生们总结想法，反思讨论期间观点的改变，评估讨论的质量。

补充材料
Helpful resources

http://www.p4c.org.cn/cn/index.aspx

P4C 活动流程

过往研究证据

Prior evidence

本研究选择了四个利用随机控制实验评价对话式教学有效性的研究，其中第1、2、4是针对 P4C 效益的评价，第三个是针对其他对话式教学模型的评估。研究结果表明，对话式教学对学生的阅读理解能力有明显的正向作用，与此同时，该项目施行一年内可能有较好的效果。

研究	年级	持续时间	证据可靠度	效应量
Gorard et al. (2018)	5, 6	1 年	3 🔒	+0.12（阅读） +0.03（写作）
Lord et al. (2021)	6	2 年	3 🔒	+0.01（阅读）
Jay et al. (2017)	5	20 周	3 🔒	+0.12（CAT4 阅读） +0.03（KS2 阅读）
Fair et al. (2015)	7, 8	24 周	2 🔒	+0.21（阅读）

参考文献

References

Fair, F., Haas, L. E., Gardosik, C., Johnson, D. D., Price, D. P., & Leipnik, O. (2015). Socrates in the schools from Scotland to Texas: Replicating a study on the effects of a Philosophy for Children program. Journal of Philosophy in Schools, 2(1).

Gorard, S., Siddiqui, N., & See, B. H. (2018). Philosophy for Children: Evaluation report and executive summary. EEF: London UK

Jay, T., Willis, B., Thomas, P., Taylor, R., Moore, N., Burnett, C., ... & Stevens, A. (2017). Dialogic teaching: Evaluation report and executive summary.

Lord, P., Dirie, A., Kettlewell, K., & Styles, B. (2021). Evaluation of philosophy for children: an effectiveness trial. EEF: London, UK.

书写价值感
Writing about values

在重大考试前，教师组织学生完成关于自我价值感的短文写作练习。该方法可以在短期内小范围提高学生的阅读成绩（效应量 0.05 至 0.1）。

Pupils facing a high-stakes test write two or three brief essays about their values. Reasonable evidence of impact for disadvantaged pupils (ES 0.05 to 0.1).

什么是书写价值？
What is writing about values?

提高自我价值感是一种心理学方法，目的是帮助一些处境不利的学生克服所经历的刻板影响。为了保护他们的自我价值感不被贬低，一些学生还发展出防御机制，例如对考试成绩表现得漠不关心。这种干预措施可以拓宽学生的视野，让他们不再觉得自己是“坏”学生。在刻板印象威胁最突出的时候，撰写关于价值观的积极声明，可以给个人带来积极的价值感，消除负面情绪，促进学习。这项干预措施最显著的优点是效益高、实施简单。

提高自我价值感是如何运作的？
How does it work?

自我价值感干预包括两到三次关于价值感的写作活动，每次持续 10 到 15 分钟。写作练习放在指定的信封里，并单独分发给学生，学生需要把写作练习纸放回信封里再上交，由教师保存。

任务开始之前，教师对任务做一个简短的、有条理的介绍，明确这个任务是一个以学生自己的想法和观点为重点的写作练习，没有正确或错误的答案，重要的是完成这个活动的过程。学生的写作不会被老师阅读或打分，也不会提供反馈，这个干预只是为了鼓励自由和开放的思考。

但需要注意的是，教师和学生都不应该确切地知道这项干预的具体原理是什么，因为有证据表明，学生在了解这项干预的原理之后会降低其效果，教师在实施该方法时应该尽一切努力模糊写作练习背后的原理。

第一次写作练习通常是最重要的，如果学生由于第一个写作活动而表现得更好，他们的自信心就会提高，亦或者是老师对他们表现出更高的期望，进而激发学生更好的表现，而这个过程就会延续下去。因此，学生高质量完成第一个写作练习至关重要，而第二次（和第三次）练习是为了强化正向的表现。

研究建议，自我价值感写作练习最好在学期开始时和大型考试之前的班会课或语文课上进行。

补充材料

Helpful resources

书写价值感写作任务

书写价值感教师自查表

过往研究证据

Prior evidence

自我价值感干预在美国已经被广泛评价，证据表明，这项干预可以在阅读及其他多个学科产生长期的积极效果，提高学习成绩，特别是针对少数民族学生和贫困学生等处境不利的学生（例如，Wu et al., 2021）。一些研究还表明，这些效果一直持续到中学，甚至大学（Borman et al., 2021；Goyer et al., 2017）。

在英格兰进行的一项大型随机控制实验表明，自我价值感干预对处在不利社会经济地位学生的成绩尤其有效（See et al.，2022）。该实验囊括了英格兰地区 11000 名 14 到 16 岁的学生，在经过两次写作任务后，实验组学生的在标准化考试中的整体成绩有明显提升（ES = +0.05）。

研究	年级	持续时间	证据可靠度	效应量
See et al. (2022)	9，10，11	1 年	4	+0.05

参考文献

References

Borman, G. D., Choi, Y., & Hall, G. J. (2021). The impacts of a brief middle-school self-affirmation intervention help propel African American and Latino students through high school. Journal of Educational Psychology, 113(3), 605–620. https://doi.org/10.1037/edu0000570

Goyer, J. P., Garcia, J., Purdie-Vaughns, V., Binning, K. R., Cook, J. E., Reeves, S. L., Apfel, N.,Taborsky-Barba, S., Sherman, D. K., & Cohen, G. L. (2017). Self-affirmation facilitates minority middle schoolers' progress along college trajectories. Proceedings of the National Academy of Sciences, 114(29), 7594–7599. https://doi.org/10.1073/pnas.1617923114

See, B.H., Morris, M., Gorard, S., Siddiqui, N., Easterbrook, M.J., Nieuwenhuis, M., Fox, K., Harris, P.R. & Banerjee, R. (2022) A conceptual replication study of a self-affirmation intervention to improve the academic achievement of low-income pupils in England, *Educational Research and Evaluation*, DOI:10.1080/13803611.2021.2022317

Wu, Z., Spreckelsen, T. F., & Cohen, G. L. (2021). A meta-analysis of the effect of values affirmation on academic achievement. Journal of Social Issues, 77(3), 702–750. https://doi.org/10.1111/josi.12415

成长型思维
Growth mindset

学生参加一场阅读讨论会来了解成长型思维的基本概念：大脑像肌肉一样是可塑的。之后，他们就如何在日常生活中运用成长型思维展开讨论。此外，成长型思维鼓励教师在教学中贯穿这一思维，并组织相关活动帮助学生培养成长型思维。研究显示，成长型思维对学生（尤其是学困生）有着积极影响（效应量 0 至 0.2）。

Students participate in one reading workshop about the idea that the brain is malleable like a muscle to get basic information about growth mindset. Then they are invited to discuss how they could implement growth mindset in their daily life. Teachers are also encouraged to embed growth mindset in daily teaching and conduct relevant activities. Some evidence of positive impacts especially for low-achieving students (ES 0 to 0.2).

什么是成长型思维
What is Growth mindset?

成长型思维相信能力是可以被培养的而不是固定的，在过往对思维的研究中，这种信念会影响人的行为。

成长型思维如何运作？
How does it work?

成长型思维干预的逻辑是：当学生认识到智商是可以通过努力和训练而提高的，而不是与生俱来的之后，学生可以更好地应对学业挑战，从而提高学习成绩、锻炼毅力。这项干预措施改变学生的心态，将学业挑战视为学习机会，而不是永久的障碍。

成长型思维通过线上模块或课堂环境传播，以个人或小组的形式对学生进行干预。活动通常在学期开始时或期中考试前后进行，分一次或多次进行，每次持续约 30 分钟。

成长型思维的干预措施包括“Expose”（暴露）和“Apply”（使用）两个部分，“Expose”让学生了解智力是可塑的，可以通过努力增长；“Apply”让学生用自己的语言描述如何应用这些成长型思维模式的信息。在第一阶段，学生会接触到这样的信息：当一个人努力解决新的或具有挑战性的问题时，他的智力就会增长。这些信息可能包括大脑可塑性研究摘要、努力解决难题的好处，或其他学生通过努力、使用有效的学习策略或寻求帮助而克服学业挑战的经历。第二阶段，学生们会描述他们将如何应用成长型思维模式的概念来支持自己或其他学生在面对学业挑战时的坚持不懈。例如，学生可以描述他们将如何利用这些概念来准备下一次考试，或者写一封信鼓励下一届学生，告诉他们可以提高自己的学习能力。

此外，本课题鼓励教师在日常教学中融合成长型思维，这需要教师相信人的智力和能力是可塑的，为每一位学生设定高目标并帮助学生达到目标。

补充材料

Helpful resources

Dweck, C. S. (2006). *《终身成长》*。Random house.

Gershon, M. (2016). *《如何在课堂中培养成长型思维》*。CreateSpace Independent Publishing Platform.

Brock, A., & Hundley, H. (2016). *《成长型思维训练：12 个月改革学生思维模式指导手册》*。Simon and Schuster.

成长型思维学生阅读材料

成长型思维教案

过往研究证据

Prior evidence

Foliano（2019）和 Yeager（2019）开展了两项大规模随机对照试验，以评估成长型思维模式的有效性。Yeager 的项目包括两节 25 分钟的课程，而 Foliano 的工作则为学生提供了更多的研讨会，并培训教师在日常实践中使用成长型思维模式。Rienzo（2015）也进行了随机对照试验，但样本量较小，流失率较高。这些研究证明，成长型思维模式是一种很有前途的提高学生阅读成果的方法。不过，由于 Yeager 是 Dweck 研究团队的成员，因此这些证据并不够令人满意。

研究	年级	持续时间	证据可靠度	效应量
Foliano, F., et al. (2019)	6	6 场学生研讨会; 2 次教师培训	3 🔒	+0.01
Yeager, D. S., et al. (2019)	9	2 次, 每次 25 分钟	3 🔒	+0.11 +0.33（学困生）
Rienzo, C., et al. (2015)	5	6 场学生研讨会，另附加 4 场智力可塑性相关课程	1 🔒	+0.18 +0.34（学困生）

过往研究证据

Prior evidence

Foliano, F., Rolfe, H., Buzzeo, J., Runge, J., & Wilkinson, D. (2019). Changing mindsets: Effectiveness trial.

Rienzo, C., Rolfe, H., & Wilkinson, D. (2015). Changing Mindsets: Evaluation Report and. Executive Summary. Education Endowment Foundation.

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., ... & Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. Nature, 573(7774), 364-369.

结语
Ending

教育在全球各个国家都至关重要。越来越多的研究关注着教学法的有效性，研究者试图摸索出效益最大的方法。然而，学界对研究证据在教学实践中的运用知之甚微。许多教育学实验研究成果在实验后就被束之高阁，犹如投石于深潭中，纵然漾起微澜，经时而复寂静。

而您正在这个世界前沿的研究中扮演着重要的角色。在这场座谈会之后，或许您对研究证据有了更进一步的认识。那么，现在轮到您在日常教学中使用这些研究证据了！

Education is highly valued world widely. An increasing amount of research evaluate the effectiveness of teaching approaches for the most promising approach. However, there is limited knowledge about using research evidence in teaching practice. Interventions during trials just likes a stone throwing in a pool, which makes a small splash and restores to tranquillity.

Therefore, you are in a pioneering study that may make a difference to the future. After this workshop, you may have a better understanding of research evidence which is not far from teaching practice at all. Now it is your turn to practice the above evidence in your classroom practice!

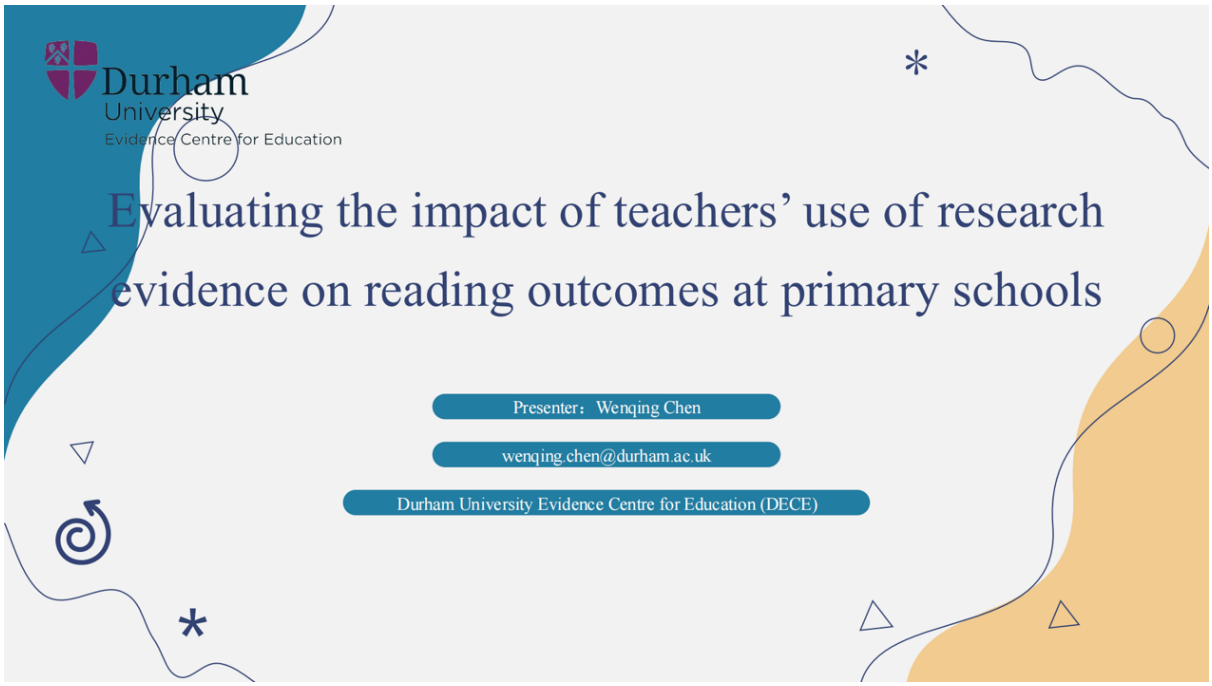
研究团队:

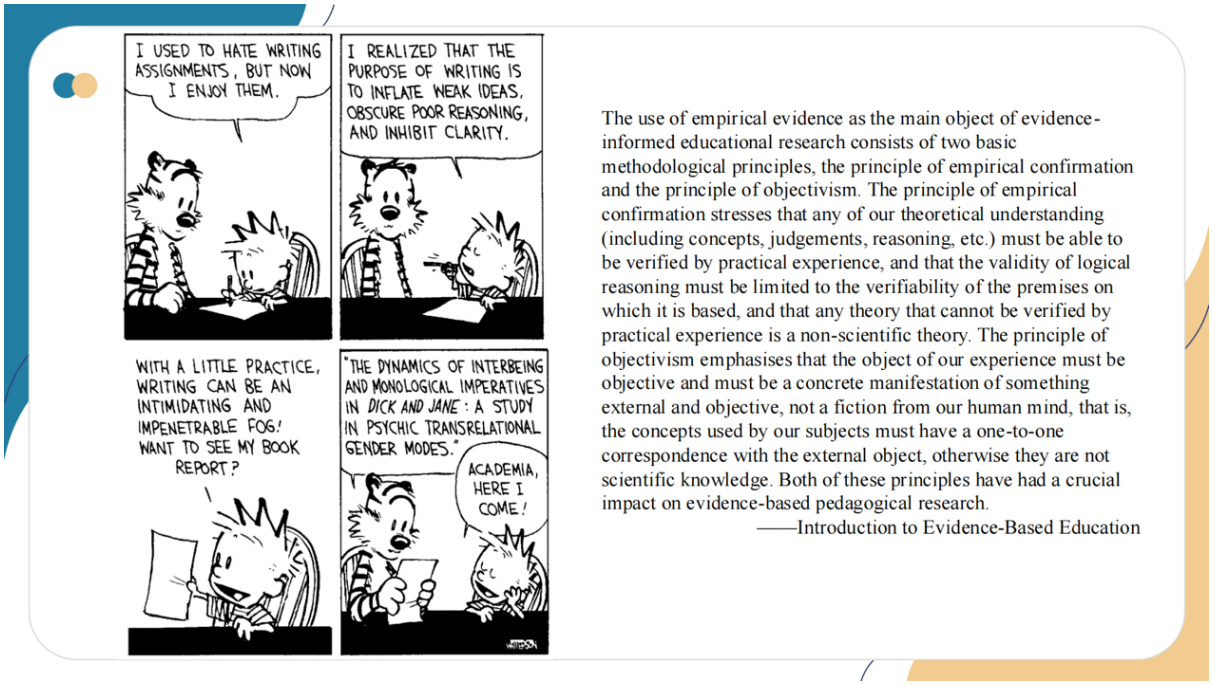
杜伦大学教育证据研究中心



Appendix C. Use-of-research-evidence training slides

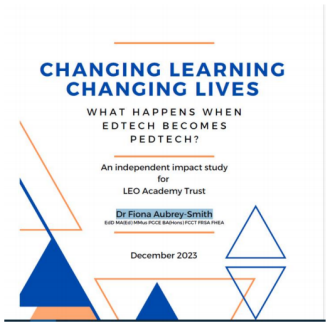
Here is the English version of the *Use-of-Research-Evidence* training slides.





The use of empirical evidence as the main object of evidence-informed educational research consists of two basic methodological principles, the principle of empirical confirmation and the principle of objectivism. The principle of empirical confirmation stresses that any of our theoretical understanding (including concepts, judgements, reasoning, etc.) must be able to be verified by practical experience, and that the validity of logical reasoning must be limited to the verifiability of the premises on which it is based, and that any theory that cannot be verified by practical experience is a non-scientific theory. The principle of objectivism emphasises that the object of our experience must be objective and must be a concrete manifestation of something external and objective, not a fiction from our human mind, that is, the concepts used by our subjects must have a one-to-one correspondence with the external object, otherwise they are not scientific knowledge. Both of these principles have had a crucial impact on evidence-based pedagogical research.

——Introduction to Evidence-Based Education



The attainment of children at LEO schools significantly outperforms national norms in national tests

This study is framed within an inductive-deductive interpretivist paradigm (Cohen, Manion & Morrison, 2007).
The design of this study adopts a phenomenological mindset - seeking to surface the implicit influences which have underpinned the evolution of LEO's 1:1 approach. the research was conducted in partnership with 16 middle, senior and executive leaders from across LEO Academy Trust rather than solely by an independent outsider-researcher.

Building Educated Leaders for Life (BELL)
Over 1,000 elementary school children who applied to BELL summer programs in New York and Boston in 2005 were randomly chosen to be in either a treatment group that was selected to participate in the BELL summer program, or a comparison group that was not.
"a well-implemented summer learning program can improve reading test scores"

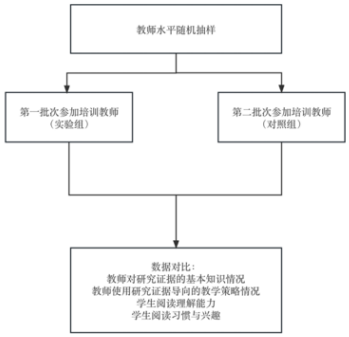
Outcome	Treat	Control	Standard deviation	Effect size	P-value	N
Total	423	422	65	+0.02	0.67	847
Vocabulary	420	422	67	-0.03	-0.73	847
Comprehension	418	417	73	+0.02	0.45	847

01

About this project

Introduction





Project title

Evaluation of teachers' use of research evidence on students' reading outcomes

Aim of the project

Explore the impact of high quality effective education research findings into real teaching and learning environments.

Stage 1: Acquire basic knowledge and processes for searching for research evidence and integrating teaching strategies into teaching and learning, using the teaching strategies chosen by the school.

Stage 2: Self-select appropriate instructional strategies based on student learning contexts

Research design

Randomised controlled experiment in which teachers were randomly divided into two cohorts and attended the training successively.

Timeline: Embedded formative assessment, peer tutoring

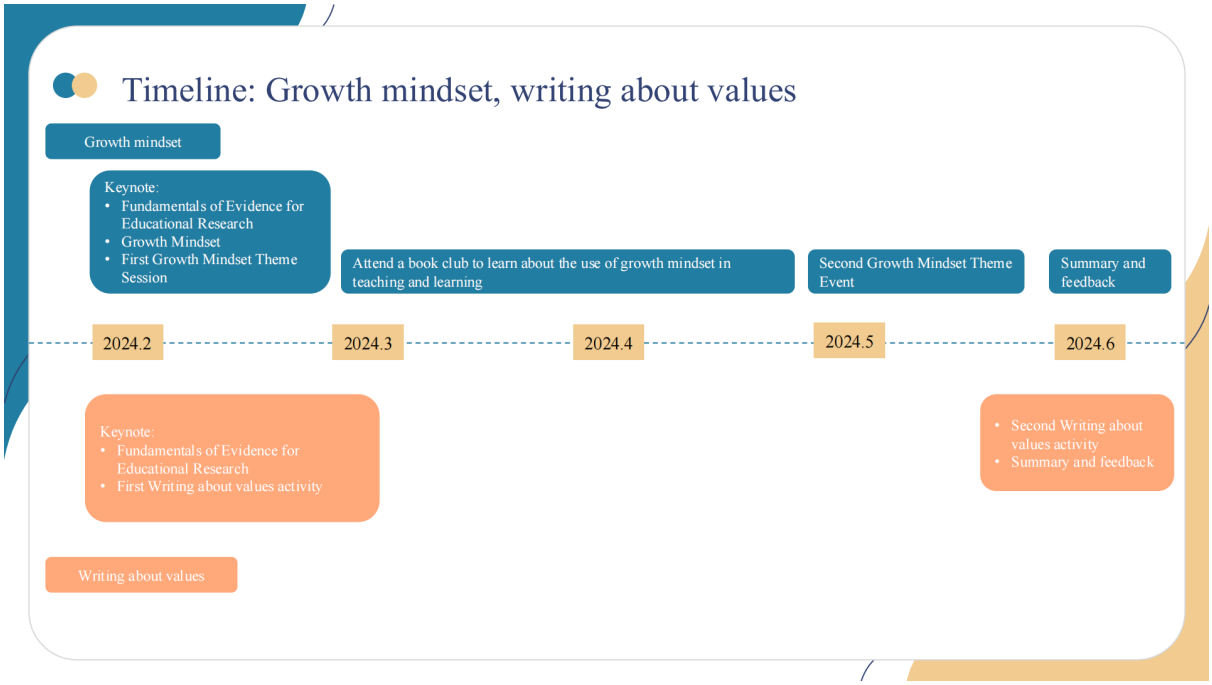
- Deeper Learning Research Evidence-Based Instructional Strategies
- Developing Research Evidence-Based Instructional Plans
- Participate in public classes
- Experiment with research evidence-based instructional strategies
- Provide feedback and make classroom adjustments

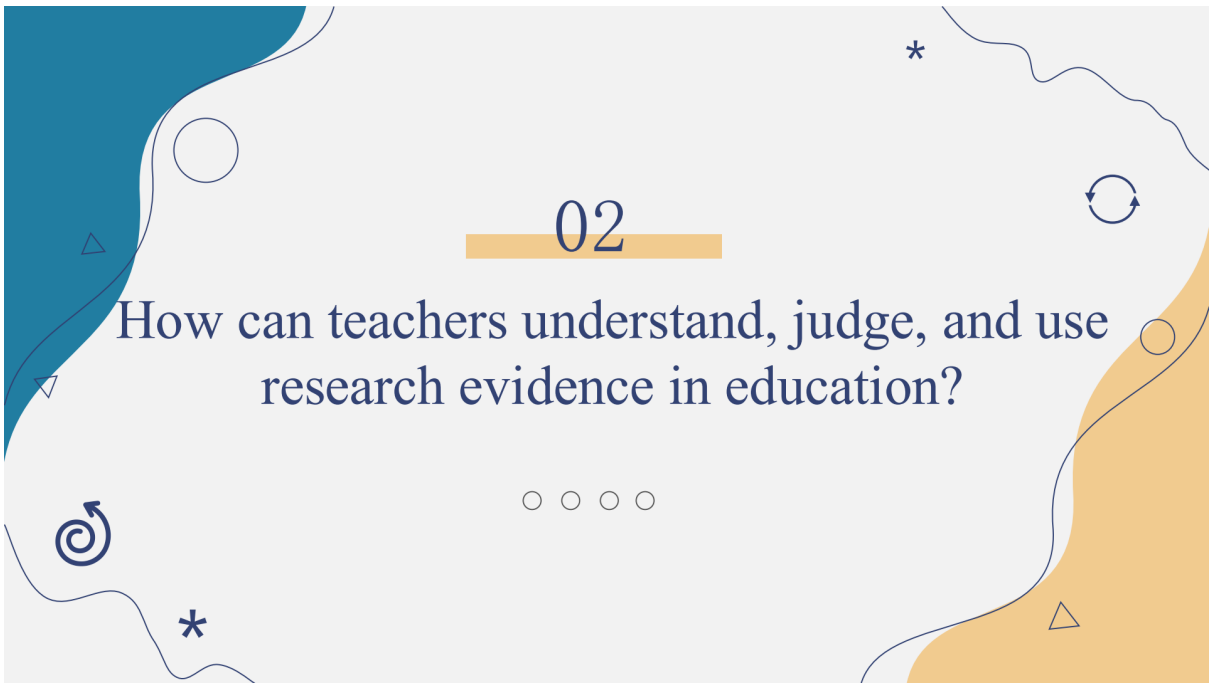
Summary and feedback

2024.2 — 2024.3 — 2024.4 — 2024.6

Keynote:
- Fundamentals of Research Evidence in Education
- Teaching strategies based on research evidence

- Daily use of teaching strategies based on research evidence
- Polishing public lessons


Timeline: Growth mindset, writing about values
Growth mindset
Keynote:
• Fundamentals of Evidence for Educational Research
• Growth Mindset
• First Growth Mindset Theme Session
Attend a book club to learn about the use of growth mindset in teaching and learning
Second Growth Mindset Theme Event
Summary and feedback
2024.2
2024.3
2024.4
2024.5
2024.6
Keynote:
• Fundamentals of Evidence for Educational Research
• First Writing about values activity
• Second Writing about values activity
• Summary and feedback
Writing about values


02
How can teachers understand, judge, and use research evidence in education?

Two dimensions of evidence-into-use, and who judges the quality of research?

Research evidence

- Research evidence in education usually refers to findings from research with a formal research question, and a systematic design, conducted by research experts in education. Such research often evaluates the effectiveness of certain programmes or approaches in education.

	Passive transfer	Engagement in transfer	(inter)Active transfer
Plain evidence	e.g. open access to journals	e.g. journal clubs	e.g. practitioner inquiry
Modified evidence	e.g. EEF Toolkit for practitioners	e.g. Think tanks	e.g. internships, research schools
Engineered evidence	e.g. lesson plans	e.g. hotlines, helpdesks	e.g. legislation for population measures

How to access research evidence?

Searching Engine：

CNKI、Wanfang Data、Google Scholar…

Educational research journal：

Teacher education research、Shanghai education research、Journal of The Chinese Society of Education、Review of Educational Research、Journal of Teacher Education…

International report：

PISA、TIMSS、PIRLS…

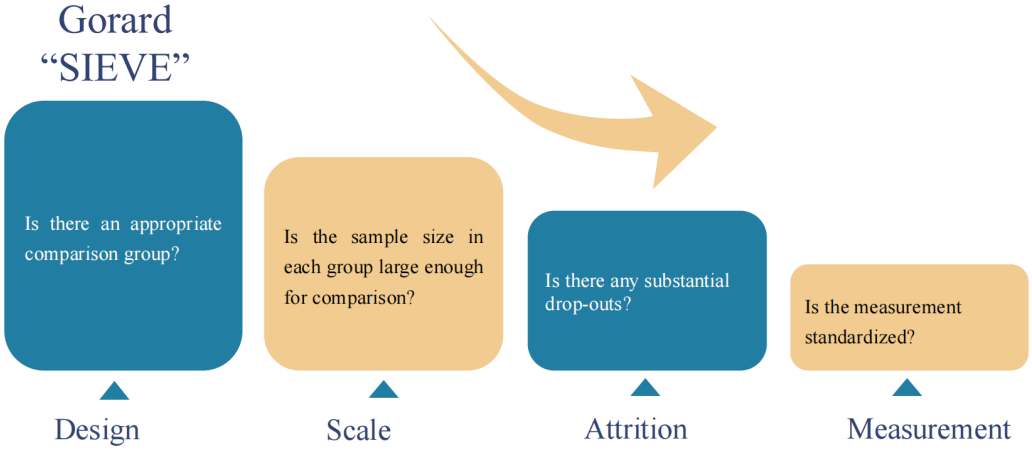
Websites of educational research evidence:

What Works Clearinghouse、 EEF Teaching and Learning Toolkit



How to assess the quality of research?

Gorard
"SIEVE"



Examples

To explore the effect of graded reading on students' language reading skills:

① The students of 30 classes were randomly divided into two groups on a class basis, where one group of students took a language reading ability test and the teacher recommended the students to read books of different tiers according to their reading ability. After a period of time, the researcher designed the test questions to compare the test scores of the two groups of students

② Interview 40 students about their perceptions of graded reading after a period of time in a school using software related to graded reading.

③ A portion of a class in a school was selected as an experimental class to use the graded reading software, and the language scores of the experimental class were compared before and after the use of the graded reading software.

④ A portion of a class in a school was selected as an experimental class to use graded reading software. After a period of time, questionnaires were used to collect evaluations from students in the experimental and regular classes about their reading abilities.

Which has the best design?

Which has the best measurement?

Which is most trustworthy?

How to understand effect size?

Common effect size

Cohen's d effect size

Pearson r

Standard deviation

…

What is effect size?

- a statistical measure of the difference between two groups
- assess the importance of a study's findings

A large effect size means?

- In the experiment
- On average, students in the experimental group achieved more than students in the control group

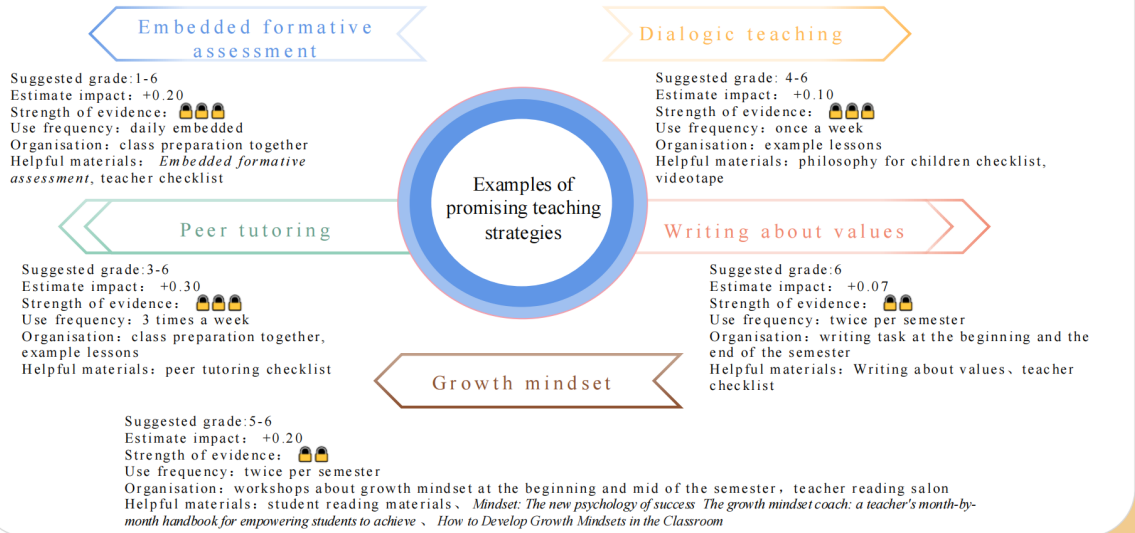
Relationship between effect size and research quality

- not necessarily related
- High effect size does not mean good study quality
- High-quality studies also do not guarantee substantial effect sizes

Process of research evidence use



What are promising evidence-led teaching strategies?



03

How do teachers implement research evidence-informed teaching strategies?

Embedding formative assessment

A recent structured reviews combining 48 studies suggests that the average effect size of formative assessment on reading achievement is +0.18 (Xuan, 2022). Another review suggests larger effect size +0.33 on reading in the U.S （Lee et al., 2020）. In 2018, a large-scale randomised control trial illustrates that using Dylan Wiliam's embedded formative assessment could benefit students' GCSE achievement with an effect size of +0.11.

Report	Year group	Duration	Outcome	Security	Effect size
Speckesser, S., et al. (2018)	Year 10, U.K.	2 years	GCSE	4	+0.11

Who is using

Purpose

Formative assessment is defined as: to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited.

Better decisions

Action

	Where the learner is going	Where the learner is now	How to get there
Teacher	Clarifying, sharing, and understanding learning intentions and success criteria	Engineering effective discussions, tasks, and activities that elicit evidence of learning	Providing feedback that moves learning forward
Peer		Activating students as learning resources for one another	
Learner		Activating students as owners of their own learning	

Peer tutoring

Many studies have evaluated the effectiveness of peer tutoring, however, the majority of the research has been limited to case studies or lacks balanced comparison groups. The following two studies were randomised control trials.

More than three thousand pupils from 125 primary schools in Scotland participated in the first study which implemented both cross-age and same-age peer tutoring over two years. The second study directly evaluates the effects of peer-assisted learning, randomising 118 students from Grade 4. Both studies have promising results in improving reading achievement of low-achieving students.

Report	Year group	Duration	Outcome	Evidence strength	Estimated impact
Topping et al. (2012)	Year 4 and Year 6	30-minute sessions, 1 or 3 times per week, for a duration of 15 weeks	Reading	3	+0.31 First Year of Implementation +0.34 Second Year of Implementation
Kung & Teo (2022)	Year 4 (Taiwan)	40-minute sessions, twice a week, for a duration of 15 weeks	Reading	2	+0.07 Vocabulary +0.11 Reading Comprehension

Set up goals

What's your purpose of using peer tutoring?
What's your students' learning status?

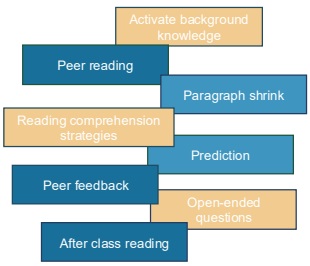
Organise and match

Pair students according to their reading ability, with a gap between peers, but not too large a gap. Students can swap roles at the end of a lesson or during the same lesson to adjust to classroom needs



Monitoring and feedback

Observe group activities and give guidance where necessary. Regular feedback with students to summarise the effectiveness of peer tutoring.



Writing about Values

Self-affirmation interventions have been widely evaluated, mostly in the U.S. The evidence suggests that self-affirmation activities can have positive and long-term results improving academic achievement, especially for ethnic minority groups (e.g. Wu et al., 2021). Some studies have also suggested that these effects persist through secondary school (Borman et al. 2021) and right up to university (Goyer et al., 2017).

Two studies conducted in England have shown that the self-affirmation approach is also effective in raising the attainment of low-SES pupils (See et al., 2022). See et al.'s (2022) involving c.11,000 pupils aged 14 to 16 in England also showed positive effects on pupils' Attainment 8 scores (ES = +0.05) for those eligible for free school meals.

Report	Age Group	Duration	Outcomes	Evidence strength	Estimated impact
See et al. (2022)	14-16	One year	Overall Attainment	4	+0.05

There are a lot of things that are important to people—things that make their lives better, more important, or special.

For example, some people find being honest important because other people can trust them. Some other people find their family important because they love and value them. Other people find being good at sport important because it makes them feel good to play well.

In the space below, please write about what **you** find important in **your** life. How important is it to you? Why is it important to you? What does it mean to you to have it in your life?

Focus on your thoughts and feelings, and don't worry about spelling, grammar, or how well written it is, or how much you can write.

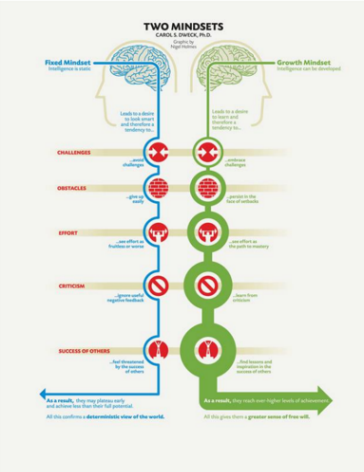
There are a lot of things that are important to people—things that make their lives better, more important, or special.

In the space below, please select two values from the list that are important to you and write about what they will do to show that these are important or how much you enjoy practicing them.

List of Values: humility, compassion, honesty, friendship, determination, courage, trustworthiness, happiness, self-reliance, curiosity, knowledge, fame, and achievement.

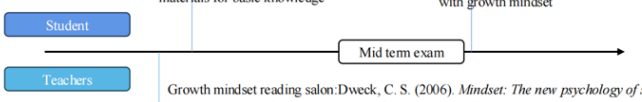
Focus on your thoughts and feelings, and don't worry about spelling, grammar, or how well written it is, or how much you can write.

Growth mindset



Foliano (2019) and Yeager (2019) conducted two large-scale randomised control trials to estimate the effectiveness of growth mindset. Yeager's work only consisted of two 25-minute sessions while Foliano's work provided more workshops to students and trained teachers to use growth mindset in daily practice. Rienzo (2015) also conducted a randomised control trial but the sample size was small and the attrition rate was high. These studies provide evidence that growth mindset is a promising approach to improve students reading outcomes. However, as Yeager is a member of Dweck's research team, the evidence is not considered as strong.

Report	Year group	Duration	Outcome	Evidence strength	Effect size
Foliano, F., et al. (2019)	Year 6, U.K.	6 workshops for students; 2 half-day training for teachers	Key stage 2 English	3	ES = +0.01
Yeager, D. S., et al. (2019)	Grade 9, U.S.	2 sessions, 25 min each	English GPA	3	Overall ES = +0.11, Low achieving ES = +0.33
Rienzo, C., et al. (2015)	Year 5, U.K.	6 workshops and 4 further sessions for students about malleability of intelligence	Progress in English (PiE)	1	Overall ES = +0.18, Low-achieving ES = +0.34

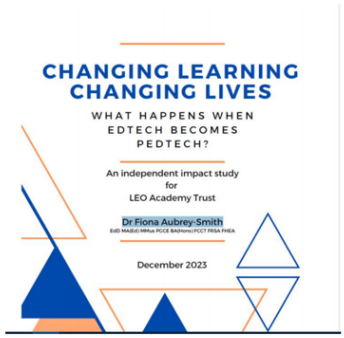


Growth mindset reading salon:Dweck, C. S. (2006). *Mindset: The new psychology of success*. Random house. Brock, A., & Hundley, H. (2016). The growth mindset coach: a teacher's month-by-month handbook for empowering students to achieve. Simon and Schuster. Gershon, M. (2016). How to develop growth mindsets in the classroom: The complete guide. CreateSpace Independent Publishing Platform.

Here is the Chinese version of the Use-of-Research-Evidence training slides, which were used in the experiment. The Chinese version not only translates the content of the English version but also adjusts the wording to enhance readability.







学校的毕业生成绩远远优于英国国家均值。

本研究采用归纳-演绎解释范式（Cohen, Manion & Morrison, 2007）。
本研究的设计采用了现象学的思维方式--试图揭示支撑 LEO 1:1 方法演变的隐性影响因素。
这项研究是与 16 名来自 LEO 学院信托基金的中层、高层和执行领导合作进行的，而不是由独立的外部研究者单独进行的。

Building Educated Leaders for Life (BELL) 夏季研学活动
2005 年，纽约和波士顿有 1000 多名小学生申请加入贝尔暑期项目，他们被随机选入被选中参加BELL 夏季研学活动的实验组，未被选中的作为对比组。
结论：一个实施良好的暑期研学活动可以提高阅读考试成绩

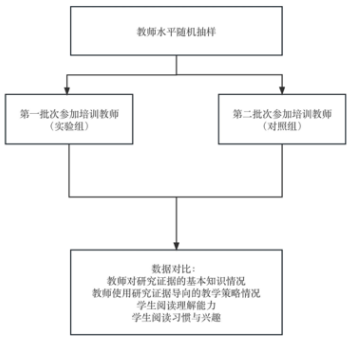
结果	实验组	对照组	标准差	效应量	显著性	人数
总分	423	422	65	+0.02	0.67	847
词汇	420	422	67	-0.03	-0.73	847
阅读理解	418	417	73	+0.02	0.45	847

01

关于本次课题

研究内容





课题名称

探索教师使用教育研究证据对小学语文教学效率的影响

课题目标

探索教师在教学中使用被实验严格论证有效的教学策略是否会对学生语文学习产生积极影响。
第一阶段：了解研究证据导向的教学策略，结合学情，在课堂中尝试使用
第二阶段：基于第一阶段的结果，稳步实施研究证据导向的教学策略

研究设计

随机控制实验，将教师随机分成两批，先后参加培训。

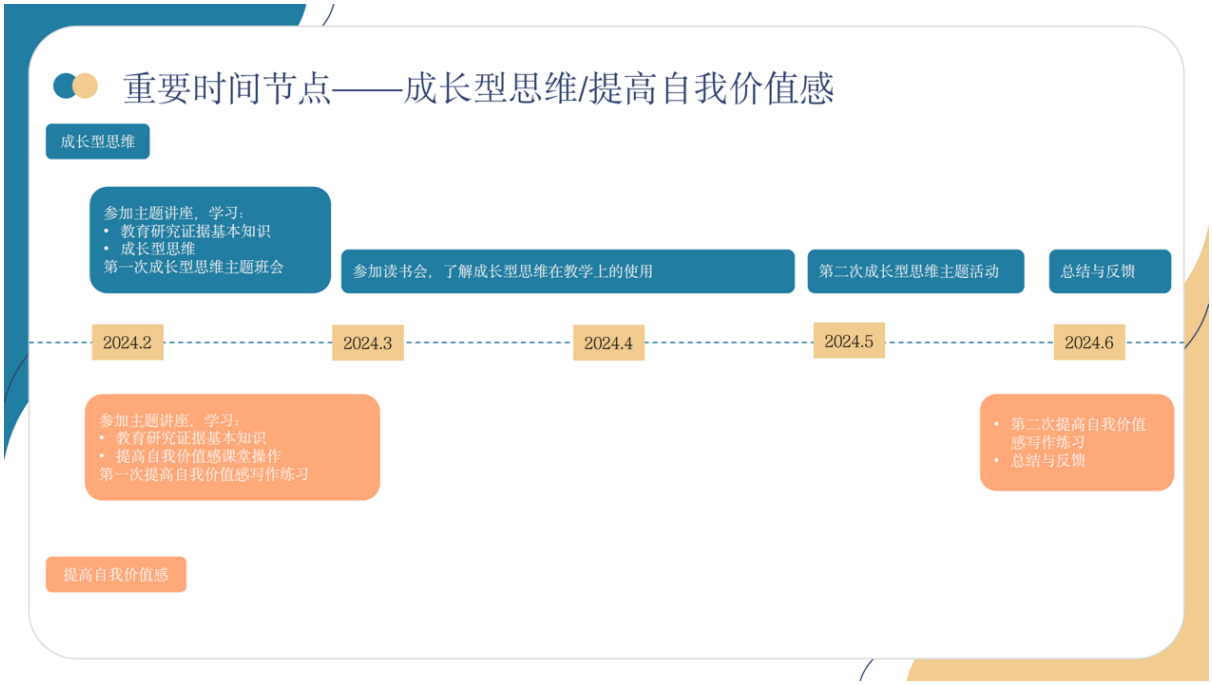
重要时间节点——同伴辅导/融于课堂的形成性评价

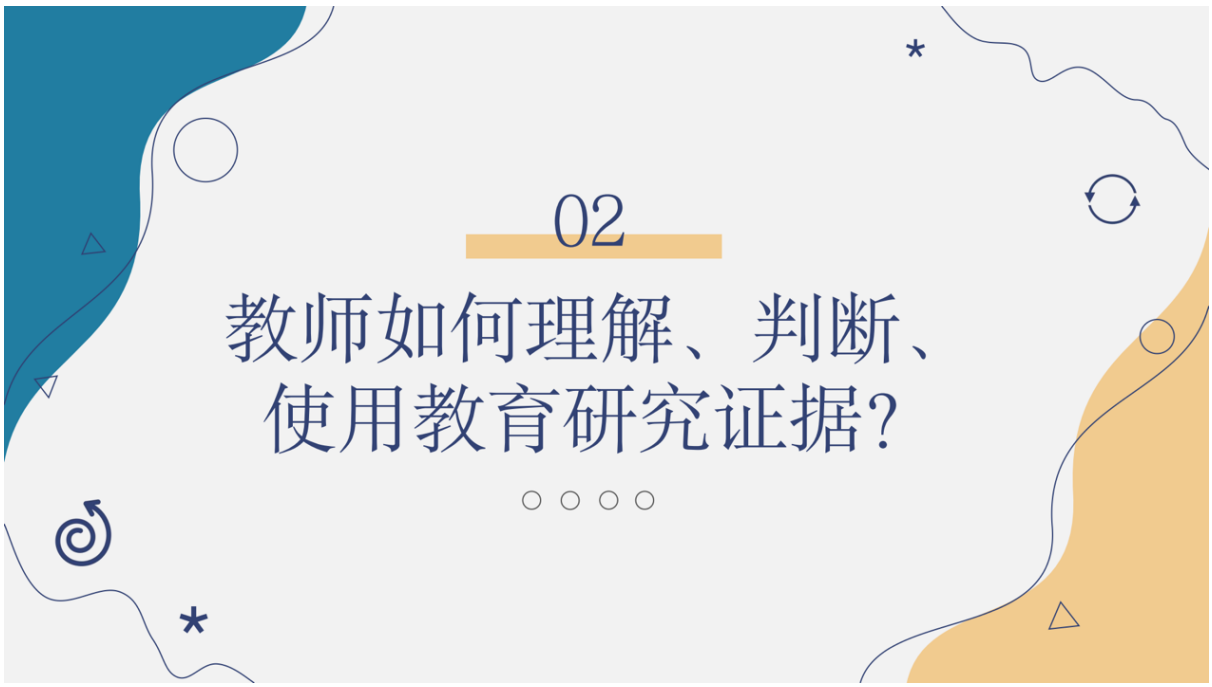
2024.2
- 参加主题讲座，学习：
- 教育研究证据基本知识
- 基于研究证据的教学策略

2024.3
- 深度学习基于研究证据的教学策略
- 制定基于研究证据的教学计划
- 参加公开课
- 尝试使用基于研究证据的教学策略
- 反馈、调整课堂情况

2024.4
- 日常使用基于研究证据的教学策略
- 打磨精品课

2024.6
总结与反馈


重要时间节点——成长型思维/提高自我价值感
成长型思维
参加主题讲座，学习：
• 教育研究证据基本知识
• 成长型思维
第一次成长型思维主题班会
参加读书会，了解成长型思维在教学上的使用
第二次成长型思维主题活动
总结与反馈
2024.2
2024.3
2024.4
2024.5
2024.6
参加主题讲座，学习：
• 教育研究证据基本知识
• 提高自我价值感课堂操作
第一次提高自我价值感写作练习
• 第二次提高自我价值感写作练习
• 总结与反馈
提高自我价值感


02
教师如何理解、判断、使用教育研究证据？

使用研究证据的两个维度

研究证据

- 研究证据是指从一项有正式研究问题、系统性设计、由研究者操作的研究中得到的研究成果，在教育学中这样的研究目的在于测评某些项目或教学方法的有效性。

	被动传输	参与传输	(交互)主动传输
原始型证据	例如，开放获取的期刊	例如，期刊俱乐部	例如，使用者行动实验
修正型证据	例如，EEF教学工具包	例如，智囊	例如，研究型学校
设计型证据	例如，课例	例如，帮助热线	例如，全民范围内立法

如何寻找研究证据?

CNKI中国知网 www.cnki.net 中国知识基础设施工程　万方数据 WANFANG DATA　VIP 维普资讯　Google 学术搜索

研究搜索引擎/数据库：
中国知网、万方数据知识、维普网、谷歌学术等

教育学研究期刊：
教师教育研究、上海教育科研、中国教育学刊、
Review of Educational Research、Journal of Teacher Education等

国际教育测评报告：
PISA、TIMSS、PIRLS等

其他教育证据网站：
What Works Clearinghouse、 EEF Teaching and Learning Toolkit



PIRLS USA　TIMSS USA　PISA OECD　ies WHAT WORKS CLEARINGHOUSE　Education Endowment Foundation

如何评价研究的质量?

Gorard
"SIEVE"

- 研究的质量可以从研究设计、样本数量、缺失样本以及测评方法四个维度进行评价
- 从左到右，研究设计对研究的质量最为重要

研究设计	样本数量	缺失样本	测量方法
设计合理的对照组：随机控制实验>准实验>前测与后测>个案分析	实验中样本数量越大，实验结果越能够普及运用到实际教学中。	实验中缺失的样本数量越小，实验结果越可靠	准确的测量：统一的、标准的测试或量表>研究者设计的测试或量表>访谈

评价研究质量示例

为了探究分级阅读对学生语文阅读能力的影响：

① 将30个班级的学生以班级为单位随机分为两组，其中一组学生参加语文阅读能力测试，教师根据学生的阅读能力推荐学生阅读不同层级的书。一段时间后，由研究者设计考试试题，对比两组学生的考试成绩。
② 在一所学校使用分级阅读相关软件，一段时间后，采访40名学生对于分级阅读的看法。
③ 在一所学校中筛选出一部分班级作为实验班，使用分级阅读软件，对比实验班和对照班使用分级阅读软件后的语文期末考试成绩。
④ 在一所学校中筛选出一部分班级作为实验班，使用分级阅读软件。实验开始前后，利用问卷收集实验班和普通班学生对自己阅读能力的评价。

研究设计最好的是?　　测量方法最好的是?　　研究结果最可信的是?

如何理解研究的效应量?

常见的研究效应量

Cohen's d 效应量
Pearson r 相关系数
标准差
...

什么是效应量?

- 效应量是用于衡量两组之间差异的一个统计指标
- 效应量通常用于评估研究结果的重要性

效应量越大意味着什么?

- 在实验环境下
- 平均来看，实验组学生的成就比对照组学生的成就更大

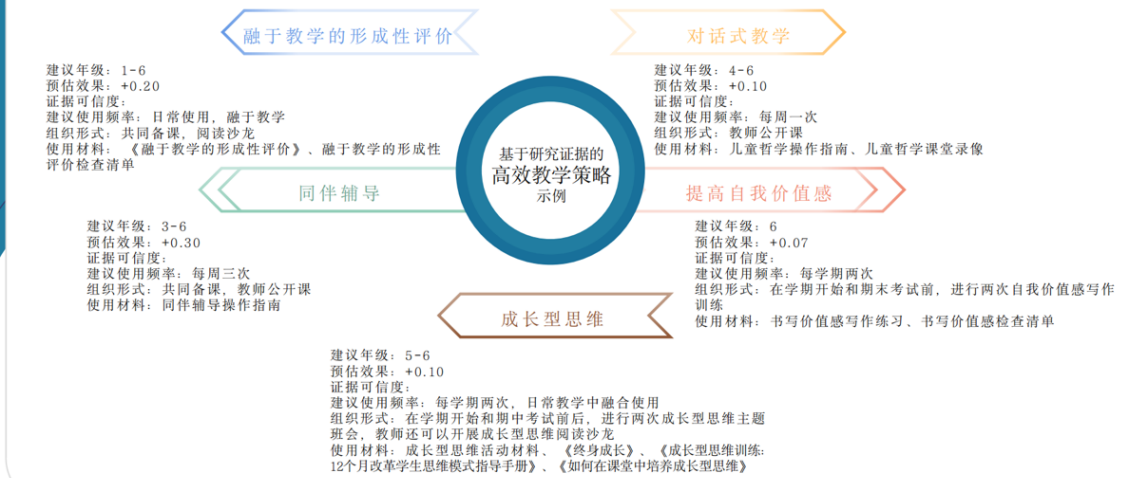
效应量和研究质量有什么关系?

- 效应量和研究质量没有必然联系
- 效应量大不代表研究质量好
- 高质量的研究也不能保证可观的效应量

寻找、总结研究证据流程



有哪些是以教育研究证据为导向的教学实践？



03

教师如何实施研究证据导向的教学策略？

融于课堂的形成性评价

一篇近期发表的系统性文献综述通过总结归纳48篇论文，发现形成性评价对学生阅读能力的平均效应量为+0.18，另一篇基于美国近年研究的文献综述发现了更大的效应量，为+0.33。2018年一个大规模随机控制实验显示，Dylan Wiliam的融于课堂的形成性评价对学生在GCSE标化考试成绩的效应量为+0.11。

Report	Year group	Duration	Outcome	Security	Effect size
Speckesser, S., et al. (2018)	Year 10, U.K.	2 years	GCSE	4	+0.11

若教师、学习者或其同伴引出，解释和运用关于学生成就的证据，以做出关于下一步教学的决策，而这种决策比缺乏证据时做出的决策可能更好或更可靠，那么这种评价就在形成性地运行。

“谁”在实施评价

收集证据的目的

重点：下一步教学的科学决策

焦点：引发的行动，而不是意向

	学习者要去哪里	学习者当前在哪里	如何去那里
教师	明晰、分享和理解学习目标与成功指标	设计有效的讨论、任务和活动，引出学习的证据	提供促进学习的反馈
同伴		激发学习者成为彼此的教学资源	
学习者		激发学习者成为自己学习的主人	

同伴辅导

过往很多研究都评估了同伴辅导的效果，但大多数研究都局限于个案分析或缺乏对照组，以下的两个实验都基于随机对照试验。
第一个研究包含了125所苏格兰小学的3000名四年级和六年级的学生，在两年间随机参加了同龄同伴辅导和混龄同伴辅导。在第二个研究中，118位四年级学生被随机分配到同伴互助阅读和对照组。结果显示，同伴辅导对辅导者和被辅导者均存在积极影响。

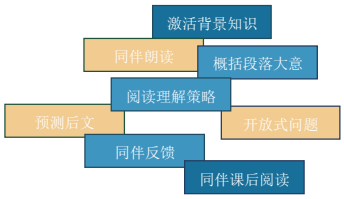
Report	Year group	Duration	Outcome	Evidence strength	Estimated impact
Topping et al. (2012)	Year 4 and Year 6	30-minute sessions, 1 or 3 times per week, for a duration of 15 weeks	Reading	3	+0.31 First Year of Implementation +0.34 Second Year of Implementation
Kung & Teo (2022)	Year 4 (Taiwan)	40-minute sessions, twice a week, for a duration of 15 weeks	Reading	2	+0.07 Vocabulary +0.11 Reading Comprehension

设定目标
同伴辅导的目的是什么？现在的学情是怎样的？

组织与匹配
根据学生的阅读能力配对，同伴之间的阅读能力要有差距，但是差距不能过大。学生可以在一节课结束后或同一节课中交换角色，根据上课需要进行调整。



监控与反馈
小组巡视、观察，并在必要时给出指导。定期与学生反馈总结同伴辅导的效果。



书写价值感

自我价值感干预在美国已经被广泛评价，证据表明，这项干预可以在阅读及其他多个学科产生长期的积极效果，提高学习成绩，特别是针对少数民族学生和贫困学生等处境不利的学生（例如，Wu et al., 2021）。一些研究还表明，这些效果一直持续到中学，甚至大学（Borman et al., 2021；Goyer et al., 2017）。在英格兰进行的一项大型随机控制实验表明，自我价值感干预对处在不利社会经济地位学生的成绩尤其有效（See et al.，2022）。该实验囊括了英格兰地区11000名14到16岁的学生，在经过两次写作任务后，实验组学生的在标准化考试中的整体成绩有明显提升（ES = +0.05）。

Report	Age Group	Duration	Outcomes	Evidence strength	Estimated impact
See et al. (2022)	14-16	One year	Overall Attainment	4	+0.05

世界上有很多有意义的事情，这些事情会让你体验到自己的重要性或是与众不同，从而丰富你的生活和阅历。

请在下列关键词中选择两个你认为重要的词，在空白处写一写你在生活中是怎么做到的。

关键词：幽默、同情心、诚实、友谊、决心、勇气、信任、高兴、坚韧、好奇、知识、名誉、成就

这个练习需要你表达你的想法和感受，在写作的时候不用担心字数限制、错别字和病句。

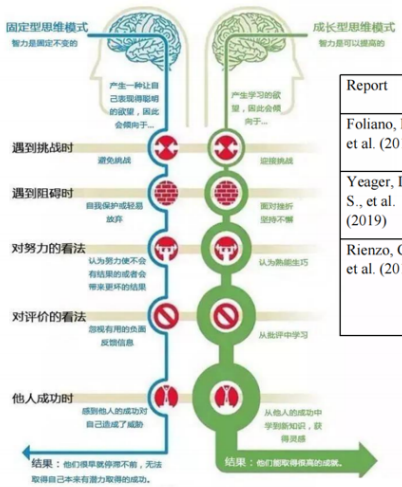
世界上有很多有意义的事情，这些事情会让你体验到自己的重要性或是与众不同，从而丰富你的生活和阅历。

例如，有些人认为诚信很重要，因为他人会相信他们。有的人认为家庭很重要，因为他们珍重他们的家人。还有些人重视体育运动，因为运动时他们可以突破自己、体验快乐。

请在下方的空白处写一件你生活中重要的事。展开写写这件事有多重要，为什么重要，对你的生活意味着什么？

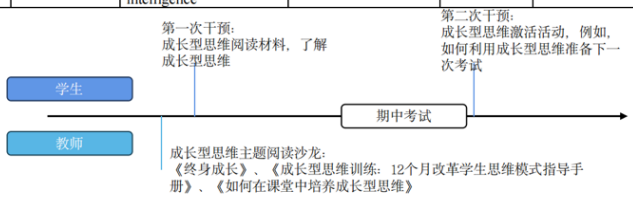
这个练习需要你表达你的想法和感受，在写作的时候不用担心字数限制、错别字和病句。

成长型思维



Foliano（2019）和Yeager（2019）开展了两项大规模随机对照试验，以评估成长型思维模式的有效性。Yeager的项目包括两节25分钟的课程，而Foliano的工作则为学生提供了更多的研讨会，并培训教师在日常实践中使用成长型思维模式，Rienzo（2015）也进行了随机对照试验，但样本量较小，流失率较高。这些研究证明，成长型思维模式是一种很有前途的提高学生阅读成果的方法。不过，由于Yeager是Dweck研究团队的成员，因此这些证据并不够令人满意。

Report	Year group	Duration	Outcome	Evidence strength	Effect size
Foliano, F., et al. (2019)	Year 6, U.K.	6 workshops for students; 2 half-day training for teachers	Key stage 2 English	3	ES = +0.01
Yeager, D. S., et al. (2019)	Grade 9, U.S.	2 sessions, 25 min each	English GPA	3	Overall ES = +0.11, Low achieving ES = +0.33
Rienzo, C., et al. (2015)	Year 5, U.K.	6 workshops and 4 further sessions for students about malleability of intelligence	Progress in English (PiE)	1	Overall ES = +0.18, Low-achieving ES = +0.34




提问环节

Appendix D. Implementation guide for evidence-informed strategies

D.1 Embedding formative assessment supplementary materials

This document provides checklist, teaching plan sheet, and student survey to help you understand and use embedding formative assessment. You can adapt these materials to your teaching practice. Additionally, you can read *Wiliam, D. (2019). Embedding formative assessment* for reference.

1. Embedding formative assessment checklist

Use the following checklists to check whether you correctly understand the strategy and use it correctly. You can use it after reading the strategy or before using the strategy to ensure correct implementation. These checklists may also help you to reflect on your practice after the class.

Reflection checklist for strategy 1: clarifying, sharing, and understanding learning intentions and success criteria

	I don't do this	I do this sometimes	This is embedded in my practice	I could support someone else
I know what the learning intention of the lesson is.				
I keep the learning intention and success criteria for a lesson context free.				
I communicate quality by using at least two pieces of anonymous work.				
At the end of a lesson I sometimes ask my students what they have learned.				
I use rubrics to discuss quality with my students.				

Reflection checklist for strategy 2: engineering effective discussion, tasks, and activities that elicit evidence of learning

	I don't do this	I do this sometimes	This is embedded in my practice	I could support someone else
I find out what every student knows at least once a lesson, by using an all-student response system.				
I make 'no hands up' a standard classroom policy.				
I ensure that all students have time to think about an answer to a question I pose before I choose who answers.				
I ask a hinge question during a lesson when I need to decide whether I could move on.				

I use exit cards or another way of collecting extended responses from students.				
I use statements rather than questions to encourage more thoughtful answers.				

Reflection checklist for strategy 3: Providing feedback that moves learning forward

	I don't do this	I do this sometimes	This is embedded in my practice	I could support someone else
I focus on student response to my feedback rather than the feedback itself.				
I give a student a scaffold response if unable to answer my question, but then I come back to that student.				
I praise students for effort rather than ability.				
I give students task-involving feedback rather than ego-involving feedback.				
I give students challenging tasks that are achievable with effort, and make it clear that failure is not only acceptable but also expected.				
I use comment-only grading.				
I provide time for students to adjust based on feedback.				

Reflection checklist for strategy 4: activating students as learning resources for one another

	I don't do this	I do this sometimes	This is embedded in my practice	I could support someone else
I have groups of students working cooperatively at times.				
Pairs of students look at anonymous work to comment on.				
My students use 'two stars and a wish,' or another technique, to give feedback to each other.				
I discuss ground rules of peer feedback with students.				
I get peers to mediate my feedback to students.				
I use group goals.				
I prioritize individual accountability in peer and group work.				

Reflection checklist for strategy 5: activating students as owners of their own learning

	I don't do this	I do this sometimes	This is embedded in my practice	I could support someone else
I use student self-reports but am careful to check on the accuracy of the reporting.				
Students self-assessment is a routine part of my work.				
Students have learning portfolios that focus on their progress.				
I use a framework to help my students see connections.				
I use learning logs to help my students reflect on their learning.				
I use both intrinsic and extrinsic orientations to motivate my students.				

2. Lesson planning sheet

The technique I am going to use:
Why I am planning to use it and the results I am hoping for:
Class and date:
Preparation for the lesson:
What I am going to do less of:
Reflecting on how the technique worked, including evidence to support my claims:
What I am going to do next:

3. Peer observation sheet

Class to be observed:
Technique to be observed:
What I want my peer to comment on:
Peer's comments:
Reflections after reading peer's comments and talking through the observation:
What I will do next:

4. Student survey

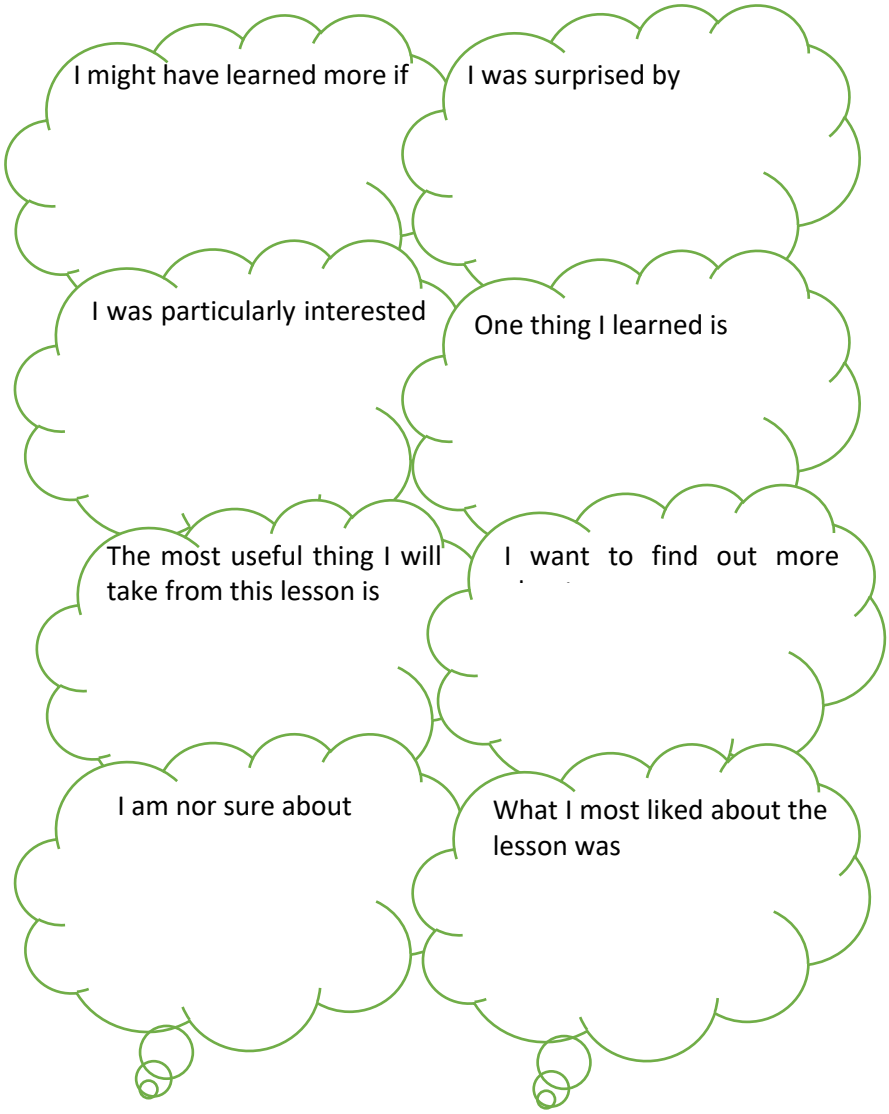
Name:	Date:
Circle one number for each line to show me what you felt about this lesson	
The pace	Slow 1 2 3 4 5 Fast
The difficulty	Easy 1 2 3 4 5 Difficult
My interest	Low 1 2 3 4 5 High
My understanding	Low 1 2 3 4 5 High
My learning	Poor 1 2 3 4 5 Good

5. Student feedback to teacher (anonymous)

	Never	Sometimes	Often
Does your teacher tell you what is expected before you start an assignment?			
Do you compare anonymous pieces of work to understand what makes a good assignment?			
Does your teacher ask what you have learned at the end of a lesson?			
Does your teacher choose students at random to answer questions?			
Does your teacher wait three seconds after asking a question to give everyone a chance to think?			
Does your teacher wait three seconds after a student answer a question to allow that students to give a fuller answer?			
Does your teacher ask all students to answer a question at the same time?			
When your teacher grades an assignment, do you get comments that help you improve?			
Do you give feedback to other students?			
Do you work with others in a group to try to help everyone improve?			

6. My learning log

Name: Date:



7. Student reflections on learning

	Disagree	Sometimes	Agree
I believe that I can learn a lot.			
I learn a lot in class.			
My teacher believes that I can learn a lot.			
I find it hard when other students look at my work.			
My classmates help me learn.			
When I work hard I get smarter.			
My teacher helps me learn.			
I enjoy helping other students learn.			

Here is the Chinese version which was used in the experiment.

以下核对清单旨在检查您对五大形成性评价策略的理解和应用是否正确，您可以在阅读完策略之后或使用某策略前进行核对，以确保正确的理解，也可以在课后根据以下清单反思教学。

策略 1:明晰、分享和理解学习目标与成功指标

	我没有这样做	我偶尔这么做	这已经融入了我的教学	我可以帮助别人做
我很清楚本节课的学习目标。				
我会向学生分享学习目标和成功标准。				
我用至少两本匿名的作业和学生展开优缺点讨论。				
下课后，我会询问学生学到了什么。				
我根据评分标准和学生探讨高质量的作业要关注什么。				

策略 2: 设计有效的讨论、任务和活动，引出学习的证据

	我没有这样做	我偶尔这么做	这已经融入了我的教学	我可以帮助别人做
我每节课至少使用一次全学生应答系统，以了解所有学生掌握情况。				
我采用'不举手'策略。				
学生在回答问题之前，我会预留充分且适当的等待时间。				
我会提出一个连接点问题来检查学生是否为下一步学习做好准备。				
我使用出口通行证等技巧收集学生的学习情况。				
我使用论断代替提问，从而引发学生的深度思考。				

策略 3:提供促进学习的反馈

	我没有这样做	我偶尔这么做	这已经融入了我的教学	我可以帮助别人做
我关注学生对反馈的反应，而不仅限于反馈本身。				
我会给遇到问题的学生支架式反馈，之后再回来帮助他。				
我赞赏学生的努力，而不是他们的能力。				
我给学生的反馈只关于任务的完成，而与他们个体能力、品质无关。				
我给学生布置挑战性任务，只要努力就能完成，并明确指出，失败是可以接受的。				
我的反馈以评语的形式而不是分数的形式出现。				
我在课堂上留出时间以运用反馈改进学生的学习.				

策略 4:激发学习者成为彼此的教学资源

	我没有这样做	我偶尔这么做	这已经融入了我的教学	我可以帮助别人做
我有时会让学生分组合作学习。				
我鼓励学生与邻座交换作业，互相评分。				
我的学生使用'两颗星加一个愿望'等方式给他们的伙伴提供改进建议。				
我和学生讨论了同伴互评的基本规则。				
我设立小组共同目标。				
在同伴和小组工作中，我同样强调对个人责任的需求。				

策略 5:激发学习者成为自己学习的主人

	我没有这样做	我偶尔这么做	这已经融入了我的教学	我可以帮助别人做
我会使用学生自我评价表，但是我会检查报告的精确度。				
学生自我评价是我的日常工作之一。				
学生们用学习档案袋记录自己的进步。				

我使用一个框架来帮助我的学生发现其中的联系。				
我使用学习日志，鼓励学生反思。				
我使用内部和外部因素提高学生的积极性。				

课堂计划表 Lesson planning sheet

班级： 日期：
我使用这一技巧的原因和目的：
我将要使用什么技巧：
教案：
教学反思
我应该避免什么？
反思该技巧如何发挥作用，写出支持以上陈述的证据：
下一步我将做什么：

听课记录表（授课教师版）Peer observation sheet

班级： 日期：
使用技巧：
课堂记录：
评价：
建议：

学生问卷 Student survey

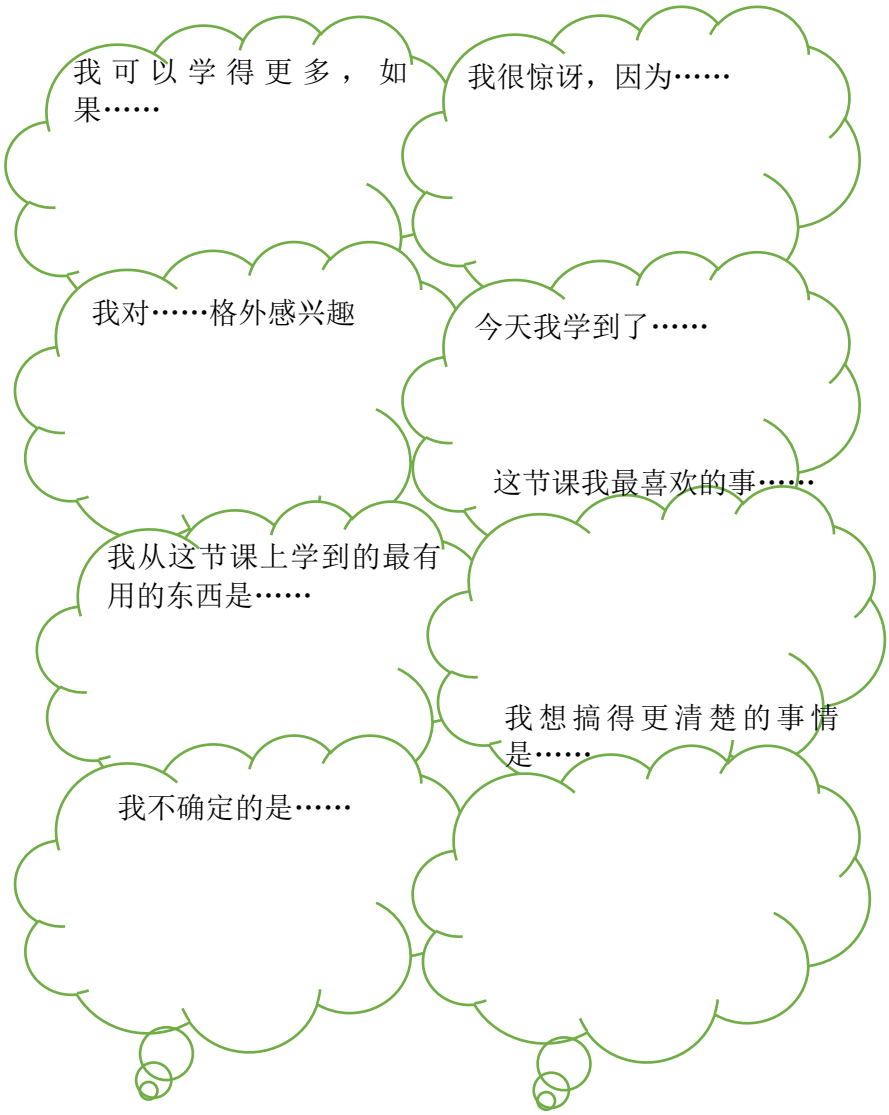
姓名：	日期：
在每一行中圈出一个数字，反馈你对本次课的体验	
课堂进度	慢 1 2 3 4 5 快
课程难度	易 1 2 3 4 5 难
我的兴趣	低 1 2 3 4 5 高
我的理解度	低 1 2 3 4 5 高
我的学习情况	差 1 2 3 4 5 好

学生反馈表 Student feedback to teacher (anonymous)

	从不		有时		总是
在做作业之前，老师会告诉你作业的要求	1	2	3	4	5
通过比较匿名作业，你明白什么是高质量的作业？	1	2	3	4	5
一堂课结束时，老师会问你学到了什么	1	2	3	4	5
你的老师会随机选择学生回答问题	1	2	3	4	5
你的老师在提问后会停顿三秒钟，给每个学生思考的机会	1	2	3	4	5
你的老师会在同学回答问题后等待三秒钟，鼓励他给出更完整的答案	1	2	3	4	5
你的老师要求所有同学一起回答同一个问题	1	2	3	4	5
老师批改作业时会留下有助于你进步的评语	1	2	3	4	5
你会给同学提供反馈	1	2	3	4	5
你是在小组中与同学积极合作，努力帮助每个人进步	1	2	3	4	5

学习日志 My learning log

姓名 Name: 日期 Date:



学生自我反思 Student reflections on learning

	不赞同				赞同
我相信我可以学很多。	1	2	3	4	5
我在课堂上学了很多。	1	2	3	4	5
我的老师相信我可以学习。	1	2	3	4	5
当其他同学看我的作业时，我觉得很难受。	1	2	3	4	5
我的同学帮助我学习。	1	2	3	4	5
学习越努力，我就会越聪明。	1	2	3	4	5
我的老师为我的学习提供帮助。	1	2	3	4	5
我享受帮助其他同学学习的过程。	1	2	3	4	5

D.2 Peer tutoring supporting document

This document provides the organisation of peer tutoring in your classroom, recommended activities, teacher observation form, student self-evaluation form, activity sheet, teaching plan sheet, and example lessons. You can adapt these materials to your teaching plans.

1. Objective for peer tutoring

It is important to be clear about the objectives, since if it is not clear what outcome(s) you are hoping for, you will not be able to tell whether the project was a success or not. You should first set up a goal before implementing peer tutoring, such as improvements in reading/ accuracy/ comprehension/ fluency, improved attitudes towards reading, enhanced self-esteem, and wide reading materials. You should reflect on your teaching and adjust instruction of peer tutoring to get closer to their goals.

2. Organisation and matching

All students are paired in their class by first ranking them on reading performance and then splitting the ranked list in half. The top-ranked student in the stronger half is paired with the strongest reader in the weaker half. Next, second-ranked students in each half are paired. This matching process continued until all students have a partner. You are then advised to inspect the pairings to determine whether one or more were socially incompatible. If such a coupling was found, it was changed. Within each pair, the role of tutor and reader (tutee) was reciprocal; that is, each student in each pair served as reader for part of the time and as tutor for an equal amount of time. To better manage the class, students could exchange their roles after one lesson finished or during the same lesson. Pairs remained together for 4 weeks, after which you announced new pairings.

3. Implementation frequency

Peer tutoring should be implemented at least once a week. If teachers expect visible change of students' achievement, the frequency should be 3 times a week at the minimum of 15 minutes for 10 weeks.

Timing of the tutoring periods and scheduling the number of tutorial contacts throughout the week should be tuned to the objectives of your program. When aiming for complex skills, such as practising the use of various reading strategies for optimising reading comprehension, you would be better to include a longer session instead of two or three short sessions. This will allow children to get deeply involved in a text to practise the different reading strategies in combination, and to get as close as possible to authentic reading and the use of strategies in a relevant way.

4. Example activity

The following paragraphs introduce the process of a typical peer-assisted learning lesson involves following activities to support teaching. You could use a time alarm to remind students time of each session so that they could better control the pace of teaching.

1). Partner reading (PALS)

The reader (or tutee) reads aloud, receiving immediate corrective feedback if words are mis-pronounced. Tutors should correct word recognition errors, which include saying the wrong word, leaving out a word, adding a word, and pausing longer than 4 seconds (self-correction). Tutors are also expected to correct errors as they occur and to encourage the reader to reread the sentence with accuracy. The program calls for the stronger reader in each pair to read first, which is designed to provide an opportunity for the weaker reader in the pair to preview the passage and review difficult words before it is his or her turn to reread the same text. Students switch roles after five-minute blocks.

2). Paragraph 'shrinking' (PALS)

The reader states the main idea (i.e., who or what the passage is about), gives a 20-word summary of the passage, and provides a sequential retelling of the important events of the passage. Then teachers pick two to three pairs and have them read aloud their 20-word paragraph summary from the paragraph 'shrinking' activity. The presentation activity was intended to enhance the participating students' joint attention, as well as to allow them to receive more feedback during the activity process.

3). Activating background knowledge (One Book for Two)

It provides a good opportunity to activate students' prior knowledge and combine it with the reading material. Students are taught to deduce the subject of the text on the basis of the title and pictures, to think about what they already know about the topic, and to write down a number of key words. Students discuss in pairs or groups and exchange ideas.

4). Prediction relay/ Advancing questions (PALS)

In fiction materials, the reader predicts what is likely to happen next on the next page, reads aloud from the page, and summarizes the just-read text, with the tutor deciding whether the predictions are accurate. Students switch roles after five-minute blocks. For non-fiction materials or when students have already read the whole story, other challenging questions could be posed by teachers or students themselves for further discussion.

5). Reading comprehension strategies (RR & One Book for Two)

In group peer tutoring activity, teachers could guide students to use certain reading comprehension strategies to explore a question. In *One Book For Two*, students were taught the strategy of summarising texts graphically by means of making a scheme such as sequence schemes, cause-effect schemes, and classification schemes. In *Reciprocal Reading*, four evidence-based strategies: predicting

(to make and explore inferences), clarifying (to explore words and phrases and evaluate critically), questioning (to focus on the main ideas), summarising (to monitor understanding) are recommended during 20 minutes.

6). Group goals in peer tutoring (Classwide peer tutoring & Best composite answer)

The teacher could organise individual class members into tutor student pairs who work together on two competing teams. Students earn points for their team by responding appropriately to the tasks. Tutors earn points from the teacher based on their implementation of the tutoring role. If the teacher does not want to identify tutor and tutee, the competition could be between groups. For example, the teacher create a group goal—all students are working towards a single, best group response.

7). Peer feedback (Embedding formative assessment)

Students can learn from each other. Peer feedback consists of eight steps:

Students answer a question individually and write down reasons for their response. → The teacher show students three attempts at a similar question (excellent/ medium/ bad). → The teacher asks students which answer is excellent and its reasons (records on the board). → Group work: create 3 possible solutions to the question together at different levels of quality. → Group presentation. → Peers assess the original solution. → Study plan (individual/ peer). → Final answer

8). After class paired reading (Paired Reading)

This activity could also be held after class. Students could use 'Five Finger Test' to select reading material. The difficulty of material should be between the reading level of the pair. The 'Five Finger Test' is as following:

Open a page at random; Spread five fingers; Place fingertips on the page at random; One student attempts to read the five words; Repeated on another four pages; If he/she has struggled on more than one word, the level of book is higher than his/her reading level.

Pairs could:

Talk about the pictures; Talk about interesting words or ideas or events; Activating prior background knowledge; Talk at some natural break; Review the main ideas or events at the end of chapters and at the end of the book.

5. Monitoring and feedback (Topping)

The peer tutoring session is not an opportunity for you to 'get on with some marking'. On the contrary, you should either be setting a good example by reading silently or with a tutee, or should be circulating round the group, observing and guiding children as necessary. You could create a list of students who may need extra help and monitor. Each lesson is expected to provide an activity sheet to introduce the main questions to focus on during peer tutoring (or at least questions to discuss should be shown

on the electronic blackboard). You and your students could use the following *Teacher observation form* and *Student self-evaluation form* to record and evaluate the process. If possible, you, together with students, could have feedback meetings to communicate.

6. Other resources

https://www.peertutoringresource.org/

7. Student self-evaluation form (revised from Reading in Pairs)

Self-evaluation Reading in Pairs Date: Pairs:		
Aims and proposals from the last-evaluation session (short comment about your improvements)		
Before reading, we…	Need improvement	Good
Explore the characteristics of the text.		
Make predictions about the contents of the text.		
Talk about what we know from the topic.		
Talk about possible difficulties.		
When [Name 1] reading,	Need improvement	Good
[Name 1] reads correctly.		
[Name 2] listens and follows the text.		
[Name 2] appreciates the [Name 1]'s reading.		
When [Name 2] reading,	Need improvement	Good
[Name 2] reads correctly.		
[Name 1] listens and follows the text.		
[Name 1] appreciates the [Name 2]'s reading.		
While reading together	Need improvement	Good
We agree on how to read before starting.		
We read together and, at the same time, listen to each other.		
We listen to each other and waits for the one who read slower.		
PPP reading	Need improvement	Good
When reading, we notice mistakes.		
When reading, we notice mistakes and correct them.		
While the [Name 1] is reading and makes a mistake, [Name 2] stops him or her (Pause) and gives clues.		
While the [Name 2] is reading and makes a mistake, [Name 1] stops him or her (Pause) and gives clues.		
We encourage each other to correct mistakes and go on reading.		
Reading comprehension	Need improvement	Good
We answer the questions together.		
We share ideas and our previous knowledge to make good answers.		
We all try to answer the questions.		
[Name 1] gives clues and helps [Name 2] with the answers.		

[Name 2] gives clues and helps [Name 1] with the answers.		
At the end we talk about the main ideas of the text.		
Things we can improve for the next sessions		
[Name 1]		
[Name 2]		

8. Teacher observation form

Teachers are suggested to have a list of students who may need extra help during peer tutoring. The following form could be used when teachers observe listed students.

Group number		1	2	3	4	5	6
Group member							
Pairs	The control time						
	They detect their own lack of knowledge						
	They acknowledge mistakes						
	They feel comfortable						
	They are interested in the task						
	They identify the meaning of new words						
	They activate previous knowledge						
	They don't look shy						
	Communication between them is fluid						
	They keep eye contact						
Teaching tasks	Correct pronunciation of new words						
	Fluent reading						
Further comments: Teacher's role, materials, organisation							

9. Teacher Checklist

	I don't do this	I do this sometimes	Embedded in my practice	Support someone else
I have groups of students working cooperatively at times.				
Pairs of students look at anonymous work to comment on.				
My students use '…[] technique' to support each other.				
I support some students with sentence starters.				
I discuss ground rules of peer tutoring with students.				
I use group goals.				
I prioritise individual accountability in peer and group work.				
Other techniques for this strategy that I used to improve student learning:				

10. Lesson planning sheet

The technique I am going to use:
Why I am planning to use it and the results I am hoping for:
Class and date:
Preparation for the lesson:
What I am going to do less of:
Reflecting on how the technique worked, including evidence to support my claims:
What I am going to do next:

11. Activity sheet

Name:

Chapter or book title:

<table>
<tr><td colspan="2">Prediction: Before you begin to read the selection, look at the title or cover, scan the pages to read the major headings, and look at any illustrations. Write down your prediction(s).</td></tr>
<tr><td>Prediction:</td><td>Support:</td></tr>
<tr><td>Main Ideas: As you finish reading each paragraph or key section of text, identify the main idea of that paragraph or section.</td><td>Questions: For each main idea listed, write down at least one question.</td></tr>
<tr><td>Main idea 1:

Main idea 2:

Main idea 3:

Main idea 4:

Main idea 5:</td><td>Question 1:

Question 2:

Question 3:

Question 4:

Question 5:</td></tr>
<tr><td colspan="2">Summarise: Write a brief summary of what you read.</td></tr>
<tr><td colspan="2">Clarify: Copy down words, phrases, or sentences in the passage that are unclear. Then explain how you clarifiied your understanding.</td></tr>
<tr><td>Word or Phrase:</td><td>Clarify:</td></tr>
</table>

Here is the Chinese version which was used in the experiment.

同伴辅导支持文件

Peer Tutoring Supporting Document

设立同伴辅导目标

Objective for peer tutoring

设立清晰的目标非常重要，如果不知道什么结果是被期待的，那么就无法判断一个项目成功与否。在实施同伴辅导之前，教师要设定一个目标，例如提高阅读能力/阅读准确度/阅读理解能力/阅读流畅度，亦或是培养积极的阅读态度，提高自我价值，甚至可以只是引导学生广泛地阅读。教师要在操作过程中不断反思、调整，从而接近目标。

组织与配对

Organisation and matching

首先教师根据学生的阅读能力配对，将学生成绩从平均线分成两组，学习小组的成员之间阅读能力的差距不宜过大。因此我们建议将平均线以上最优的学生和平均线以下最优的学生配对，以此类推。教师要在教学过程中检查学生的合作效果，通过进一步调整，找到最合适的小组配对。每对学生中，'老师'和'学生'的角色是对等的；也就是说，每对学生在部分时间内担任学习者，在同等的时间内担任辅导老师，学生可以在一节课结束后或同一节课中交换角色。配对的学生在一起学习 4 周，之后教师宣布新的配对。

实施频率

Implementation frequency

每周应至少开展一次同伴辅导。如果教师希望学生的成绩有明显的变化，那么每周要开展 3 次同伴辅导活动，每次至少 15 分钟，持续 10 周。

同伴辅导时间的安排和一周内的辅导次数应与课程目标相适应。如果目标是培养复杂的技能，如练习使用各种阅读策略来优化阅读理解能力，最好安排较为持续的辅导，而不是简短的辅导活动。这样，孩子们就能深入阅读课文，综合练习不同的阅读策略，尽可能接近真实的阅读，并以相关的方式使用技巧。

活动案例

Example activity

以下的段落将介绍一些典型的同伴互助学习活动。对于任何活动，教师注意控制时间和课堂进度（例如在白板上显示计时器）。

A. 同伴朗读 Partner reading

结对的两位同学分别朗读课文（通常是一两段，根据文本长短安排），在朗读的过程中，另一位同学要及时给予同伴反馈。倾听的一方要纠正同伴朗读的错误，例如错字、漏字、增字、停顿过长（超过 4 秒）等。倾听方要及时指出错误并鼓励朗读方准确地重新朗读。这个教学方法建议两个人中阅读水平较强的学生先读，这样可以给予阅读水平较弱的学生提供示范。在聆听同伴朗读的过程中，阅读水平较弱的学生可以复习段落里较难的词语。

B. 概括段落大意 Paragraph ‘shrinking’

学生总结段落大意（例如，这段话和谁有关系，发生了什么），在小组内部进行讨论，最后用 20 个字概括段意，复述故事的情节，教师可以选两到三组学生分享他们的答案。这样可以提高参与学生的注意力，这也是一个很好的给予学生反馈的时机。

C. 激活背景知识 Activating background knowledge

学生们根据文章的标题、配图来推论文章的主题，激活背景知识并将知识和阅读材料结合起来。学生们先独立思考，记录下关键词，之后在小组讨论并交流文章的背景知识。

D. 预测后文/ 挑战性问题 Prediction relay/ Advancing questions

对于虚构型文本，一名学生先预测接下来一段的内容，之后朗读并概括段落，另一名学生判断他的伙伴预测得是否准确。五分钟后学生对调角色，预测之后段落的内容。对于非虚构文本，或者有些学生在课前已经通读了整篇文章，可以由教师或学生提出挑战性问题，并在小组内讨论。

E. 阅读理解策略 Reading comprehension strategies

在小组讨论过程中，教师可以指导学生用某些阅读理解策略来探讨问题。例如，*One Book For Two*（两人共读一本书）活动教会了学生使用顺序图、因果图和分类图等方案，图解性地总结文本的策略。*Reciprocal Reading*（相互阅读）推荐学生在小组讨论的 20 分钟内使用四种策略，包括：预测（推断与探索），阐述（词语和短语，批判性评价）、提问（聚焦主要观点）、总结（理解）。

F. 小组目标 Group goals in peer tutoring

教师可以把学生划分成’老师’组和’学员’组，当学员完成了他们的任务，就可以为’学员’组加分，当’老师’给予’学员’适当的指导，担任起小老师的责任时，可以为’老师’组加分。或者，竞争也可以存在于学习小组之间。例如，老师提出一个小组学习问题，所有的学生在小组里讨论以获得最佳答案，根据小组表现给各个小组评分。

G. 同伴反馈 Peer feedback

同伴是最好的学习资源，同伴反馈通常有八个步骤：

学生自主写下问题的答案→教师展示'好、中、差'三个示范答案→教师提问哪一个答案是好答案，并询问原因→小组合作：写出三个不同评级的答案→小组展示→同伴互改→最终答案

H. 课后组队阅读 After class paired reading

同伴阅读活动可以在课后进行。学生用'五根手指测试'来筛选阅读材料，这个材料的难度应该介于同伴的阅读水平之间。'五根手指测试'如下：

随即打开一页；张开五根手指；随机点在页面有字的地方；读出选中的汉子；重复四次；如果他（她）发现了一个以上的生字，那么这本书就比较困难了。

同伴之间可以：

描述插图；谈论词语的含义；交流背景知识；回顾整个故事的主题。

监控与反馈

Monitoring and feedback

同伴辅导课不是教师 '批改作业 '的机会。相反，教师应以身作则，默读或与被辅导者一起阅读，或者在小组内巡视，观察并在必要时指导孩子们。教师可以列出一份可能需要额外帮助和监督的学生名单，在小组活动期间，关注他们的行为。每节课都应提供一张活动表，展示同伴辅导的核心问题（或至少应在白板上显示讨论的问题）。教师和学生可使用以下教师观察表和学生自我评价表来记录和评价这一过程。如有可能，教师和学生可以召开反馈会议进行交流。

学生自评表

同伴朗读自评表/Self-evaluation Reading in Pairs 日期： 同伴：		
阅读目标（说说你的进步）		
在朗读之前，我们……	待提升	要保持
探索生字、词语 预测文章内容 谈论我们已经知道什么 谈论可能的困难		
当【同学 1】朗读时	待提升	要保持
【同学 1】正确地朗读 【同学 2】认真倾听 【同学 2】为【同学 1】的朗读点赞		
当【同学 2】朗读时	待提升	要保持
【同学 2】正确地朗读 【同学 1】认真倾听 【同学 1】为【同学 2】的朗读点赞		
共同朗读时	待提升	要保持
我们对朗读任务的划分达成一致 我们一起朗读时也在互相倾听 我们互相倾听并等待读得较慢的一方		
互助朗读	待提升	要保持
在朗读时，我们注意到了错误 在朗读时，我们注意到错误并及时改正 当【同学 1】朗读出错的时候【同学 2】制止了他并给一些提示 当【同学 2】朗读出错的时候【同学 1】制止了他并给一些提示 我们互相鼓励纠正错误，并继续朗读		

阅读理解	待提升	要保持
我们一起回答问题 为了做出最好的回答，我们分享想法和知识 我们都努力尝试回答问题 【同学 1】给【同学 2】提示和帮助 【同学 2】给【同学 1】提示和帮助 最后我们探讨了文章的主题		
下次我们可以做得更好		
【同学 1】		
【同学 2】		

教师课堂观察表

为了更好地帮助学生，教师可以提前设定好一个名单，记录可能需要额外帮助的小组。利用下面的表格，在课堂上观察他们课堂交流情况。

	组号						
同伴互助行为	他们控制讨论时间。						
	他们发现了他们不足的知识。						
	他们发现了错误。						
	他们的交流是自然的。						
	他们对任务很感兴趣。						
	他们知道了生词的含义。						
	他们激活了已有的相关知识						
	他们并不害羞。						
	他们的交流是流畅的。						
	他们有眼神交流。						
	…						
	…						
	…						
教学任务	生字、生词的正确发音						
	朗读流利						
	…						
	…						
	…						
其他课堂记录：							

教师检查清单

	我不使用	我偶尔使用	融入教学	我还帮助其他老师
这段时间内，我的学生合作学习了。				
学生们一起阅读、评价了匿名作业。				
我的学生用了【……】方法来互相支持。				
我提供给学生互助对话技巧。				
我和学生们一起讨论了同伴互助的规则。				
我使用了小组目标。				
我强调了同伴辅导的个人责任。				
我使用的其他策略：				

备课单

我将使用的教学技巧：
为什么我要使用这一技巧，以及我期待什么：
班级： 日期：
备课记录：
我下一次会减少：
反思技巧的使用过程，有哪些证据能证实以上观点：
我将如何改进：

预测后文活动表

姓名

书名/章节：

<table>
<tr><td colspan="2">预测：
在你阅读之前，先观察文章标题和插图，浏览文章的小标题，写下你的预测。</td></tr>
<tr><td>预测：</td><td>理由：</td></tr>
<tr><td>主旨：在你读完全文后，请概括段落大意。</td><td>提问： 针对每一个段落大意，提出至少一个问题。</td></tr>
<tr><td>主旨 1：

主旨 2：

主旨 3：

主旨 4：

主旨 5：</td><td>问题 1：

问题 2：

问题 3：

问题 4：

问题 5：</td></tr>
<tr><td colspan="2">总结：对你阅读的内容写一个简短的总结</td></tr>
<tr><td colspan="2">解释：在左边空格内记录你在阅读中有困难的词句，并在右边写下对他们的解释</td></tr>
<tr><td>词语/句子：</td><td>解释：</td></tr>
</table>

In addition, the pilot study provided two teaching plans for the main experiment.

A. 合肥市阳光小学：《小英雄雨来》第一课时教学设计

教材分析

这篇略读课文节选自作家管桦写的同名中篇小说，选入课文时有改动，讲的是在那战火连天、枪炮轰鸣的抗日战争时期，晋察冀边区的少年雨来，聪明勇敢，游泳本领高强，为了掩护革命干部，机智地同敌人作斗争的故事。课文以序号为标志自然分成六大部分。本文虽是略读课文，但由于篇幅较长，教学中要充分利用课文本身提供的有利条件，练习用比较快的速度阅读课文，培养学生的自读能力。

学情分析

这篇课文内容通俗易懂，故事性强，故事情节清晰，引人入胜。教学时，应放手让学生自读课文，在统揽全文的基础上，抓住重点，积极探究，把握课文主要内容。关于体会课文各部分内容之间的关系，目的是品味作者布局谋篇的匠心。所以要引导学生揣摩作者是如何布局谋篇的，感悟到写作文也要注意各段之间的关系。本文篇幅较长，因此要训练学生用较快的速度阅读课文，了解内容。抓住重点部分导读。教师应引导学生积极地与文本'对话'，扩大学生的生活视野，提升学生的人生境界，丰富学生的精神底蕴，确认价值指向的导读目标。引领学生在与文本的对话中，有思想的交流和碰撞，有情意的体悟与对流，有美感的激发与生成。

教学目标

1.认读 8 个字。

2.默读课文，给各部分内容加小标题，体会课文各部分内容与表现雨来品质的联系。（学习重点）

3.了解课文内容，学习雨来热爱祖国、机智勇敢的品质。（学习难点）

教学过程：

一、新课导入，激发兴趣

1.激发情感，参与预习。

同学们最喜欢读小英雄的故事，今天我们一起来学习一篇新的课文——《小英雄雨来》（板书课题），这个故事发生在抗日战争相持阶段，根据地的人民配合八路军对敌人展开了英勇的斗争，其中出现了无数的英雄人物和可歌可泣的事迹,雨来就是其中的一个。

2.理解课题：

小——指的是雨来的年龄小，是一个少年。

英雄——指的是雨来的特点。

雨来——课文中的主人公。

课题表现出文章的中心——英雄

3.引导学生质疑：看到课题后，你想知道什么？（为什么称雨来是小英雄？）（从课文中的哪些地方可以看出雨来是一位小英雄？）

二、初读课文，整体感知。

1．用自己喜欢的方式读课文，把课文读通顺。

2．用课件学习字词：个人读，同桌相互检查读

3．你还有什么不会的、不懂的字词，要向大家请教。（全班交流）

注意：①几个象声字的读音：’吧嗒吧嗒’’唔里哇啦’’哐啷’。

4.小组合作朗读课文，相互评价

5、思考：课文的主要内容是什么？让学生按以下内容填空。

本文讲述了在_____时期，晋察冀边区的少年_____为了掩护_____，机智地同_____作斗争的故事。

三、快速地读课文，看看课文一共分为几部分，请给每一部分加上小标题。（结合课件）

1.引导学生分部分阅读，每部分都给出阅读时间，促使学生提高阅读速度，并概括每部分的主要内容。

2.指导拟小标题：怎样编写小标题？

可以是句子，也可以是词语，可以用书上的语句，也可以自己概括。

3.学生合作拟小标题。

4.交流，注意引导学生使用不同的方法。

如：第一部分，自己概括，可用’游泳本领高’，用书上的话可用’望着妈妈笑’，如果用词语作小标题，可写’游泳’。

归纳主要内容，提炼小标题

第一部分：望着妈妈笑（或游泳）

第二部分：爱自己的祖国（或读书）

第三部分：夜归见闻（或见闻）

第四部分：掩护李大叔（或掩护）

第五部分：与鬼子斗争（或斗争）

第六部分：雨来没有死（或脱险）

五、作业

1.读书，体会小雨来的品质。

2.体会文章各部分之间的联系。

B. 合肥市屯溪路小学：《骑鹅旅行记》循证教育研究

同伴辅导设计

一、文本教学解读

本节课属于’整本书阅读’学习任务群。结合单元内容和语文要素，学生在单元学习中需要理解的大概念是：选择阅读适宜译本的外国文学名著可以促进整本书阅读经验的积累；参与读书共同体可以借鉴有益的阅读经验。本文重点是结合描写尼尔斯变成小狐仙后内心世界变化的句子，立体的多元的感受人物形象；借助目录里的标题，如’鹤之舞表演大会’或’大海中的白银’，猜测故事情节，产生阅读原著的兴趣。

《骑鹅旅行记》是瑞典女作家塞尔玛·拉格洛芙创作的一部童话作品，讲述了一个名叫尼尔斯的十四岁小男孩的故事。他调皮、贪玩、任性、懒惰，不爱学习，喜欢捉弄小动物，因为对小精灵不恭敬、出尔反尔，受到了惩罚，被小精灵变成了一只只有拇指般大小的小人儿，随后又为了保护自家飞走的雄鹅而被带上高空。从此，他骑在鹅背上，跟随大雁走南闯北，周游各地，历时八个月才返回家乡。在整个旅行中，尼尔斯不仅欣赏到瑞典的美丽自然风光，听到了许多古老的传说故事，增长了见闻，还饱尝了风险和苦难。他逐渐改正了自己的缺点，蜕变成了一个懂事、勇敢、善良、乐于助人的好孩子。

教材节选部分描述的是尼尔斯变成小人儿后去牛棚找小狐仙，他先遇到了麻雀，麻雀嘲笑他是’拇指大的小人儿’；接着是受过他欺负的鸡没完没了地追着他说’活该，活该’；随后尼尔斯遇到了猫，猫也不愿意告诉他小狐仙的去向，并且当恼怒的尼尔斯又要去揪猫尾巴时，平时温柔和善的猫却将他狠狠扑倒；吃尽苦头的尼尔斯又到了牛棚中，引得牛棚里吼声四起，牛都想好好教训下这个平日捣蛋的小男孩。又后悔又害怕的尼尔斯意识到院子里没有人愿意帮他找小狐仙，于是他无奈地爬上围墙，恰逢看到自己家的雄鹅要和大雁一起飞走，为了让家里不受损失，他一跃跳上鹅背，开启了此次旅行。

根据上述文本解读，本节课的学习目标设计如下：

1.通过用较快的速度默读课文，梳理古诗情节，初步了解课文的主要内容。

2.围绕节选内容，能借助学习可视化工具，表述出尼尔斯变成小狐仙后，他的世界发生的变化。

3.借助原著目录里的标题，展开想象，猜想情节，产生阅读原著的兴趣。

二、学习活动设计

板块一 了解相关信息，激发阅读兴趣。

1.谈话导入。

这节课我们一起走进另一部外国文学名著（出示《骑鹅旅行记》作品封面）。

2.师生交流，了解相关信息。

（1）教师介绍交流，激发学生的阅读兴趣。

书名《骑鹅旅行记》又名《尼尔斯骑鹅旅行记》，作者塞尔玛·拉格洛芙（1858-1940）是瑞典19世纪末新浪漫主义文学的代表。这部童话作品出版以后，被指定为全瑞典学童的必读课外读物，并被译成50多种文字流传世界。由于这部作品的成功和广为流传，许多喜欢这部作品的读者甚至从丹麦、挪威和德国来看望作者拉格洛芙。瑞典文学院将1909年的诺贝尔文学奖授予了她。在瑞典，最重要的儿童文学奖就是用尼尔斯这个名字命名的。

板块二 品读节选内容，感受人物形象

1.自主阅读，明确学习要求：文中出现了哪些'人'物？

2.请四人小组选择其中一个角色，完成以下任务：

（1）请自由读小组选定的内容。

（2）挑选大家认为感兴趣的段落进行朗读。要求：组长先读，其他成员依次读，做到认真倾听，互帮互助。

（3）小组成员试着概括这部分内容。要求：20字以内，组长先说，其他成员再说，做到认真倾听，互帮互助。

3.批注阅读，体会尼尔斯的世界发生了什么变化，有理有据地说说尼尔斯给自己留下了怎样的印象。

板块三 猜想故事情节，产生阅读原著兴趣

1.运用策略，大胆预测。

（1）使用预测策略，结合节选片段，大胆猜测尼尔斯'变形'前是个什么样的人。

预设：

尼尔斯经常欺负小动物，调皮捣蛋，还经常惹父母生气。

（2）使用预测策略，结合节选片段，大胆猜测跟随大雁进行了一段神奇的旅程之后，尼尔斯会有哪些变化。

2.拓展资源，验证猜测。

（1）出示原著中描写尼尔斯'变形'前的相关语句。

爸爸埋怨他太慢吞吞而且懒得要命，说他不顶用，连叫他去看管鹅群都叫人不大放心。

不过她最烦恼伤心的还是他的粗野和顽皮。

他对牲口非常凶狠，对待人也很厉害。

（2）出示原著中相关资料验证猜测。

'哞，哞，大家都说你已经变了，变好了，我还不相信哩。喔！欢迎你回家来，尼尔斯·豪尔耶松，欢迎你回家来！我真太高兴啦，我有好久没有这样高兴过啦！'

'要想叫我交出大拇指，那是休想，'阿卡回答说，'我们愿意保卫他而献出自己的生命。'

3.借助目录，大胆想象。

（1）出示《骑鹅旅行记》目录，同桌互相想象尼尔斯的经历。

结合课前阅读提示要求，引导学生从’鹤之舞表演大会’和’大海中的白银’中选择一个标题，想象故事情节。

（2）制订阅读计划，阅读原著内容，验证自己的想象。

导语：从目录中我们可以看到各种飞禽走兽、各种人物和不同的地点，再加上作者有着无穷的想象力，阅读这本书一定会让你看到一个你意想不到的神奇世界。赶快去看看这本书吧！

D.3 Philosophy for Children in the classroom

This guidance describes the original idea, steps, and procedures of Philosophy for Children (P4C). The main emphasis of the intervention is to facilitate in asking open-ended questions and encourage pupils to develop the same. Teachers should avoid implanting their own views and values to give students the space to construct their own. With guidance from the teacher, the dialogue is focused not only on the chosen questions but also on the assumptions that lie behind the answers and the criteria used to make judgements.

1. Preparation

Teachers' role: sometimes being a participant, sometimes guiding students to deeper involvement, sometimes ensuring that the agreed rules of dialogue are adhered to. In P4C classroom, the teacher is not the only holder of knowledge and ideas or even perhaps the best ideas.

Development of the inquiry: Focusing attention on important points; modelling good questioning, for example by asking for clarification, reasons, and evidence; encouraging appropriate student behaviours, such as how to listen to each other and how to respond to each other; rewarding positive contributions with praise; not being content with just conversation; directing the discussion towards truth.

What is philosophy: * see Developing behavioural social targets

Moral structure: fucus attention on the speaker; do not 'put down' others; are not forced to speak; respect others' views; are truthful; are open minded. * see reminder cards & dialogic target

Skilful questioning: * see Thinking questions guidance sheet

Clarifying: what reasons do you have for saying that? Can you explain more about that?

Seeking evidence: how do you know that? What is your evidence?

Exploring alternative views: Is there another point of view? Can you put it another way?

Probing the superficial: why do you think that? What is the cause of that?

Scaffolding: if … then what do you think about…? You said… but what about…?

Testing implications: how can we test that in practice? Is that consistent with what you first said?

Evaluating: can anyone summarise the main points for us? Where has our thinking taken us?

2. Stages of developing a P4C session

The first two stages are preparation stages for students to help them dialogical skills and construct rules. The targets are as following: *see Planning template_Stage 1-3

Stage 1: teach and practice dialogical skills;

Stage 2: add think/ pair/ share and connections and tensions;

Stage 3: have students frame personal and philosophical questions (real-time introduction of P4C).

3. Process of a P4C session

The following are the main ten stages of a P4C session, extracted from Gorard, S., Siddiqui, N., & See, B. H. (2018). Philosophy for Children: Evaluation report and executive summary. EEF.

1). Getting set

Pupils and teacher sit in a circle so everyone can see and hear one another. The teacher negotiates guidelines on conduct, and the aims of enquiry with children or reminds them of previously negotiated guidelines. Sometimes the guidelines are revised in the light of ongoing dialogue. Warm-up activities are sometimes used to aid co-operation, speaking, listening, and thinking.

2). Presentation of stimulus

The teacher introduces the planned material she has chosen in order to provoke pupils' interest, puzzle them or prompt their sense of what is important. Stimuli intended for P4C should reference the sort of 'big ideas' – such as truth, fairness, rights, knowledge, and friendship – that are likely to excite philosophical dialogue. The range of stimuli used in P4C sessions includes short stories, poems, images, picture books, passages from novels, short video clips, newspaper articles, and material taken from other curriculum areas.

3). Thinking time

A minute of silent, individual thinking followed by pupils in pairs sharing interesting issues and themes or jotting down key words. The teacher often records some of the key words and ideas that emerge.

4). Question making

The teacher may suggest a question based on the outcome of thinking time. This is appropriate if the children are very young or new to P4C. More commonly, the class is split into small groups and asked to decide on a question they think is interesting, worth discussing, and that requires an answer based on reasoned judgement.

5). Question airing

Question airing Children present their group's question so all can see and hear it. When all the questions are collected and recorded, children are invited to clarify, link, appreciate or evaluate the questions prior to choosing one for discussion.

6). Question choosing

When the listing of questions is complete, the next phase is to select a one as a dialogue starter. The selection is made by pupils using one of a range of voting methods. Depending on the age of the children and the level of their experience of doing P4C, the teacher may help them with their selection.

7). First thoughts

The teacher asks pupils to share their thoughts on the question with their thinking partner(s). The discussion floor is then open for all to share their views. The teacher is moderator of the group so their role is to ensure that all pupils have a chance to speak. Confident and talkative pupils are likely to dominate the discussion but the teacher must encourage all pupils both to speak and to listen so that all may contribute to the 'community of enquiry'.

8). Building

Pupils participate in the discussion, building on other pupils' contributions, clarifying them, questioning them, and stating their own opinions. Whether agreeing or disagreeing the rule is to justify opinions with reasons. The teacher's role at this stage is to support pupils' reasoning, motivate them to question, and encourage them to take part in a dialogue with their peers. Teachers will often prompt pupils to make moves such as imagining alternatives and consequences, seeking evidence, quantifying with expressions like 'all', 'some' or 'most', offering examples and counter examples, and questioning assumptions. Teachers promote and model a stance of 'fallibilism' – willingness to amend or abandon one's opinion in the face of a good argument to do so. It is recommended in the P4C method to use some short gaps of silence or partner talk so that pupils can organise their thoughts and practise arguments with peers before sharing with the whole group. The teacher can also draw diagrams or make notes to keep track of significant arguments.

9). Last thoughts

The closing of the session involves last words from all pupils. Pupils might have the same opinion as they had at the beginning, or their view could have changed as a result of the dialogue. Pupils are invited to sum up their views concisely and without contradiction from others. They can sum up their views in a few words. This activity could either be a verbal statement or for a detailed reflection whereby a teacher could ask pupils to write a summary of their views.

10). Review

The teacher invites reflective and evaluative comments about the enquiry with reference to broad criteria such as the guidelines the group has adopted (see stage 1). The teacher asks: 'What went well?' 'What could we improve on?' 'What do we need to do next?' The teacher could point to issues of pupils' behaviour and turn-taking in the session and ask them to reflect on their progress. The review could include suggestions on what else needs to be focused in the next P4C sessions.

4. Classroom observation sheet

In observing a philosophy session, jot down strengths, weaknesses and points for discussion. Remember to note - *good comments by children; evidence or reasons given by children, linking of several ideas by an individual or several people, range of questions by children, has the group*

'moved forward' through dialogue? What threads were followed by the facilitator? Exactly what did you see and hear?

The calming exercise:
The stimulus:
Individual/ pair/ group work:
The enquiry through dialogue: (longest section and most detailed)
General comments:

5. Assessing classroom dialogue

Class: **Date:**

Duration:

Behavior	Need to improve	Good
Occurrence of pupils asking a question.		
Occurrence of pupil supporting their view / opinion with a reason.		
Occurrence of a pupil agreeing or disagreeing with the view of another pupil and giving a reason.		
Occurrence of pupil offering a reviewing or evaluating comment. (See ideas under Skilful Questioning section).		
Occurrence of pupil directly addressing another pupil.		
Occurrence of teacher asking a question requiring a one word or factual response.		
Occurrence of teacher asking an open- ended question. (This includes follow-up questions).		

6. Dialogic teaching DIT indicators

Indicators	Ratings		
	Monologic	Mid	Dialogic
Authorities	The teacher has exclusive control over discussion content and processes. She or he nominates students, asks questions, initiates topical shifts, and evaluates the answers.	There are occasional opportunities for students to freely engage in the discussion. These are rare and involve only a few students. Most of the time, the teacher controls turn-taking, prescribes topic choice, and reshapes the discussion to align with specific fixed content.	Students share major responsibilities for the process and substance of the discussion. They manage turns, ask questions, react to each other's ideas, suggest topical shifts, and propose procedural changes.
Questions	Teacher questions target recall of specific facts from the story. These are simple 'test' questions with one right or wrong answer known from the story or other sources.	The teacher asks questions of mixed quality, including complex, openended questions. Open questions are often designed to 'lead' students to a narrow range of interpretations of the text deemed	The discussion centers on truly open and cognitively challenging questions. The questions target higher order thinking, involving students in critical evaluation and analysis.

		acceptable by the teacher.	
Feedbacks	The teacher uses short, formulaic, or ambiguous feedback. The feedback does not invite students to further develop their answers (e.g., 'Umm. OK. Tracy?')	The quality of teacher follow-up is mixed. The teacher often listens to and works with student responses, but occasionally misses important opportunities to help the group to advance their inquiry further.	The teacher consistently works with student answers to inspire further exploration. He or she praises or questions the process of reasoning, not the conclusions (e.g., 'But how is cheating different from lying?').
Meta-level reflection: Connecting students' ideas	The teacher does not relate student answers to each other.	The teacher sometimes misses opportunities to connect students' ideas.	The teacher does not miss opportunities to make visible the connections among student ideas and prompt students to relate their ideas to what's been said by others. He or she often attributes student ideas and questions to specific speakers (e.g., 'Bill, do you want to respond to Kim's example?').
Explanation	Students do not explain what they think and why. Their responses are brief and factual, consisting of a word or a phrase.	Students occasionally share opinions and provide good justification for them. Longer student responses may represent simple retelling of events from the story.	Students take personal positions on the issue (e.g., 'I think,' 'I believe,' 'I feel') and support them with reasons and examples. They make elaborate, lengthy contributions, explaining their thinking to others.
Collaboration	Student responses are short, disjointed, and unrelated to each other. Students primarily 'report' about established, known facts.	Students occasionally build on each other's ideas. The collaboration often involves sharing of similar experiences, rather than a critical analysis of each other's ideas (e.g., 'This happened to me, too! I was visiting my aunt in Boston....').	Students engage in critical and collaborative 'co-construction of ideas.' Their responses are 'chained together,' as they react to each other's ideas.

Note: extracted from Reznitskaya, A. (2012). Dialogic teaching: Rethinking language use during literature discussions. *The reading teacher*, *65*(7), 446-456.

7. Assessing Philosophical Inquiry

0 = not observed, 1 = observed once, 2 = observed a few times, 3 = observed often

	Items	0	1	2	3
Social skills	Did students respond to the questions being asked?				
	Did students focus their attention on the speaker?				
	Did students avoid interrupting the speaker?				
	Did the group encourage each other to speak?				
	Did students respond to the previous speaker?				
	Were the contributions brief and to the point?				
	Did students recall the ideas of others and name them?				
	Did people try to build on the idea of others?				
	Did students listen to ideas different from their own?				
	Was there a willingness for people to change their minds?				
Thinking together	Did students ask open and inviting questions?				
	Did anyone ask for clarification?				
	Did students question assumptions?				
	Did anyone ask for examples or evidence?				
	Did students ask for reasons or criteria?				
	Did anyone give an example or counter-example?				
	Did anyone give a reason or justification?				
	Did people offer or explore alternative viewpoints?				
	Did students make comparisons or analogies?				
	Did students make distinctions?				

8. Example lesson plan

Marvin Gets Mad – Session 1

This has been written in 2 sessions – the first for children to enjoy the story, the second to explore 'ANGER'. Other possible themes are PATIENCE, LONELINESS, BEING SORRY, WANTING THINGS, CAN ANIMALS HAVE FEELINGS?

Focusing Exercise/ Activity	Get ready to listen. Gather the class together in a circle. Pass round a 'Hello, welcome to our philosophy group'. Practise sitting still and connecting with the listening until the class falls quiet. What did they notice? Remind the class of the rules for Philosophy, particularly listening to each other and 'no put-downs'.
Last week's thought	*Talk about the last Philosophy session and the 'Thought for the week.'* Did they remember it? Can they say anything about it? Encourage children to say what they found.

Story, Poem, Activity	*Gather the class round about you, if possible. Introduce the story as you normally would, taking as many different ideas as possible -* *e.g.* Look at cover. What is in the picture? How do you think he is feeling? What do you think the story could be about? (*Take different ideas)*' Have you ever been mad? *Read the story to the class, enjoying the pictures and talking about facial expressions and feelings as you read, without losing the thread of the story.* AFTER READING Recap events in stories with children Get them to act out facial expressions. How did Marvin feel? e.g. Finding the apple tree. (surprised, pleased) Not being able to reach the apple.(disappointed) Waiting for the apple to fall. (patient) Molly had eaten the apple. (disappointed) Marvin getting madder, stamping and shouting 'BAAAA'(angry) Ground swallowing him up.(surprised) In the hole.(lonely) Molly appears. (relieved, sorry) Back to everything perfect.(happy) Wanted the pear.
Preparation for next session	Tell the children that we'll do the story again tomorrow and talk about being angry.

Marvin Gets Mad – Session 2

Theme – ANGER

Focusing Exercise/ Activity	Repeat the opening from session 1.
Story, Poem, Activity	Ask the class what they remember about the story 'Marvin Gets Mad'. Recall the story together and all the emotions Marvin experienced. (Any one of them could be the basis for a philosophy session and if you feel confident, go with what interests the class. If you use ANGER, use some of the following questions to start the dialogue)
Think, Pair, Share	Construct questions that students can think about then talk to a partner about.
Dialogue/ Discussion Plan	Theme: ANGER <u>Story</u> Why do you think Marvin got mad? How do you think Molly felt when Marvin got mad at her? Why? How do you think the chickens, ducks and cows felt? Why? If you had been Marvin, would you have got mad? What would you have done? <u>Personal</u>

	Have you ever been angry like Marvin? (*Take examples)* What does it feel like? Can you stop being angry if you want to? How do you feel if someone is angry at you? Are you ever sorry after you have been angry? Why? Do you ever pretend to be angry? <u>Philosophical.</u> Is anger a good thing? What is anger?
What have we talked about?	What did we talk about today? How well did we do in our thinking and talking? (Use thumbs)
Thought for the week.	This week think about being angry. What makes you angry? What happens when we get angry? Can we stop being angry?

Here is the Chinese version used in the main experiment.

儿童哲学在课堂

Philosophy for Children in the classroom

这份指南描述了儿童哲学的原始想法、步骤和过程。儿童哲学强调的是帮助教师提出开放性问题，并鼓励学生也发展提问能力。教师应该避免把自己的观点和价值观强加在学生头上。在教师的指导下，对话不仅限于选定问题，同时还要关注答案背后的假设以及做出判断的标准。

准备

Preparation

教师角色：教师有时候是课堂的参与者，是带领学生深度讨论的引领者，有时还是规则的维护者。在儿童哲学课堂里，教师不是唯一的知识拥有者，他们的观点未必是最好的。

课堂构建：把注意力放在重要的观点上；建立良好的提问结构，例如请学生解释含义、阐述原因和证据；鼓励学生的正确行为，比如如何倾听其他人的观点以及如何回复；表扬学生对课堂的贡献；不要局限于对话本身，还要将对话引向对真理的探讨。

预备课程：哲学是什么

在课程正式开始之前，教师可以给从来没有接触过儿童哲学课的学生上一节关于哲学是什么的预备课程。《发展社会行为目标》是一个示例。

课堂规则：注意倾听发言者；不要'镇压'别人的观点；发言不是必须的；尊重他人的观点；开放度。

提问技巧

澄清：你所说的理由是什么？你可以给我们解释一下吗？

寻找证据：你是怎么知道的？你的证据是什么？

探寻其他观点：还有同学有其他观点吗？

探究表面：你为什么会这么想？原因何在？

帮扶：如果……，那你会怎么想？你说……，但是你怎么看待……？

联系实际：我们如何在现实生活中测试？这与你最初所说的一致吗？

评价：有哪位同学可以为我们总结主要观点吗？我们的讨论到哪里了？

儿童哲学课的三个阶段

Stages of developing a P4C session

前两个阶段教会学生对话技巧，并确立课堂规则，是对真正的儿童哲学课的预备课。

阶段 1:训练对话技巧

阶段 2:思考、配对、分享、联系观点

阶段 3:学生独立提出哲学问题

儿童哲学课堂的步骤

Process of a P4C session

热身活动 Getting set

学生和教师围坐一圈，以便每个人都能看到和听到对方。教师与学生们商议这项教学活动的行为准则和目的，或者回顾以前讨论过的准则，有时候可以根据正在进行的对话修改准则。可以使用热身活动来帮助合作、对话、倾听和思考。

展示启发材料 Presentation of stimulus

教师介绍事先准备的材料，激发学生的兴趣，令他们感到困惑，或促使他们意识到什么是重要的。为 P4C 准备的启发性材料需要涉及一些宏大的概念，例如真理、公平、权利、知识和友情，这些可能引发哲学对话的材料。启发材料的形式包括短篇故事、诗歌、图片、绘本、小说段落、视频、报纸文章以及来自其他课程领域的材料。

思考时间 Thinking time

留出一分钟的时间，要求学生们安静地思考。之后以小组为单位分享他们想到的有趣的问题、话题或关键词。教师在旁听时，经常记录一些出现的关键词和想法。

问题形成 Question making

对于低龄段学生或刚开始使用 P4C 的班级，教师可以在思考时间之后率先提出一个问题。对于高年级段学生或熟练使用该方法的班级，更常见的做法是，教师将全班分成几个小组，要求学生商议出一个他们认为有趣的、值得讨论的、需要根据推理判断来回答的问题。

问题提出 Question airing

每组的学生分别展示他们的问题，这样每组的想法都可以被听到，他们提出的所有问题都会被收集记录下来。而在选择本次讨论的核心问题之前，教师要邀请学生阐释、关联、欣赏或评价这些问题。

问题选择 Question choosing

列出所有问题后，学生们需要投票选择一个问题作为对话的开始。根据孩子们的年龄和班级使用 P4C 的经验，教师可以帮助他们进行选择。

最初想法 First thoughts

构建想法的最初阶段是一个开放讨论的平台，就是要让所有人分享对这个问题的想法。教师是班级讨论的主持人，他们的作用是确保所有的学生有机会发言。自信和健谈的学生可能会主导讨论，但教师应该鼓励所有的学生发言和倾听，这样所有的人都能为 '探究式学习共同体'作出贡献。

构建 Building

学生们参与讨论，以其他学生的发言为基础，明确地理解他们的观点，提出质疑，并陈述自己的观点。无论是否同意，都要用理由来支持自己的观点。教师在这个阶段的作用是支持学生对观点进行论证、质疑，鼓励他们参与到对话中。教师可以设想选择和结果，寻求证据，使用精确的词语量化表达，提供正面的或反向的例子，或者对假设提出质疑，从而帮助学生推动对话。教师要推崇一种'不确定主义'的立场，认为所有知识都有可能是不完全的、有限的、暂时的或可更正的，在面对有力的论据时，愿意修改或放弃自己的观点。教师可以在讨论中穿插短暂的沉默或同伴讨论，以便学生在分享观点之前有时间组织想法，并与同伴练习论证。教师也可以用画图或做笔记的方法，记录重要的论点。

最终想法 Last thoughts

在课程的最后，教师邀请所有学生的轮流发言。学生们可能持有和讨论开始时一致的观点，也有可能在对话后发生改变。学生们简明扼要地总结他们的观点，这个过程教师需要维持秩序，不加入其他学生评价或反驳的环节。这项活动既可以是口头陈述，也可以书写观点总结。

回顾 Review

教师请学生参照最开始制定的讨论准则，对调查进行反思和评价性评论。教师提问：'哪些地方做得好？"哪些地方可以改进？"我们接下来需要做什么？'教师可以指出学生在这一环节中的行为和轮流发言的存在的问题，并要求他们反思自己的进步，提供在下一个 P4C 课程中还需要关注什么的建议。

课堂观察表 Classroom observation sheet

在观察一节哲学课时，记下优点、缺点和讨论要点。记住要记下--孩子们的精彩评论；孩子们给出的证据或理由；个人或几个人将几个观点联系起来；孩子们提出的问题范围；小组是否通过对话‘向前推进’？教师遵循了哪些主线？你到底看到和听到了什么？

热身环节
启发材料
个人/同伴/小组合作
对话内容：
评价

评价课堂对话

班级： **日期：**

持续时间：

	行为	经常出现	偶尔出现	没有出现

学生	学生提问			
	学生阐述观点并叙述理由			
	学生表示赞同或否定他人的观点并阐述理由			
	学生回顾或评价其他人的观点			
	学生直接和其他学生对话			
教师	教师提出了一个唯一答案的问题			
	教师提出了一个开放式问题			

对话式教学指标

指标	评价		
	单向教学	半对话式教学	对话式教学
教师权威	老师对讨论的内容和过程拥有独占控制权。他/她指名学生、提问、引导话题转移，并评估答案。	偶尔会给学生一些自由参与讨论的机会，但这种情况很少，只涉及少数学生。大部分时间，老师控制着轮流发言，规定话题选择，并调整讨论内容以与特定固定内容相一致。	学生在讨论的过程和内容方面承担重要责任。他们管理发言轮次、提出问题、对彼此的想法做出回应、推动话题，并提出程序性变更。
提问	教师对事实进行提问，就像考试一样有唯一正确或错误的答案。	教师提出的问题中有一些是复杂的、开放式问题。但是开放式问题往往存在教师预设的正确答案。	讨论的核心式开放式的、有挑战的问题。这些问题目的在于引导高阶思维，需要学生批判性评价、分析。
反馈	教师使用简短的、公式化的、模糊的反馈。这些反馈并不邀请学生深入考虑他们的回答。	教师会给予学生后续反馈，并倾听学生的回答，但有时候会忽略重要的、可以推进思考的机会。	教师对学生的回答进行持续性关注，并且鼓励学生进一步的探索。反馈会更关注过程和原因，而不是结论。
关联学生的想法	教师不会把学生的观点关联起来。	教师有时候会错过关联学生观点的机会。	教师不会错过任何可以把学生的想法关联起来的机会，他会提示学生其他学生的看法并引导他们交流。
解释	学生不会解释他们在思考什么以及为什	学生偶尔会分享他们的观点并解释。但是	学生会从个人角度出发（比如'我认为……',''

	么。他们的回答都是简略的、事实性的。	又一些学生只是复述文章内容。	我相信……’）。他们会详尽地向别人解释观点
合作	学生的回答简短、互不关联。学生主要‘报告‘既定的已知事实。	学生偶尔会把自己的观点建立在别人的观点之上。合作通常包含了分享类似经历而不是批判性分析其他人的观点。（例如，’这件事也发生在了我的身上！我去拜访我姑姑的时候……’）	学生积极参加批判性的、合作性的’观点共建’。因为会回应其他人的观点，所以他们的回答串连在一起。

评价哲学性问题

0=没有出现，1=出现一次，2=出现少许几次，3=总是出现

	指标	0	1	2	3
社交技能	学生们回答了问题吗?				
	学生们专注倾听发言者了吗?				
	学生们避免打断发言者的话了吗?				
	小组讨论鼓励每个人都发言了吗?				
	学生有没有回应之前发言者的话?				
	学生的发言是简洁的吗?				
	学生们是否回想联系了其他人的看法?				
	学生们是否尝试在别人的观点上构建自己的观点?				
	学生们是否倾听了和自己不同的见解?				
	学生们是否愿意改变他们的想法?				
共同思考	教师有没有询问开放性问题?				
	有没有学生要求解释清楚?				
	学生们有没有质疑猜想?				
	有没有人询问例子或证据?				
	学生们是否询问原因或标准?				
	有没有人给出示例或反例?				

	有没有人阐述理由或证明了观点？				
	有没有人提出了其他观点？				
	有没有学生进行对比或类比？				
	学生有没有进行区分？				

示例教案

Marvin Gets Mad 马文气疯了– 第一课时

这次儿童哲学课分为两个部分，第一课是为了学生能够享受阅读故事的过程，第二课是为了探究’愤怒’。其他可能的话题有：耐心、孤独、愧疚、渴望、动物会有感觉吗等。

热身活动	准备好倾听 全班聚拢坐成一个圆圈，互相说’你好，欢迎来到我们的哲学小组。’ 提示儿童哲学课的规则，尤其是倾听且不要阻拦其他人说话。
上周思考	*讨论上一次儿童哲学课以及’每周思考’。* 他们还记得吗？他们可以对上次的内容说些什么吗？鼓励学生说一说他们发现了什么。
故事	让全班聚集在你周围，介绍故事内容，尽可能囊括更多的想法。 例如，看封面的插图，图片里有什么？你觉得他有什么感受？你觉得这个故事说的是什么（囊括不同的想法）？你有没有生气疯狂过？ 对着全班读故事，欣赏图片，在不跳出故事情节的基础上，聊一聊最表层的想法和感受。 阅读后，给孩子们复述故事情节，让他们表演出主人公的面部表情。马文的感受如何？例如， 找到苹果树。(意外的、愉悦的) 够不到苹果。 (失落的) 等着苹果掉下来。 (耐心的) 莫利吃掉了苹果。(失落的) 马文更疯狂了，跺脚大喊：’啊——’(生气的) 大地把他吞下去。(惊讶的) 在洞里。(孤单的) 莫利出现了。 (放松的、抱歉的) 一切归于正常。 (高兴的)
下节课准备	告诉学生我们明天将继续讨论生气。

Marvin Gets Mad 马文气疯了– 第二课时

主题–愤怒

热身活动	重复第一课

故事	询问学生们还记得《马文气疯了》中哪些情节，一起回忆故事内容和马文的心情波动（都可以作为课程的主要内容）。
思考、组队、分享	提出一些问题，让学生和同伴讨论。
对话/讨论计划	主题：愤怒 <u>故事</u> 你觉得马文为什么气疯了？ 你觉得当马文对莫利发火时，莫利怎么想？为什么？ 你觉得鸡、鸭、牛感受如何？为什么？ 如果你是马文，你会怎么做？ <u>个人</u> 你有没有像马文一样生气过？是什么样子？请举例子说明。 你可以控制你的脾气么？ 当别人对你发火的时候，你感受如何？ 在你发火之后你会感觉抱歉吗？为什么？ 你有没有假装生气过？ <u>哲学</u> 愤怒是好事么？ 什么是愤怒？What is anger?
总结	我们今天谈论了什么？ 我们的讨论好不好？（竖起大拇指）
本周思考	本周我们讨论了愤怒，什么会让你愤怒？你愤怒的时候会发生什么？ 我们怎么控制生气？

D.4 Writing about value

Writing about your life

There are a lot of things that are important to people—things that make their lives better, more important, or special.

For example, some people find being honest important because other people can trust them. Some other people find their family important because they love and value them. Other people find being good at sport important because it makes them feel good to play well.

In the space below, please write about what **you** find important in **your** life. How important is it to you? Why is it important to you? What does it mean to you to have it in your life?

Focus on your thoughts and feelings, and don't worry about spelling, grammar, or how well written it is, or how much you can write.

Writing about your Values

There are a lot of things that are important to people—things that make their lives better, more important, or special.

In the space below, please select two values from the list that are important to you and write about what they will do to show that these are important or how much you enjoy practicing them.

List of Values: humility, compassion, honesty, friendship, determination, courage, trustworthiness, happiness, self-reliance, curiosity, knowledge, fame, and achievement.

Focus on your thoughts and feelings, and don't worry about spelling, grammar, or how well written it is, or how much you can write.

Here is the Chinese version.

First Writing Exercise

第一次写作练习

姓名：

日期：

世界上有很多有意义的事情，这些事情会让你体验到自己的重要性或是与众不同，从而丰富你的生活和阅历。

例如，有些人认为诚信很重要，因为他人会相信他们。有的人认为家庭很重要，因为他们珍重他们的家人。还有些人重视体育运动，因为运动时他们可以突破自己、体验快乐。

请在下方的空白处写一件你生活中重要的事。展开写写这件事有多重要，为什么重要，对你的生活意味着什么？

这个练习需要你表达你的想法和感受，在写作的时候不用担心字数限制、错别字和病句。

Second Writing Task

第二次写作练习

姓名：

日期：

世界上有很多有意义的事情，这些事情会让你体验到自己的重要性或是与众不同，从而丰富你的生活和阅历。

请在下列关键词中选择两个你认为重要的词，在空白处写一写你在生活中是怎么做到的。

关键词：幽默、同情心、诚实、友谊、决心、勇气、信任、高兴、坚韧、好奇、知识、名誉、成就

这个练习需要你表达你的想法和感受，在写作的时候不用担心字数限制、错别字和病句。

D.5 Growth mindset

1. Materials

Student 1st intervention

Aim: get the idea of Growth mindset

Choose one:

Standard intervention material:

You can grow your intelligence: New research shows the brain can be developed like a muscle

Other useful materials:

Dweck's talk:

https://www.ted.com/talks/carol_dweck_the_power_of_believing_that_you_can_improve?language=en

Explanation:

https://www.youtube.com/watch?v=WtKJrB5rOKs

Student 2nd intervention (choose one)

Aim: how to use Growth mindset in their life

Workshop: how to use Growth mindset to prepare for their next exam

Writing activity: write a letter encouraging students in lower grades that they can improve their academic abilities

Workshop: is there anything you've always wanted to do, but were afraid to because you weren't good enough at it? Set up a plan to make it happen.

Writing activity: is the hero in you someone who succeeded solely based on extraordinary ability and with little effort? Go find out how he/she worked.

Active activities (Embedded Growth mindset)

Book list

After the fall, Dan Santat;

Anything is possible, Giulia Belloni;

The bad seed, Jory John;

Bears don't read, Emma Chichester Clark;

Beautiful Oops!, Barney Saltzberg;

Bob the artist, Marion Deuchars;

The book of mistake, Corinna Luyken;

The busy beaver, Nicolas Oldland;

Change Sings: A Children's Anthem, Amanda Gorman;

The Day You Begin, Jacqueline Woodson;

The Dot, Peter H. Reynolds;

The Dreamer, Il Sung Na;

Drum Dream Girl: How One Girl's Courage Changed Music, Margarita Engle and Rafael López;

Elephants Cannot Dance!, Mo Willems;

Emmanuel's Dream: The True Story of Emmanuel Ofosu Yeboah; Laurie Ann Thompson;

Flight School, Lita Judge;

Giraffes Can't Dance, Giles Andreas;

The Girl Who Never Made Mistakes, Mark Pett;

Malala's Magic Pencil, Malala Yousafzai;

The Most Magnificent Thing, Ashley Spires;

My Beautiful Voice, Joseph Coelho;

Ruby's Wish, Shirin Yim;

https://childrenslibrarylady.com/growth-mindset-books-for-children/

https://msjordanreads.com/developing-growth-mindset-reading/

Video

https://biglifejournal.com/blogs/blog/growth-mindset-videos-for-children

https://www.youtube.com/watch?v=rf8FX2sI3gU

Growth mindset for teachers -- lesson plan

Week 2: https://www.youtube.com/watch?v=Q_EyPX3CD-g

Week 3: https://www.youtube.com/watch?v=lBBge0bX2Ng

https://www.youtube.com/watch?v=wa-0yJXFeKk

Others

27 Growth mindset examples:

https://helpfulprofessor.com/growth-mindset-examples/

Mindset tools:

https://www.mindtools.com/asbakxx/dwecks-fixed-and-growth-mindsets

Resources collection:

https://www.edutopia.org/article/growth-mindset-resources/

Free online courses:

https://www.mindsetkit.org

2. You can grow your intelligence: New research shows the brain can be developed like a muscle

Reading for Activity Option 1 or 2

You Can Grow Your Intelligence

New Research Shows the Brain Can Be Developed Like a Muscle

Many people think of the brain as a mystery. They don't know much about intelligence and how it works. When they do think about what intelligence is, many people believe that a person is born either smart, average, or dumb—and stays that way for life.

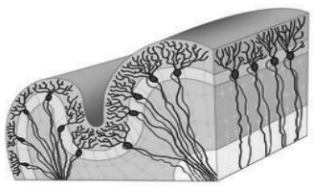
But new research shows that the brain is more like a muscle—it changes and gets stronger when you use it. And scientists have been able to show just how the brain grows and gets stronger when you learn.

Everyone knows that when you lift weights, your muscles get bigger and you get stronger. A person who can't lift 20 pounds when they start exercising can get strong enough to lift 100 pounds after working out for a long time. That's because the muscles become larger and stronger with exercise. And when you stop exercising, the muscles shrink and you get weaker. That's why people say "Use it or lose it!"



© 2010 Mindset Works

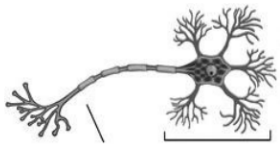
But most people don't know that when they practice and learn new things, parts of their brain change and get larger a lot like muscles do when they exercise.



© Fotosearch

A section of the cerebral cortex

Inside the cortex of the brain are billions of tiny nerve cells, called neurons. The nerve cells have branches connecting them to other cells in a complicated network. Communication between these brain cells is what allows us to think and solve problems.



© Fotosearch

A typical nerve cell

When you learn new things, these tiny connections in the brain actually multiply and get stronger. The more that you challenge your mind to learn, the more your brain cells grow. Then, things that you once found very hard or even impossible to do—like speaking a foreign language or doing algebra—seem to become easy. The result is a stronger, smarter brain.

2

Brainology® Curriculum Guide for Teachers

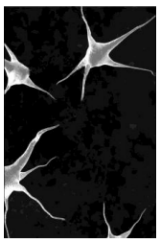
 Copyright © 2002-2014 Mindset Works, Inc. All rights reserved.

How Do We Know the Brain Can Grow Stronger?

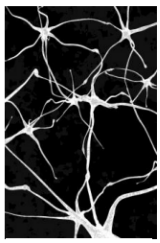
Scientists started thinking that the human brain could develop and change when they studied animals' brains. They found out that animals who lived in a challenging environment, with other animals and toys to play with, were different from animals who lived alone in bare cages.

While the animals who lived alone just ate and slept all the time, the ones who lived with different toys and other animals were always active. They spent a lot of time figuring out how to use the toys and how to get along with the other animals.

Effect of an Enriched Environment



Nerves in brain of animal living in bare cage



Brain of animal living with other animals and toys

© 2010 Mindset Works

These animals had more connections between the nerve cells in their brains. The connections were bigger and stronger, too. In fact, their whole brains were about 10% heavier than the brains of the animals who lived alone without toys.

The animals who were exercising their brains by playing with toys and each other were also "smarter"—they were better at solving problems and learning new things.

Even old animals got smarter and developed more connections in their brains when they got the chance to play with new toys and other animals. When scientists put very old animals in the cage with younger animals and new toys to explore, their brains also grew by about 10%!

3

Children's Brain Growth

Another thing that got scientists thinking about the brain growing and changing was babies. Everyone knows that babies are born without being able to talk or understand language. But somehow, almost all babies learn to speak their parents' language in the first few years of life. How do they do this?

The Key to Growing the Brain: Practice!

From the first day they are born, babies are hearing people around them talk—all day, every day, to the baby and to each other. They have to try to make sense of these strange sounds and figure out what they mean. In a way, babies are exercising their brains by listening hard.

Later, when they need to tell their parents what they want, they start practicing talking themselves. At first, they just make goo-goo sounds. Then, words start coming. And by the time they are three years old, most can say whole sentences almost perfectly.

Once children learn a language, they don't forget it. The child's brain has changed—it has actually gotten smarter.

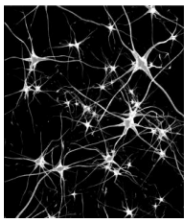
This can happen because learning causes permanent changes in the brain. The babies' brain cells get larger and grow new connections between them. These new, stronger connections make the child's brain stronger and smarter, just like a weightlifter's big muscles make them strong.

4

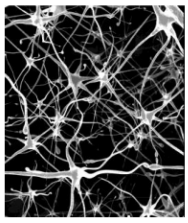
Brainology® Curriculum Guide for Teachers

 Copyright © 2002-2014 Mindset Works, Inc. All rights reserved.

Growth of neuron connections in a child from birth to 6 years old



At birth



At age 6

© 2010 Mindset Works

The Real Truth About "Smart" and "Dumb"

No one thinks babies are stupid because they can't talk. They just haven't learned how to yet. But some people will call a person dumb if they can't solve math problems, or spell a word right, or read fast—even though all these things are learned with practice.

At first, no one can read or solve equations. But with practice, they can learn to do it. And the more a person learns, the easier it gets to learn new things—because their brain "muscles" have gotten stronger!

The students everyone thinks as the "smartest" may not have been born any different from anyone else. But before they started school, they may have started to practice reading. They had already started to build up their "reading muscles." Then, in the classroom, everyone said, "That's the smartest student in the class."

They don't realize that any of the other students could learn to do as well if they exercised and practiced reading as much. Remember, all of those other students learned to speak at least one whole language already—something that grownups find very hard to do. They just need to build up their "reading muscles" too.

5

What Can You Do to Get Smarter?

Just like a weightlifter or a basketball player, to be a brain athlete, you have to exercise and practice. By practicing, you make your brain stronger. You also learn skills that let you use your brain in a smarter way—just like a basketball player learns new moves.

But many people miss out on the chance to grow a stronger brain because they think they can't do it, or that it's too hard. It does take work, just like becoming stronger physically or becoming a better ball player does. Sometimes it even hurts! But when you feel yourself get better and stronger, all the work is worth it!

6

www.mindsetworks.com Copyright © 2002-2014 Mindset Works, Inc. All rights reserved.

Here is the Chinese translated version.

你可以变得更聪明

研究新发现：大脑像肌肉一样可以锻炼成长

很多人觉得大脑很神秘，他们并不了解什么是智力，也不知道它是怎么运作的。很多人以为，一个人生下来就是聪明的、普通的，或者愚蠢的，而且一辈子都不会改变。但新的研究发现，大脑其实更像是一块肌肉——你用得越多，它就会变得越强壮。科学家已经证明，在你努力学习的时候，大脑真的会成长，也会变得更强。

众所周知，举重能让肌肉变紧实，身体变得更强壮。一个一开始连 20 磅都举不动的人，只要坚持锻炼，最后也可能举起 100 磅。这是因为肌肉在锻炼中变得越来越有力。相反，如果你不锻炼，肌肉就会萎缩，力量也会变弱。这就是人们常说的用进废退。

但是，大多数人不知道，当我们练习和学习新事物时，大脑的某些部分也会发生变化，就像肌肉锻炼时会变紧实一样。在大脑的皮层里，有数十亿个非常小的神经细胞，叫做“神经元”。这些神经元有很多分支，把它们和其他细胞连接起来，形成一个复杂的网络。大脑里的这些细胞通过信息传递，让我们能够思考和解决问题。

当你学习新知识时，大脑中这些连接会增加，而且变得更强。你越努力动脑学习，大脑细胞就长得越多、越结实。以前你觉得很难、甚至不可能做的事情，比如外语或者代数，现在会变得越来越容易。这样一来，你的大脑就变得更强、更聪明了。

科学家怎么知道大脑真的能变强？

一开始，科学家是在研究动物的大脑时才注意到这一点的。他们发现，生活环境复杂的动物，比如有玩具、有同伴一起玩的，和那些单独关在空笼子里的动物，大脑是不一样的。独自生活的动物，大部分时间用来吃饭和睡觉；而有玩具和伙伴的动物总是在动脑筋，比如想着怎么玩玩具、怎么和其他动物相处。这些动物的大脑里有更多神经细胞之间的连接，而且这些连接也更大、更强。事实上，它们的大脑比那些没有玩具、独自生活的动物的大脑重了大约 10%。这些通过玩玩具和和同伴互动来“锻炼大脑”的动物，也变得更“聪明”了，尤其是在解决问题和学习新东西方面表现得更好。

就连年纪很大的动物，在有机会和其他动物一起玩、接触新玩具的时候，它们也变得更聪明，大脑的神经连接也变多了。科学家把一些年老动物和年轻动物放在一起，给予它们新玩具，这些动物的大脑容量大约增长了 10%！

儿童的大脑也会成长

还有一件事让科学家开始思考大脑是否真的会成长和变化，那就是婴儿。大家都知道，宝宝出生时不会说话，也听不懂别人的语言。但几乎所有的宝宝在几岁之内都能学会说爸爸妈妈的语言。这是怎么做到的呢？

让大脑成长的关键：多练习！

从出生的第一天起，宝宝每天都会听到周围的人在说话。有人对他们说话，也有人互相交谈。他们需要努力听这些陌生的声音，并试着搞懂这些声音是什么意思。可以说，宝宝是在通过“认真听”来锻炼自己大脑的能力。等到他们想告诉爸爸妈妈自己想要什么的时候，就开始练习说话了。一开始只是发出“咕咕”声，慢慢地开始说出一个个词，最后就能说出整句话。到三岁左右，大多数孩子已经能很顺利地表达自己的意思了。

一旦孩子学会了语言，他们通常就不会忘记，因为学习会让大脑发生永久的变化。大脑细胞变大了，也长出了新的连接。这些新的、更强的连接，就像举重运动员的肌肉一样，变得更有力、更聪明。

关于“聪明”和“愚蠢”

没有人会因为婴儿不会说话而觉得他们蠢笨，他们只是还没有学会怎么说话罢了。可是，有些人看到别人不会解数学题、拼错单词或者读得很慢，就会说这个人很笨。但其实这些都是可以通过练习学会的。一开始，没人会读书或解方程式，但只要多练习，就能学会。而且，练习越多，学新东西就越容易，因为他们的大脑“肌肉”变得更强壮了！

大家认为最“聪明”的学生，可能一开始和其他人没什么不同。但是，在他们上学之前，他们也许已经开始练习阅读了。他们已经锻炼了自己的“阅读肌肉”。于是，在课堂上他们表现得很好，大家就说：“那是班上最聪明的学生。”他们没有意识到，其他学生只要像他们一样多练习阅读，也能学得一样好。他们也只是需要锻炼自己的“阅读肌肉”罢了。

怎样才能变得更聪明？

就像举重运动员或篮球运动员一样，要成为大脑的运动员，你必须锻炼和练习。通过练习，你让大脑变得更强大。你还会学到一些技巧，让你能更聪明地使用大脑——就像篮球运动员学会了新的技巧。

很多人错失了锻炼大脑的机会，因为他们觉得自己做不到，或者觉得太难了。确实，这需要付出努力，有时候甚至会感到痛苦，这和成为更强壮、更好的球员一样。但是，当你发觉自己变得更好、更强时，所有的努力都是值得的！

Appendix E. Teacher Use-of-Research-Evidence Survey

Teacher Use-of-Research-Evidence Survey (Pre-test)

This survey asks for your views on the value of research evidence for the classroom and basic knowledge about using research evidence.
The survey is anonymous. The data in the survey will only be used for this project. However, we want to be able to link your responses here to your later views, so we ask for your grade and class later just to link the two sets of responses.
This survey has two sections. The first section is about your perceptions, understanding, and use of research evidence in the classroom. There is no right or wrong answer for each item. The second section asks for some background information (for example, so that we can compare the views of more or less experienced teachers).

Part 1: About your sources of information

Research evidence in science is always a hot topic. In education, there is also research evidence. We would like to know how you define research evidence in education and how you might use it in the classroom.

For each of the following statements below indicate how much you agree with it, using the scale from 0 (strongly disagree) to 10 (strongly agree).

1. How much do you agree that the following sources are trustworthy when making decisions to improve your teaching practice?

	Strongly disagree ……. Strongly agree										
	0	1	2	3	4	5	6	7	8	9	10
Promising projects and materials from your school, or the local system											
Colleagues e.g. informal discussion, classroom observation feedback											
High profile education figures e.g. authors of edu-books, education consultants											
Academic researchers											
China Education Guidance from Department of Education											
Educational news reports/ articles e.g. Teachers in primary schools in China											
None of these sources											
Other (please specify)											

2. How much do you agree with the following descriptions of educational research evidence?

	Strongly disagree ……. Strongly agree										
	0	1	2	3	4	5	6	7	8	9	10
Books and articles by educators and psychologists disseminating their views											
Knowledge from observations, information analysis, and discussions in my school											

Findings from robust research conducted by experts in education											
Other (please specify)											

3. I am confident that:

	Strongly disagree ……. Strongly agree										
	0	1	2	3	4	5	6	7	8	9	10
I can understand the findings of research											
I can decide whether the findings of research are trustworthy											
I can select appropriate research evidence to support my teaching practice											
I can adapt research for my teaching practice											
I can conduct my own robust research											
Research findings are useful to improve my teaching											

4. To get access to trustworthy research, I frequently use:

	Strongly disagree ……. Strongly agree										
	0	1	2	3	4	5	6	7	8	9	10
Searching engines e.g. Baidu											
Academic Databases e.g. China National Knowledge Infrastructure											
Reports, books, and articles published by the government, schools or teachers											
Research articles or summaries published by researchers											
Professional development conferences											
Social media e.g. TED, WeChat blogs, Zhihu											

5. If education research shows that an intervention leads to a positive impact on student attainment, what does this mean?

	Strongly disagree ……. Strongly agree										
	0	1	2	3	4	5	6	7	8	9	10
The study was of high quality											
The intervention was effective in this research environment											
All students in the intervention group performed better than those in the control											

6. Is this a robust research approach to evaluate the impact of peer tutoring on reading?

	Strongly disagree ……. Strongly agree										
	0	1	2	3	4	5	6	7	8	9	10
Selecting one class as a ‘peer tutoring’ class and compare their reading scores with other classes.											
Interviewing 40 students about their experience of peer tutoring											

7. I could better access, understand, and use research evidence in my practice if:

	Strongly disagree ……. Strongly agree										
	0	1	2	3	4	5	6	7	8	9	10
School provides a more evidence-based syllabus											
I have less managerial or teaching workload											
I get free access to the latest education research evidence											
Researchers give us specific training and instructions											
I observe other teachers' class using evidence-based practice											

8. Imagine that the school provided you with teaching materials which are said to be evidence-based. However, your students do not make progress. Would you?

	Strongly disagree ……. Strongly agree										
	0	1	2	3	4	5	6	7	8	9	10
Stop using these materials immediately											
Compare your use of materials with colleagues and ask for support											
Look again at the research evidence, and make a judgement on whether to continue											
Trust the school and carry on using the materials as instructed											

Part 2 About You

Grade:
Class:
(This will only be used to link your responses here to any later responses.)

What was your major subject in your highest qualification? Tick all that apply:

Teacher Education - Chinese	
Teacher Education - Other	
Curriculum and Pedagogy - Chinese	
Curriculum and Pedagogy - Other	
General education	
Chinese language and literature	
Applied Linguistics	
Other (please specify):	

Thanks for your co-operation.

Teacher Use-of-Research-Evidence Survey (Post-test)

This survey is asking for your views on the value of research evidence for the classroom. As before, the survey is anonymous. The data in the survey will only be used for this project.

Grade:
Class:

About your sources of information

Research evidence in science is always a hot topic. In education, there is also research evidence. We would like to know how you define research evidence in education and how you might use it in the classroom.
For each of the following statements below indicate how much you agree with it, using the scale from 0 (strongly disagree) to 10 (strongly agree).

1. How much do you agree that the following sources are trustworthy when making decisions to improve your teaching practice?

	Strongly disagree ……. Strongly agree										
	0	1	2	3	4	5	6	7	8	9	10
Promising projects and materials from your school, or the local system											
Colleagues e.g. informal discussion, classroom observation feedback											
High profile education figures e.g. authors of edu-books, education consultants											
Academic researchers											
China Education Guidance from Department of Education											
Educational news reports/ articles e.g. Teachers in primary schools in China											
None of these sources											
Other (please specify)											

2. How much do you agree with the following descriptions of educational research evidence?

	Strongly disagree ……. Strongly agree										
	0	1	2	3	4	5	6	7	8	9	10
Books and articles by educators and psychologists disseminating their views											
Knowledge from observations, information analysis, and discussions in my school											
Findings from robust research conducted by experts in education											
Other (please specify											

3. I am confident that:

	Strongly disagree ……. Strongly agree										
	0	1	2	3	4	5	6	7	8	9	10

I can understand the findings of research											
I can decide whether the findings of research are trustworthy											
I can select appropriate research evidence to support my teaching practice											
I can adapt research for my teaching practice											
I can conduct my own robust research											
Research findings are useful to improve my teaching											

4. To get access to trustworthy research, I frequently use:

	Strongly disagree ……. Strongly agree										
	0	1	2	3	4	5	6	7	8	9	10
Searching engines e.g. Baidu											
Academic Databases e.g. China National Knowledge Infrastructure											
Reports, books, and articles published by the government, schools or teachers											
Research articles or summaries published by researchers											
Professional development conferences											
Social media e.g. TED, WeChat blogs, Zhihu											

5. If education research shows that an intervention leads to a positive impact on student attainment, what does this mean?

	Strongly disagree ……. Strongly agree										
	0	1	2	3	4	5	6	7	8	9	10
The study was of low quality											
The intervention was ineffective in this research environment											
All students in the intervention group performed worse than those in the control											

6. Is this a robust research approach to evaluate the impact of peer tutoring on reading?

	Strongly disagree ……. Strongly agree										
	0	1	2	3	4	5	6	7	8	9	10
1,023 students randomised to read extra books for 25 minutes per day, or not											
Collecting reading scores and extracurricular reading time of 300 students for correlation analysis											

7. I could better access, understand, and use research evidence in my practice if:

	Strongly disagree ……. Strongly agree										
	0	1	2	3	4	5	6	7	8	9	10
School provides a more evidence-based syllabus											
I have less managerial or teaching workload											

I get free access to the latest education research evidence											
Researchers give us specific training and instructions											
I observe other teachers' class using evidence-based practice											

8. Imagine that you always encourage students to list their mistakes in a booklet and reflect on them. However, in a professional development conference, the expert from a prestigious university said that this was ineffective and so a waste of time. Would you?

	Strongly disagree ……. Strongly agree										
	0	1	2	3	4	5	6	7	8	9	10
Trust your experience and carry on											
Conduct your own research to decide whether to continue											
Look again at the research evidence, and make a judgement on whether to continue											
Trust the expert and change your teaching approach											

Thanks for your co-operation.

Here is the Chinese version which was used in the experiment.

关于教师使用研究证据情况的问卷

尊敬的老师：

这份问卷旨在了解教师如何看待研究证据在课堂教学中的使用以及使用研究证据的基本素养。这份问卷是匿名的，收集到的所有数据只用于本研究项目。我们需要您在后续问题中提供您的任课年级和班级，从而将与您相关的所有数据串联起来。本问卷有两个部分。第一部分包含您对于在课堂教学中使用研究证据的观点、理解和运用，这些问题没有标准答案。第二部分是关于您的背景信息。

第一部分：关于您的信息来源

科学界的研究发现总会引起热议，而教育学领域同样有丰富的研究证据。我们想要了解您如何定义教育学研究证据以及您怎样在课堂中使用。请用 0 到 10 表明您对以下这些叙述的同意度，0 表示强烈不同意，10 表示强烈同意。

1. 当您决定要改进教学的时候，您非常相信……

	非常不赞同…… 非常赞同										
	0	1	2	3	4	5	6	7	8	9	10
学校或者教育局推荐开展的项目或阅读的材料											
同事（例如课后闲谈，公开课评课等）											
有名的教育人士（例如教育类书籍作者，教育顾问等）											
学术研究者											
教育部发布的指导性文件											
教育研究报告或杂志（例如《OECD 中国教育质量研究报告》，《中国教师报》等）											
以上资源都没有											
其他（请明确具体的资源）											

2. 您有多赞同以下对教育研究证据的描述？

	非常不赞同…… 非常赞同										
	0	1	2	3	4	5	6	7	8	9	10
教育学家和心理学家用来传播他们观点的书籍和文章											
从学校观察、信息分析和讨论中获得的知识											
从教育专家开展的严谨的研究中得出的结果											
其他（请明确）											

3. 我有信心……

	非常不赞同…… 非常赞同										
	0	1	2	3	4	5	6	7	8	9	10
理解研究发现。											
判断研究发现是否可信。											

选择合适的研究证据支持教学。											
改编研究发现，使它与教学内容相适配。											
自行开展严谨的研究。											
使用研究发现可以提高教学效果。											

4. 为了获取可信的研究证据，我经常使用：

	非常不赞同…… 非常赞同										
	0	1	2	3	4	5	6	7	8	9	10
搜索引擎（如百度）											
学术数据库（如中国知网）											
政府、学校或老师发表的报告、书籍和文章											
研究者发表的文章或总结											
教师职业发展会议											
社交媒体（如 TED，微信公众号，知乎）											

5. 如果一个教育学研究发现某个干预可以给予学生成绩正向的影响，这意味着什么？

	非常不赞同…… 非常赞同										
	0	1	2	3	4	5	6	7	8	9	10
这项研究的质量很好。											
这个干预在研究环境中有效。											
所有接受干预的学生都表现得比没接收干预的更好。											

6. 请判断以下研究方法是否严谨地评价了同伴互助学习在阅读上的效果。

	Strongly disagree ……. Strongly agree										
	0	1	2	3	4	5	6	7	8	9	10
选择一个班级为'同伴互助'实验班，一段时间后，将他们的阅读成绩和其他班级学生的成绩对比。											
采访 40 名学生了解他们参加同伴互助学习的经历。											

7. 我能够更好地获取、理解研究证据，并运用到教学中，如果:

	非常不赞同…… 非常赞同										
	0	1	2	3	4	5	6	7	8	9	10
学校提供基于研究证据的课程。											
我有更少的行政或教学任务。											
我可以免费获取最新教育学研究证据。											
研究者给予我们明确的培训和指导。											
观察其他老师在课堂中基于证据的实践。											

8. 设想这一情景，学校提供了一套声称是研究证据导向的教学材料。但是，在使用后您的学生并没有进步，您会怎么做？

	非常不赞同…… 非常赞同										
	0	1	2	3	4	5	6	7	8	9	10

立即停止使用该材料。											
对比自己和同事使用材料的差异并寻求帮助。											
回顾材料的研究证据以判断是否继续使用。											
相信学校的判断并继续按要求使用。											

第二部分 关于您

任课年级：

任课班级:

您最高学历对应的专业是？请选择所有适用选项：

师范类-语文	
师范类-其他	
课程与教学-语文	
课程与教学-其他	
教育学	
汉语言文学	
应用语言学	
其他（请明确）	

感谢您的配合。

关于教师使用研究证据情况的问卷

尊敬的老师：

这份问卷旨在了解教师如何看待研究证据在课堂教学中的使用以及使用研究证据的基本素养。和之前一样，这份问卷是匿名的，收集到的数据只用于本研究项目。

任课年级：

任课班级:

关于您的信息来源

科学界的研究发现总会引起热议，而教育学领域同样有很多研究证据。我们想要了解 您如何定义教育学研究证据以及您怎样在课堂中使用。请用 0 到 10 表明您对以下这些叙述的同意度，0 表示强烈不同意，10 表示强烈同意。

1. 当您决定要改进教学的时候，您非常相信……

	非常不赞同…… 非常赞同										
	0	1	2	3	4	5	6	7	8	9	10
学校或者教育局推荐开展的项目或阅读的材料											
同事（例如课后闲谈，公开课评课等）											
有名的教育人士（例如教育类书籍作者，教育顾问等）											
学术研究者											
教育部发布的指导性文件											
教育研究报告或杂志（例如《OECD 中国教育质量研究报告》，《中国教师报》等）											
以上资源都没有											
其他（请明确具体的资源）											

2. 您有多赞同以下对教育研究证据的描述？

	非常不赞同…… 非常赞同										
	0	1	2	3	4	5	6	7	8	9	10
教育学家和心理学家用来传播他们观点的书籍和文章											
从学校观察、信息分析和讨论中获得的知识											
从教育专家开展的严谨的研究中得出的结果											
其他（请明确）											

3. 我有信心……

	非常不赞同…… 非常赞同										
	0	1	2	3	4	5	6	7	8	9	10
理解研究发现。											
判断研究发现是否可信。											
选择合适的研究证据支持教学。											
改编研究发现，使它与教学内容相适配。											
自行开展严谨的研究。											

使用研究发现可以提高教学效果。											

4. 为了获取可信的研究证据，我经常使用：

	非常不赞同…… 非常赞同										
	0	1	2	3	4	5	6	7	8	9	10
搜索引擎（如百度）											
学术数据库（如中国知网）											
政府、学校或老师发表的报告、书籍和文章											
研究者发表的文章或总结											
教师职业发展会议											
社交媒体（如 TED，微信公众号，知乎）											

5. 如果一个教育学研究发现某个干预可以给予学生成绩负面的影响，这意味着什么？

	非常不赞同…… 非常赞同										
	0	1	2	3	4	5	6	7	8	9	10
这项研究的质量很差。											
这个干预在实验环境中没有作用。											
所有接受干预的学生都表现得比没接收干预的差。											

6. 请判断以下研究方法是否严谨地评价了课外阅读对学生阅读水平的影响。

	非常不赞同…… 非常赞同										
	0	1	2	3	4	5	6	7	8	9	10
从 1023 名学生中随机抽取一部分学生进行每天 25 分钟的课外阅读，一段时间后测试。											
收集 300 名学生的阅读成绩和课外阅读时间进行相关性分析。											

7. 我能够更好地获取、理解研究证据，并运用到教学中，如果:

	非常不赞同…… 非常赞同										
	0	1	2	3	4	5	6	7	8	9	10
学校提供基于研究证据的课程。											
我有更少的行政或教学任务。											
我可以免费获取最新教育学研究证据。											
研究者给予我们明确的培训和指导。											
观察其他老师在课堂中基于证据的实践。											

8. 设想这一情景：您一直鼓励学生收集并反思错题，但在一场教师培训会议上，一位知名学者提出收集错题没有效果，反而会浪费学生的时间，您会怎么做？

	非常不赞同…… 非常赞同										
	0	1	2	3	4	5	6	7	8	9	10
相信自己的经验，要求学生继续使用错题本。											
设计自己的小实验探究收集错题的有效性，以决定是否使用错题本。											

回顾学者的研究证据，再判断是否继续使用错题本。											
相信专家，不再要求学生收集并反思错题。											

感谢您的配合。

Appendix F. Student Enjoyment of Reading Survey

Student Survey about Enjoyment of Chinese Reading

Dear students,

This survey is asking for your enjoyment and habit of reading in Chinese. The survey is anonymous. We collect your ID just to link it with your later views. In this questionnaire, there are no right or wrong answers. Your answers should be the ones that are right for yourself.

School: (select)
Grade: (select)
Class: (select)
Student ID: (type-in a number)

1. How much do you agree with the statement that I enjoy reading…

	Strongly disagree ……. Strongly agree										
	0	1	2	3	4	5	6	7	8	9	10
Chinese textbooks											
Magazines, Newspapers											
Comic books											
Fiction											
Science											
Social Science											

2. How much do you agree with the statement that I often…

	Strongly disagree ……. Strongly agree										
	0	1	2	3	4	5	6	7	8	9	10
borrow books from school library.											
read online.											
read paper books.											
read for information.											
read for enjoyment.											

3. How much you agree with each of these statements about reading?

	Strongly disagree ……. Strongly agree										
	0	1	2	3	4	5	6	7	8	9	10
I read only if I have to.											
Reading is one of my favourite hobbies.											
I like talking about books with other people.											
I would be happy if someone gave me a book as a present.											
For me, reading is a waste of time.											
I read only to get information that I need.											

4. How much you agree with each of these statements about how well you read in Chinese?

	Strongly disagree ……. Strongly agree										
	0	1	2	3	4	5	6	7	8	9	10
I usually do well in reading.											
Reading is harder for me than for many of my classmates.											
Chinese is easier for me than any other subject.											
I have to read a text several times before completely understanding it.											
I find it easy to answer questions about a text.											

5. How much time do you usually spend on Chinese homework?

_____minutes per day

Here is the Chinese Version.

学生语文阅读兴趣习惯问卷

亲爱的同学：
我们邀请你填写这份关于语文阅读兴趣和习惯的问卷。这份问卷是匿名的，我们收集你的学号只是为了关联和你相关的数据。每道题目的回答没有正确或错误之分，只需要能代表你的观点即可。

学校：
年级：
班级：
学号：

1. 我喜欢读……

	完全不赞同----------完全赞同										
	0	1	2	3	4	5	6	7	8	9	10
语文书											
杂志，报纸											
漫画书											
文学作品（小说、散文等）											
科普类书籍											
人文类书籍（传记、历史等）											

2. 我总是……

	完全不赞同----------完全赞同										
	0	1	2	3	4	5	6	7	8	9	10
从学校图书馆借阅书籍。											
网上阅读。											
阅读纸质书籍。											
为了获取信息而阅读。											
为了享受而阅读。											

3. 你有多赞同以下关于中文阅读的陈述？

	完全不赞同----------完全赞同										
	0	1	2	3	4	5	6	7	8	9	10
我只在必须阅读的时间阅读。											
阅读是我的兴趣之一。											
我喜欢和别人讨论书里的内容。											
如果有人送给我一本书作为礼物，我会很开心。											
对我来说，阅读是浪费时间。											
我只在需要获取信息时阅读。											

4. 你有多赞同以下关于你语文阅读能力的陈述？

	完全不赞同----------完全赞同										
	0	1	2	3	4	5	6	7	8	9	10
我很擅长阅读。											
相比于我的同学，我觉得语文阅读对我的挑战更大一些。											
对我来说，语文比其他学科都要简单。											
我必须反复阅读一段话才能理解其含义。											
我认为回答阅读理解题很困难。											

5. 每天你花多长时间完成语文作业？

_____分钟/天

Appendix G. Classroom observation form

Date	
Class	
Topic	
Applied strategy& technique	
Record & summary	
Evaluation	Did the teacher use the strategy fluently? Were activities/ techniques applied correctly? Did the teacher organise the time properly? Could the teacher manage the classroom? How were students' reaction? Were they engaged in the activities?
Comments	
Teacher feedback	

Appendix H. Interview questions

1. Headteacher Interview

1) How do you view the relationship between educational research outcomes and evidence-informed decision-making in schools?

2) The project could not have been carried out without the support of the school. How did you coordinate and organise the teachers' participation in the project? What difficulties did you encounter during the process?

3) Have you observed any changes among the teachers in the treatment group?

4) After this project, do you have any new ideas regarding the school's future project applications?

The following statements show the Chinese version.

1）您如何看待教育学研究成果和学校科学决策的关系？

2）课题的开展离不开学校的支持，您是如何统筹安排老师们开展这一课题的？过程中有遇到哪些困难？

3）您是否观察到实验组老师的变化？

4）这次课题后，您对学校未来申报课题有什么新的想法？

2. School leads Interview

1) How do you view the relationship between educational research outcomes and teaching? 2) What has this project brought to your teaching practice?

3) What efforts have you made to organize and encourage teachers to use these strategies?

4) What difficulties did you encounter during the implementation process, and how did you address them?

5) During this semester's implementation, have you observed any changes in the students?

The following statements show the Chinese version.

1）您如何看待在教育学研究成果和教学的关系？这项课题为您的教学带来了什么？

2）您做出了哪些努力组织老师们使用这些策略？

3）在实践过程中，您遇到过哪些困难，您是如何解决的？

4）在本学期的实践中，您是否观察到学生的变化？

Appendix I. Ethical approval

1. Ethical approval for the overall study

Wednesday, March 5, 2025 at 17:27:38 Greenwich Mean Time

Subject: Ethical Approval: EDU-2023-04-03T14_41_08-lfpx85
Date: Monday 17 April 2023 at 14:19:19 British Summer Time
From: Ethics <no-reply@sharepointonline.com>
To: CHEN, WENQING <wenqing.chen@durham.ac.uk>
CC: ED-ETHICS <ed.ethics@durham.ac.uk>, GORARD, STEPHEN A.C. <s.a.c.gorard@durham.ac.uk>

Please do not reply to this email.

Dear Wenqing,

The following project has received ethical approval:

Project Title: *Teachers' Use of Research Evidence for the Reading Attainment;*
Start Date: *01 May 2023*;
End Date: *30 September 2025;*
Reference: *EDU-2023-04-03T14_41_08-lfpx85*
Date of ethical approval: *17 April 2023*.

Please be aware that if you make any significant changes to the design, duration or delivery of your project, you should contact your department ethics representative for advice, as further consideration and approval may then be required.

If you have any queries regarding this approval or need anything further, please contact ed.ethics@durham.ac.uk

---

If you have any queries relating to the ethical review process, please contact your supervisor (where applicable) or departmental ethics representative in the first instance. If you have any queries relating to the online system, please contact research.policy@durham.ac.uk.

1 of 1

2. Ethical approval for the main experiment

Wednesday, March 5, 2025 at 17:27:52 Greenwich Mean Time

Subject: Ethical Approval: EDU-2023-11-09T10_26_22-lfpx85
Date: Thursday 16 November 2023 at 07:05:31 Greenwich Mean Time
From: Ethics <no-reply@sharepointonline.com>
To: CHEN, WENQING <wenqing.chen@durham.ac.uk>
CC: ED-ETHICS <ed.ethics@durham.ac.uk>, GORARD, STEPHEN A.C. <s.a.c.gorard@durham.ac.uk>

Please do not reply to this email.

Dear Wenqing,

The following project has received ethical approval:

Project Title: *Evaluation of Teachers' Use of Research Evidence for Reading Attainment;*
Start Date: *01 February 2024*;
End Date: *01 September 2024;*
Reference: *EDU-2023-11-09T10_26_22-lfpx85*
Date of ethical approval: *16 November 2023*.

Please be aware that if you make any significant changes to the design, duration or delivery of your project, you should contact your department ethics representative for advice, as further consideration and approval may then be required.

If you have any queries regarding this approval or need anything further, please contact ed.ethics@durham.ac.uk

---

If you have any queries relating to the ethical review process, please contact your supervisor (where applicable) or departmental ethics representative in the first instance. If you have any queries relating to the online system, please contact research.policy@durham.ac.uk.

3. Information sheet

Participant Information Sheet

Project title: Teachers' Use of Research Evidence for the Reading Attainment

Researcher(s): Wenqing Chen

Department: School of Education, Durham University

Contact details: wenqing.chen@durham.ac.uk

Supervisor name: Stephen Gorard, Nadia Siddiqui

Supervisor contact details:

s.a.c.gorard@durham.ac.uk

nadia.siddiqui@durham.ac.uk

Thank you for your interest in my research project. My project aims to find out the most promising approach to help teachers transform research evidence into practice and improve students' reading achievement in Key Stage 2. The project is supervised by Professors Stephen Gorard and Nadia Siddiqui from the Durham University Evidence Centre for Education (https://www.dur.ac.uk/dece/).

The project will be a teacher-level randomised controlled trial from February 2024. The study will randomly select half of the teachers who consent to participate to receive the resources in terms of using research evidence:

- An intervention menu including promising research evidence to improve reading comprehension.
- A presentation about use of research evidence.
- Workshops to help teachers adapt research evidence into daily practice.
- Online follow-up during 4 months.

The presentation and the workshop will be recorded. The rest of the teachers will receive all materials at the end of the experiment.

This research requires:

- Willingness to take part in the February 2024 term.
- Willingness to select a research leader in the teaching team.
- Willingness for consenting teachers to take part in the workshop and receive ongoing support.
- Willingness to allow the collection of survey data from consenting teachers in the school.
- Willingness to allow the collection of class mean scores of key stage 2 reading test.

All data from the school, teachers, and students will be only used for the research. The name of your school, any staff names, and pupils name will not appear in any of my reports or publications. Any contact details will only be shared within the research team for contact purpose. Participation in data collecting will be entirely voluntary, and teachers will have the option to quit at any time during the experiment.

If you have any questions or express your interest to take part in this study, please feel free to contact me.

Email: wenqing.chen@durham.ac.uk

Or Professor Stephen Gorard (s.a.c.gorard@durham.ac.uk).

4. Consent form

Consent Form

Project title: Teachers' Use of Research Evidence for the Reading Attainment

Researcher(s): Wenqing Chen

Department: School of Education, Durham University

Contact details: wenqing.chen@durham.ac.uk

Supervisor name: Stephen Gorard, Nadia Siddiqui

Supervisor contact details:

s.a.c.gorard@durham.ac.uk

nadia.siddiqui@durham.ac.uk

This form is to confirm that you understand what the purposes of the project, what is involved and that you are happy to take part. Please initial each box to indicate your agreement:

I confirm that I have read and understand the information sheet dated [9/11/2023] and the privacy notice for the above project.	
I have had sufficient time to consider the information and ask any questions I might have, and I am satisfied with the answers I have been given.	
I understand who will have access to personal data provided, how the data will be stored and what will happen to the data at the end of the project.	
I consent to being audio recorded / being video recorded / having my photo taken and understand how recordings / photos will be used in research outputs.	
I understand that my words may be quoted in publications, reports, and other research outputs.	
Please choose one of the following two options: EITHER I agree to my real name being used in the above OR I do not agree to my real name being used in the above	
I agree to take part in the above project.	
I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason.	

Participant's Signature_____ Date_____

(NAME IN BLOCK LETTERS)_____

5. briefing sheet

Debriefing Sheet

Project title: Teachers' Use of Research Evidence for the Reading Attainment

Thank you for taking part in this study. What this research investigates is whether the active multi-component approach potentially facilitate teachers' use of research evidence by providing promising research evidence and guidance. The data you have provided is to measure your perceptions, understanding, and actions of research evidence. These data are automatically anonymised and cannot be traced back to your identity. You have the right to change your mind about treatment participation at any period of the project. You could drop out of the intervention and move to the control group before and during the trial.

If you would like further information about the study or would like to know about what my findings are when all the data have been collected and analysed then please contact me on wenqing.chen@durham.ac.uk. I cannot however provide you with your individual results.

5. Privacy notice

Durham University's responsibilities under data protection legislation include the duty to ensure that we provide individuals with information about how we process personal data. We do this in a number of ways, one of which is the publication of privacy notices. This privacy notice provides a general description of the broad range of processing activity in addition there are tailored privacy notices covering some specific processing activity.

To ensure that we process your personal data fairly and lawfully we are required to inform you:

- Why we collect your data
- How it will be used
- Who it will be shared with

We will also explain what rights you have to control how we use your information and how to inform us about your wishes. Durham University will make the Privacy Notice available via the website and at the point we request personal data.

Our privacy notices comprise two parts – a generic part (ie common to all of our privacy notices) and a part tailored to the specific processing activity being undertaken.

PART 1 – GENERIC PRIVACY NOTICE

Please access our General Privacy Notice online.

PART 2 – TAILORED PRIVACY NOTICE

This section of the Privacy Notice provides you with the privacy information that you need to know before you provide personal data to the University for the particular purpose(s) stated below.

Project Title: Teachers' Use of Research Evidence for the Reading Attainment

Type(s) of personal data collected and held by the researcher and method of collection:

Personal data will be collected through school record, questionnaire, and classroom observation. This will include:

- Pupil's background characteristics such as year of birth, sex, ethnicity, and Hukou will be collected in the school record.
- Teacher's background characteristics such as experience, degree qualification, initial teacher training, and working status will be collected by teacher survey.
- Photographs, audio, or video recordings of teachers attending to the workshops and implementing the evidence-based practice in the classroom.

Lawful Basis

The legal basis is the University's public task: Under data protection legislation, we need to tell you the lawful basis we are relying on to process your data. The lawful basis we are relying on is public task: the processing is necessary for an activity being carried out as part of the University's public task, which is defined as teaching, learning and research.

For further information see https://durham.ac.uk/research.innovation/governance/ethics/governance/dp/legalbasis/

How personal data is stored:

- All personal data will be held securely and strictly confidential to the research team.
- The data will be anonymised or pseudonymised:

You will be allocated an anonymous number for data collection which will not be connected to your name or identity. Signed consent forms will be stored separately to project data.

- The storage arrangements for different types of personal data is as following:

All personal data in electronic form will be stored on a password protected computer, and any hardcopies will be kept in locked storage. Data will not be available to anyone outside the research team.

- The storage of video / audio-recording or photography is as following:

The conversation will be recorded and stored on an encrypted device until it has been transcribed by the researcher. No-one else will have access to the recording, and it will be erased once the transcript has been completed.

How personal data is processed:

This project evaluates the effectiveness of teachers' use of research evidence on students' reading attainment. I will collect your perceptions, understanding, and actions by questionnaire to measure teachers' use of research evidence.

Information will be entered into a database for analysis. After six months the data will be completely anonymised and the original records, including any information which can identify you personally, will be destroyed.

Withdrawal of data

Teachers have the right to withdraw any identifiable data at any stage of the research by emailing wenqing.chen@durham.ac.uk .

Who the researcher shares personal data with:

- Supervisors of the researcher
- Publications or other project outputs

How long personal data is held by the researcher:

I will hold personal data for six months, after which it will be anonymised.

How to object to the processing of your personal data for this project:

If you have any concerns regarding the processing of your personal data, or you wish to withdraw your data from the project, contact Wenqing Chen, wenqing.chen@durham.ac.uk .

Further information:

Researcher: Wenqing Chen

Contact details: wenqing.chen@durham.ac.uk

Supervisor name: Stephen Gorard, Nadia Siddiqui

Supervisor contact details:

s.a.c.gorard@durham.ac.uk

nadia.siddiqui@durham.ac.uk

Appendix J. Trends of teachers' access to educational research information by countries

Appendix table J-1 Trend of access to education conferences or seminars

	2008	2013	2018
Australia	64	58	59
Austria	50		50
Flemish Community (Belgium)	34	29	53
Bulgaria	43	40	40
Brazil	64	38	47
Alberta (Canada)		75	79
Chile		30	37
Shanghai (China)		80	85
Czech Republic		23	29
Denmark	43	40	45
England (United Kingdom)		30	36
Spain	35	25	50
Estonia	50	51	53
Finland		36	35
France		20	49
Georgia		19	24
Croatia		80	72
Hungary	42		34
Israel	42	46	58
Italy	43	34	53
Japan		56	63
Korea	47	46	43
Lithuania	67		60
Latvia		61	72
Mexico	34	38	45
Malta	53		43
Malaysia	34	33	
Netherlands	52	45	40
Norway	41	39	36
New Zealand		40	50
Poland	64	55	
Portugal	53		47
Romania		35	43
Russian Federation		54	70
Singapore		62	59
Slovak Republic	39	26	27
Slovenia	76		57
Sweden		45	47

	2008	2013	2018
Turkey	72		56
United States		50	55
Total	50	43	53

Appendix table J-2 Trend of access to professional literature

	2008	2018
Australia	82	86
Austria	90	89
Belgium (Flemish Community)	80	79
Brazil	84	59
Bulgaria	93	79
Denmark	78	67
Estonia	88	90
Hungary	90	89
Italy	67	25
Korea	52	85
Lithuania	93	94
Malta	63	45
Mexico	68	64
Netherlands	84	89
Norway	64	75
Portugal	73	58
Slovak Republic	93	80
Slovenia	86	92
Spain	68	47
Turkey	82	72
Total	79	73

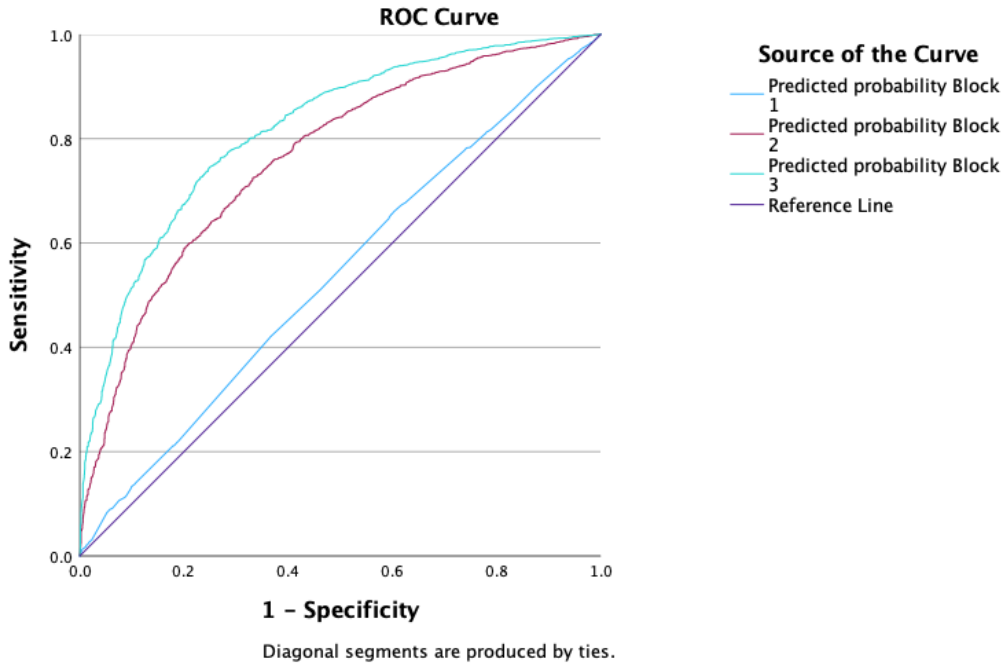
Appendix table J-3 Trend of access to Individual or collaborative research

	2008	2013
Australia	36	38
Belgium (Flemish Community)	33	19
Brazil	57	45
Bulgaria	27	23
Denmark	52	21
Estonia	26	34
Italy	57	47
Korea	50	43
Malaysia	22	25
Mexico	65	49

	2008	2013
Netherlands	44	37
Norway	13	16
Poland	40	38
Portugal	48	36
Slovak Republic	12	12
Spain	49	42
Total	38	35

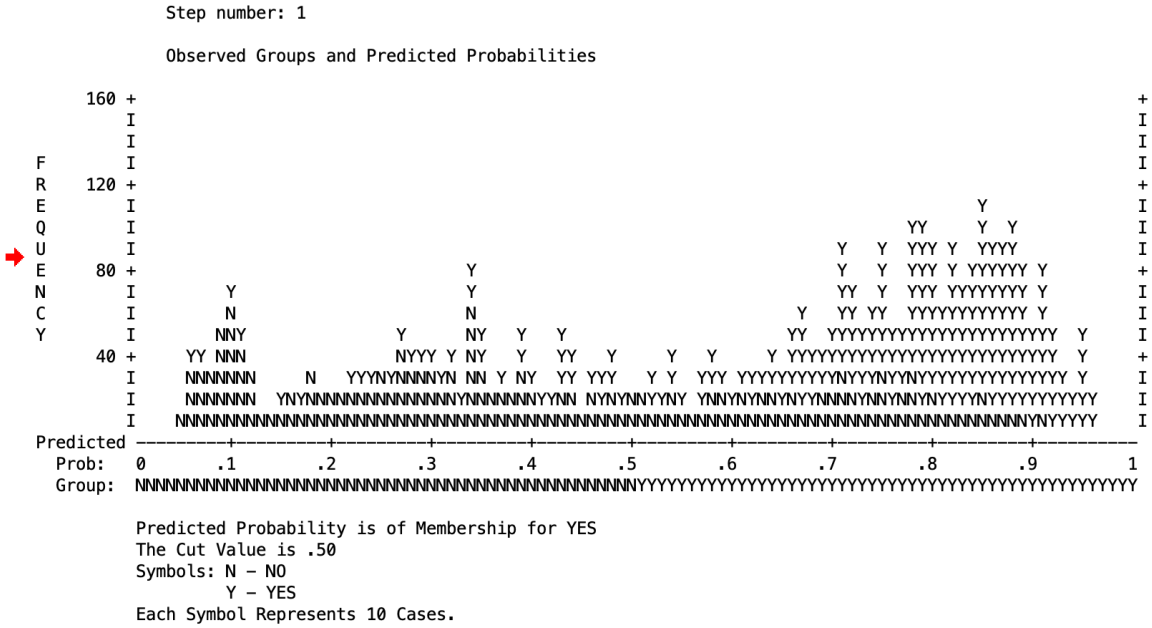
Appendix K. Refinement of the logistic regression model

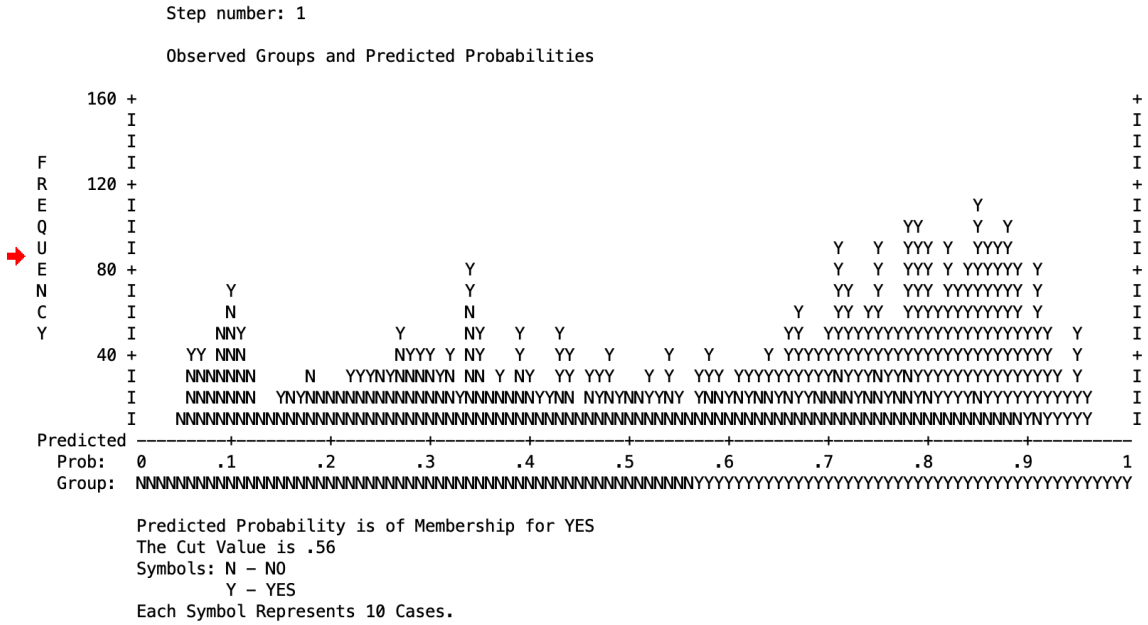
The receiver operating characteristic (ROC) curves from the original model, which included only teacher background variables, to the final, fully adjusted model are presented in Appendix figure K-1. The original model had a low discriminatory power, with an area under the curve (AUC) of 0.537, indicating performance only slightly better than random guessing. After incorporating Block 2 variables on confidence and support, the AUC increased significantly to 0.762. The final model, which further included teachers' perceptions of conducting their own research and inquiries, achieved an AUC of 0.817. This progressive improvement suggests that each set of added variables contributed meaningful predictive value, and the final model was more effective in distinguishing between teachers who perceived research evidence as useful and those who did not.



Appendix figure K-1 Receiver operating characteristic curve

The classification plot for the model illustrates the distribution of prediction results of the model, as shown in Appendix figure K-2. Y (perceiving research evidence useful) and N (not perceiving research evidence useful) in the x-axis showed the cut-off point of 0.5, while Y and N in the area of the plot were the actual outcomes.


Step number: 1
Observed Groups and Predicted Probabilities
160 +
120 +
80 +
40 +
F R E Q U E N C Y
Predicted
Prob: 0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1
Group: NNNN



Appendix figure K-3 Classification plot-amended

When the cut value was set to 0.56, the eighth N in the plot area was just diagnosed as the group 'not perceiving usefulness of research evidence'. The model could keep a balance to distinguish two categories of teachers and offer proper help. It is noticeable that when the cut value was 0.58, the percentage correct rose up to 74.9%. But in the classification plot, there would be more teachers who had already found research evidence categorised into the opposite category. The cut value of 0.58 could not be the best amended value because the model was over strict to discriminate teachers to 'not perceiving the usefulness', which may be a waste of resources.

Appendix L. Structured review data extraction

Title	Author	Transfer	Evidence type	School phase	Design	Scale	Attrition	Outcome	Rating	Results
Evidence-Based Literacy Support: The "Literacy Octopus" Trial. Evaluation Report and Executive Summary	Lord, Pippa; Rabiasz, Adam; Roy, Palak; Harland, Jennie; Styles, Ben; Fowler, Katherine	engagement in transfer	modified evidence	elementary, UK.	School level RCT; 4	540 schools (treatment) and 283 schools (control); 4	0; 4	Key Stage 2 attainment; 4	4*	No impact
Literacy Octopus' Dissemination Trial: Evaluation Report and Executive Summary	Lord, Pippa; Rabiasz, Adam; Styles, Ben	passive transfer	Modified evidence	elementary, UK.	School level RCT; 4	10,000 schools (treatment), 2,500 (control); 4	0.8%; 4	Key Stage 2 attainment; 4	4*	No impact, improved teacher knowledge and attitude
Research Learning Communities: Evaluation Report and Executive Summary.	Rose, J., Thomas, S., Zhang, L., Edwards, A., Augero, A., & Roney, P.	engagement in transfer	Modified evidence	Elementary, UK.	School level RCT; 4	60 schools (treatment), 59 school (control); 5,462 students; 4	school n=3, student n=176; 4	Key stage 2 attainment; 4	4*	ES = +0.02, improved teachers' positive disposition, no impact on knowledge

Title	Author	Transfer	Evidence type	School phase	Design	Scale	Attrition	Outcome	Rating	Results
Teachers' use of research evidence in practice: a pilot study of feedback to enhance learning	See, B. H., Gorard, S., & Siddiqui, N.	passive transfer	plain evidence	elementary, UK.	quasi-experiment; 3	9 schools (treatment), 49 schools (comparison); 3	0; 3	Key Stage 2 attainment; 3	3*	ES=+0.06 (grade 6), ES=-0.04 (all), Feasibility of evidence use
The RISE project: evidence-informed school improvement. Evaluation report	Wiggins, M	engagement in transfer	Modified evidence	Secondary, UK.	School level RCT; 4	20 schools (treatment), 20schools (control); 4	9~13%; 3	GCSE; 3	3*	ES=+0.05 First year, ES=+0.03 Second year
Evidence of efficacy of the Integrated Literacy Study Group professional learning program to enhance reading instruction for students with emotional and behavioral disorders	Benner, GJ; Filderman, MJ; Barnard-Brak, L; Pennefather, J; Smith, JLM; Strycker, LA	engagement in transfer	engineered evidence	elementary, US.	Teacher level RCT; 4	74 teachers, 148 students; 2	0; 2	Dynamic Indicators of Basic Early Literacy Skills;2	2*	ES = +0.56, improved teachers' attitude, no impact on knowledge

Title	Author	Transfer	Evidence type	School phase	Design	Scale	Attrition	Outcome	Rating	Results
Implementation and effects of explicit reading comprehension instruction in fifth-grade classrooms	Andreassen, R; Braten, I	engagement in transfer	engineered evidence	elementary, Norway	Quasi-experiment; 3	103 students(treatment);,113 student (comparison); 2	13%; 2	Researcher-generated test (RGT); 2	2*	No impact, mix impact on practice
Literacy Design Collaborative 2018-2019 Evaluation Report. CRESST Report 867. Revised.	Wang, J., Herman, J. L., Epstein, S., Leon, S., La Torre, D., & Bozeman, V.	engagement in transfer	engineered evidence	elementary and secondary, US.	Matching comparison group; quasi-experiment; 3	(Teacher, Administor, Schools) Cohort 1: 154,34,20; Cohort 2: 199,42,31;| 264, 34; 3	C1: 81%, 59%, 45%; C2:39%, 36%, 26%; 2	State ELA Assessment Scores; 2	2*	S.D.=0.021 (primary), S.D. = 0.281 (secondary), improved teacher attitude and practice
When evidence is not enough: Findings from a randomized evaluation of Evidence-Based Literacy Instruction (EBLI)	Jacob, B	engagement in transfer	engineered evidence	elementary, US.	Teacher level RCT; 4	22 teachers (treatment), 41 teachers (control); 2	5; 2	MAP reading test; 2	2*	No impact, no impact on teachers' attitude and practice

Title	Author	Transfer	Evidence type	School phase	Design	Scale	Attrition	Outcome	Rating	Results
A building-based case study of evidence-based literacy practices: Implementation, reading behavior, and growth in reading fluency, K-4	Greenwood, CR; Tapia, Y; Abbott, M; Walton, C	engagement in transfer	Modified evidence	kindergarten and elementary, US.	Quasi-experiment; 2	36 students; 1	9 (replaced); 1	Reading CBM; 1	1*	Claimed success (no accurate effect size reported), improved teaching practice
The CIERA school change framework: An evidence-based approach to professional development and school reading improvement	Taylor B.M., Pearson P.D., Peterson D.S., Rodriguez M.C.	passive transfer	modified evidence	elementary, US.	Pre/post test; 1	13 schools, 733 students; 1	0; 1	Gates–MacGinitie, words correct per minute, and a directed writing assessment;1	1*	ES = +0.29 reading comprehension, ES = +0.38 reading fluency, ES = +0.38 writing
Utilizing implementation data to explain outcomes within a theory -driven evaluation model	Edmonds, Meaghan Suzanne	passive transfer	engineered evidence	elementary, US.	School level RCT; 4	5 schools, 16 teachers, 309 students; 2	20 students; 1	Gates MacGinitie Reading Tests 4th edition (GMRT); 1	1*	No impact

Title	Author	Transfer	Evidence type	School phase	Design	Scale	Attrition	Outcome	Rating	Results
The Word Nerds project: Findings from a research–practice partnership focused on spelling instruction.	Tortorelli, Laura S.; Bruner, Lori	engagement in transfer	Modified evidence	elementary, US.	Pre/post test; 1	17 teachers, 205 students; 1	not reported; 0	Primary Spelling Inventory (PSI); Alternative Primary Spelling Inventory (APSI)I; 0	0	No impact

Appendix M. Impact analysis for the pilot

1. Sample description

Two pilot schools were recruited in Hefei, a city in the southeast of China. They used the same curricula and attended to the same standardised final tests. One primary school had 8 classes per grade while another had 3 classes per grade with approximately 45 students in a class. 46 teachers from grade 3 to 6 at two primary schools participated in the pilot study.

1) School background

Two public primary schools in Hefei, Anhui, China participated in this pilot study. Both of them are under the authority of the Baohe District, which should mean that the same final examination and rating standard is used. The two schools locate at the centre of the city, therefore, almost all students have urban 'hukou'. The two schools belong to the same school union with a good reputation in the area.

In 2004, School I selected teachers and administrative staff to monitor a new campus which became School II today. School I contains 126 teachers and 52 classes from grade 1 to 6, while School II has 44 teachers and 18 classes. The two schools involved in the pilot are of different school sizes. The pilot study focused on students from grade 3 to grade 6 who were at the same stage as Key Stage 2 students in England. 46 teachers and 2,033 students from these two schools were involved in this project.

2) Teacher background

46 teachers in total participated in this pilot study, involving 33 teachers from School I and 13 teachers from School II. All teachers were Han ethnicity, and 96% of them were female. 91% teachers had an undergraduate degree, while 6.5% of them had a master's. Only one teacher had a degree from a 'dazhuan' (college) below the undergraduate level. None of them had a doctoral degree.

In their highest degree, a majority of them (39%) majored in Chinese literacy, with teacher education (Chinese) and general education coming in second place (28%). 10% teachers had more than one major, 80% of whom are relevant to Chinese.

Teachers ranged in instructional practice experience, from those who have been in the field for more than 40 years to others who have just been employed there for a year (mean of 20 years). Approximately 96% teachers have signed the long-term contract (with 'bianzhi'). The two teachers

who do not have 'bianzhi' are all from School I. Five teachers had roles in both teaching and research, such as instructional leaders. Seven teachers were also principals or heads of grade departments, among other administrative positions. The two groups had balanced allocation in terms of ethnicity, highest degree qualification, employment status, and research work (Appendix table M-1).

Appendix table M-1 Teacher background characteristics

		Treatment group	Control group	Total
Sex	female	46	50	96
	male	4	0	4
Ethnicity	Han	50	50	100
Highest degree qualification	dazhuan / college	0	2	2
	undergraduate	46	46	92
	master	4	2	6
Major	Teacher education_Chinese	13	15	28
	Teacher education_others	4	7	11
	Curriculum and pedagogy_Chinese	2	0	2
	Curriculum and pedagogy_others	0	2	2
	General education	17	11	28
	Chinese language and literature	22	17	39
	Applied linguistics	0	2	2
	others	2	2	4
	More than one_Chinese	4	4	8
	More than one_non Chinese	2	0	2
Employment status	short term/ without 'bianzhi'	0	4	4
	long term/ with 'bianzhi'	50	46	96
Research work		7	4	11
Administrative role		11	4	15
Total number		23	23	46

3) Student background

1,553 students from School I and 529 students from School II participated in the study. 54% of them were male. 97% of the pupils were identified as Han Chinese. These two schools were not constructed for rural migrant workers' children only, therefore, only 2% students have rural 'hukou', none of whom had agricultural 'hukou'. Appendix table M-2 shows students characteristics and the group allocation. Students were well-allocated in sex, ethnicity, and 'Hukou'.

Appendix table M-2 Students' background characteristics

		Treatment (%)	Control (%)	N
Sex	female	44.3	45.1	931
	male	54.5	53.6	1,125
Ethnicity minority		1.8	2.1	40
Rural Hukou		2.1	1.7	40
System missing		1.2	1.3	26
Total		1,078	1,004	2,082

2,033 students participated in the mid-term examination, involving 1,060 students from the treatment group and 973 from the comparison group. The pre-test scores for two groups of students were nearly balanced (Appendix table M-3).

Appendix table M-3 Comparison of Chinese pre-test scores, pilot

	N	Mean	Std. Deviation	Effect size
Treatment	1,060	0.01	0.99	+0.02
Control	973	-0.01	1.01	
Total	2,033	0.00	1.00	

4) Attrition

2,033 students are included in the impact analysis because 49 students' pre-test scores were missing. They did not participate in the mid-term examination for authorised absence such as illness and personal leave. No class dropped out of the study. 3 students from the treatment group and 8 students from the comparison group missed the final term examination. The NNTD to observe an effect size larger than +0.05 would be 49 which suggested that 11 absentees could hardly make any major difference in the results.

46 teachers answered the pre-test teacher survey while 42 teachers answered the post-test survey. The response rate of the teacher survey was 91%. 1,678 students answered the student joy of reading survey with a response rate of 80%.

2. Impact evaluation

The impact evaluation includes student Chinese literacy attainment as the primary outcome, while students' reading attitudes and teachers' perception and knowledge of using research evidence are analysed as the non-cognitive outcome and the second outcome.

1) Impact on students' Chinese literacy

No major difference in students' final examination scores was found (-0.01). Treatment students did not make more progress than control group students (-0.06). Appendix table M-4 illustrates the effect sizes of z-score.

Appendix table M-4 Comparison of standardised mid-term, final, and gain scores, all pilot pupils

	N	Mean			Standard deviation		Effect size	
		Mid	Final	Gain	Final	Gain	Final	Gain
Treatment	1,057	0.01	-0.01	-0.02			-0.01	-0.06
Control	965	-0.01	0.01	0.02				
Total	2,022	0.00	0.00	0.00	1.00	0.78		

The following table illustrates the effect sizes of the strategies respectively (Appendix table M-5). Dialogic teaching, peer tutoring, and Writing about Values had the effect sizes of +0.40, -0.12, and +0.03 respectively for the post-test.

Appendix table M-5 Summary of effect sizes, by strategies

Intervention	Number of pupils	post-test	gain
Use of research evidence	2,022	-0.01	-0.06
Dialogic teaching	111	-0.12	+0.40
Peer tutoring	1,271	+0.05	-0.12
Writing about Values	640	-0.11	+0.03

2) Impact on students' enjoyment of reading

Student enjoyment of Chinese reading is the non-cognitive outcome of this study. 1,678 students from grades 3 to 6 completed the survey at the end of the semester. The survey involved three aspects: favourite reading material, reading attitudes, and time for Chinese homework and leisure reading. Appendix table M-6 compares students' reading enjoyment between the control group and the treatment group. No positive impact was observed at the end of the intervention. As the study did not include a pre-test, this difference may have stemmed from pre-existing differences in students' enjoyment of reading.

Appendix table M-6 Comparison of students' enjoyment of reading

	Treatment	Control	Standard deviation	Effect size
Chinese textbook	8.01	8.27	2.20	-0.12
Magazine, newspaper	7.16	7.22	2.80	-0.02
Comic books	6.61	6.94	3.25	-0.10

Fiction	7.99	8.08	2.31	-0.04
Science	7.60	7.97	2.51	-0.15
Social Science	7.50	7.79	2.57	-0.11
I read not only if I have to.	7.28	7.22	2.60	-0.02
Reading is one of my favourite hobbies.	7.81	7.42	2.56	-0.15
I like discussing the books with others	7.46	6.98	2.81	-0.17
For me, reading is not a waste of time.	8.19	8.23	1.85	+0.02
I read not only to get information that I need.	7.30	7.27	2.54	-0.01
I do not have to read a text several times before completely understanding it.	6.27	6.14	2.95	-0.04
I find it easy to answer questions about a text.	6.71	6.58	2.61	-0.05
I read fluently.	7.28	7.22	2.60	-0.02

Intervened students had longer leisure reading time and Chinese homework time compared with controlled students, while the proportion of leisure reading, and total Chinese learning time were almost the same for the two groups (Appendix table M-7).

Appendix table M-7 Comparison of Chinese homework time and leisure reading time

	Control	Treatment	Standard deviation	Effect size
Leisure reading time	1.30	1.40	0.90	+0.10
Chinese homework time	0.99	1.08	0.61	+0.15
Ratio	0.57	0.56	--	--

3) Impact on teachers' knowledge and ability of using research evidence

The secondary outcome of this study was about teachers' perceptions and basic knowledge of research evidence, collected by teacher surveys before and after the intervention. The survey comprised 36 items including the most trustworthy sources for decision-making, description of research evidence, confidence, barriers to use research evidence, and access and understanding of research evidence. Teachers rated from 0 to 10 to express their agreement with the statement.

The pre-survey also demanded teachers' majors of their highest degree qualification. Teachers from the treatment group and the control group were compared with post-test scores and gain scores. School II used the paper survey in the pre-test and had lots of missing values. Therefore, the gain scores were mainly based on teachers at School I.

Teachers from the treatment group were more suspicious of all sources used for decision-making. The treatment group might have misunderstood research evidence, which suggests that the intervention lacked a clear introduction to research evidence. Intervened teachers reported more frequent access

to research articles (+0.2), professional development conferences (+0.7), and social media (+0.6) compared with control group teachers, on average. The intervention enhanced teachers' confidence in understanding research findings (+0.5), judging research quality (+0.3), and adapting research evidence for teaching practice (+0.2). Treatment group teachers were more likely to perceive research findings as useful (+0.4). To better support the incorporation of research evidence into teaching, treatment group teachers reported a greater need for guidance from researchers (+0.1). The following Appendix table M-8 reports the impact on teachers' perceptions of research evidence.

Appendix table M-8 Effect sizes of teachers' perceptions

Items	Gain effect size
Projects and materials from school/ local system	-0.2
Colleagues	-0.2
High profile education figures	-0.3
Academic researchers	-0.1
China Education Guidance from Department of Education	+0.1
Educational news reports/ articles	+0.2
None of these sources	+0.4
Books and articles by educators and psychologists disseminating their views.	-0.2
Knowledge from observations, information analysis, and discussions in my school.	+0.4
Findings from robust research conducted by experts in education.	-0.2
Searching engines e.g. Baidu	+0.1
Academic Databases e.g. China National Knowledge Infrastructure	0.0
Reports, books, and articles published by the government, schools or teachers	-0.1
Research articles or summaries published by researchers	+0.2
Professional development conferences	+0.7
Social media e.g. TED, WeChat blogs, Zhihu	+0.6
I can understand the findings of research.	+0.5
I can decide whether the findings of research are trustworthy.	+0.3
I can select appropriate research evidence to support my teaching practice.	0.0
I can adapt research for my teaching practice.	+0.2
I can conduct my own robust research.	-0.2
Research findings are useful to improve my teaching	+0.4
School provides a more evidence-based syllabus	-0.2
I have less managerial or teaching workload	-0.5
I get free access to the latest education research evidence	-0.3
Researchers give us specific training and instructions	+0.1
I observe other teachers' class using evidence-based practice	-0.1

The use-of-research-evidence intervention improved teachers' skills in interpreting research findings (+0.46), judging research quality (+0.46) and making evidence-led decisions (+0.55). The results are as following (Appendix table M-9).

Appendix table M-9 Comparison of teachers' skills of research evidence

Ability		N	Mean		Standard deviation		Effect size	
			Post	Gain	Post	Gain	Post	Gain
Impact	Treatment	18	2.45	0.06	1.01	0.74	+0.15	+0.46
	Control	14	2.30	-0.50	1.08	1.03		
	Total	32	2.38	-0.28	1.06	1.20		
Research quality	Treatment	21	0.18	-0.10	0.39	0.55	+0.08	+0.46
	Control	18	0.15	-0.40	0.37	0.74		
	Total	30	0.17	-0.23	0.38	0.65		
Decision making	Treatment	17	1.73	0.24	1.61	1.89	-0.46	+0.55
	Control	14	2.40	-0.71	1.19	1.44		
	Total	31	2.05	-0.19	1.45	1.74		

Appendix N. Impact analysis including the quasi-experiment school

1 . Chinese literacy

1,116 students were involved in the experiment. However, the quasi-experiment school only provided the pre-and the post-tests scores of 302 students. 45 of them were at grade 2, which did not have an adequate comparison group from other schools. This analysis only included 257 students from grades 3 and 4. Students' Chinese literacy records were transformed to z-scores, merged with RCT schools, labelled as treatment students, and compared to the control group. The missing case rate was high at this QE school, and it did not have a matched comparison group, meaning that the findings from this school should be treated with caution.

The group consisted of a slightly higher proportion of male students, making up 55% of the total, while 97% of them were identified of Han ethnicity. Compared to students from RCT schools (Section 10.1.1), participant students at QE school had a more privileged background. Only one student had rural Hukou, and none of the students were migrants or from low-income backgrounds. Appendix table N-1 presents grades of students from QE school in the pre-and the post-test. In the post-test, there was no students getting C- or D. However, the proportion of students obtaining A+ and A decreased.

Appendix table N-1 Student grades distribution, QE school only

	%-pre	%-post
A+	5.1	1.6
A	54.9	47.1
A-	26.1	26.8
B+	7.4	10.1
B	3.1	8.6
B-	1.6	3.9
C+	0.0	0.8
C	1.2	1.2
C-	0.4	0.0
D	0.4	0.0
Total N		257

The average scores of pre-, post-tests, and gain were compared, as shown in Appendix table N-2. Consistent with the main analysis, there was no major difference in students' post-test performance and small positive impact on gain scores (+0.06). Combining all students receiving the intervention, treatment group students still made more progress during the experiment semester.

Appendix table N-2 Comparison of pre, post-tests, and gain scores, QE included

	N	Mean			Standard deviation		Effect size	
		pre	post	gain	post	gain	post	gain
Treatment	3,622	-0.01	0.00	0.02			0.00	+0.06
Control	3,426	0.02	0.00	-0.03				
Total	7,048	0.00	0.00	0.00	0.99	0.75		

However, compared to the main analysis (gain effect size =+0.07), the progress was slightly smaller when including the quasi-experiment school. The average gain score of the QE school students was smaller than the treatment group from RCT schools. This may suggest that the impact on quasi-experiment school students was not as large as other RCT schools. Unlike the RCT schools, the QE school appeared to have managed the experiment independently and did not complete the two required workshops. Teachers' participation, school management for the experiment, and implementation in the classroom could have contributed to the smaller effect, which are further discussed.

The results for the QE school should be interpreted carefully. They selected 30 classes to participate in the treatment, but only reported grades of 7 classes. Even within a quasi-experiment, missing of grades from the majority of participant classes may introduce bias in the evaluation of the intervention. Additionally, the grading criteria were unclear, and the school research lead did not directly respond to inquiries about the standards. The z-score was computed, therefore, from these reported classes, which may lead to issues with comparability. This may cause issue of incomparable in data. The ideal approach should be getting all students' scores for z-scores.

2. Impact on students' enjoyment of reading

593 students at the quasi-experiment school participated in the survey at the beginning of the survey and 406 students finished the post-test survey. However, only 131 students from grade 3 to 6 completed both the pre- and the post-surveys. Thus, the result of this additional analysis should be interpreted with caution. Overall, the findings were almost consistent with the findings based on RCT schools (Appendix table N-3).

Appendix table N-3 Effect sizes of students' reading enjoyment, including the QE school

Items	pre	post	gain
Chinese textbooks	+0.01	+0.02	+0.01
Magazines	-0.01	-0.05	-0.04
Comic books	+0.05	-0.08	-0.13
Fiction	+0.01	+0.04	+0.02

Science	+0.03	-0.07	-0.09
Social Science	+0.01	-0.02	-0.03
borrow books from school library.	-0.09	-0.12	-0.03
read online.	-0.12	-0.07	+0.05
read paper books.	0.00	+0.01	+0.01
read for information.	0.00	-0.03	-0.02
read for enjoyment.	+0.06	+0.03	-0.03
I read not only if I have to.	+0.05	+0.07	+0.01
Reading is one of my favourite hobbies.	0.00	+0.08	+0.08
I like talking about books with other people.	+0.03	+0.06	+0.02
I would be happy if someone gave me a book as a present.	0.00	+0.02	+0.02
For me, reading is not a waste of time.	+0.02	+0.05	+0.02
I read not only to get information that I need.	+0.02	+0.09	+0.05
I usually do well in reading.	+0.05	+0.04	-0.01
Reading is easier for me than for many of my classmates.	+0.04	+0.08	+0.03
Chinese is easier for me than any other subject.	+0.01	+0.02	+0.02
I do not have to read a text several times before completely understanding it.	+0.01	+0.08	+0.06
I find it easy to answer questions about a text.	+0.01	-0.09	-0.08

3. Impact on teachers' views and skills for using research evidence

11 teachers in the quasi-experiment (QE) completed the pre-test and 18 teachers participated in the post-test survey. 8 teachers from the QE group completed both the pre-and the post-surveys. This section described teachers' characteristics in the quasi-experiment and reported the impact analysis including teachers from the quasi-experiment.

1) Teachers in the quasi-experiment(QE)

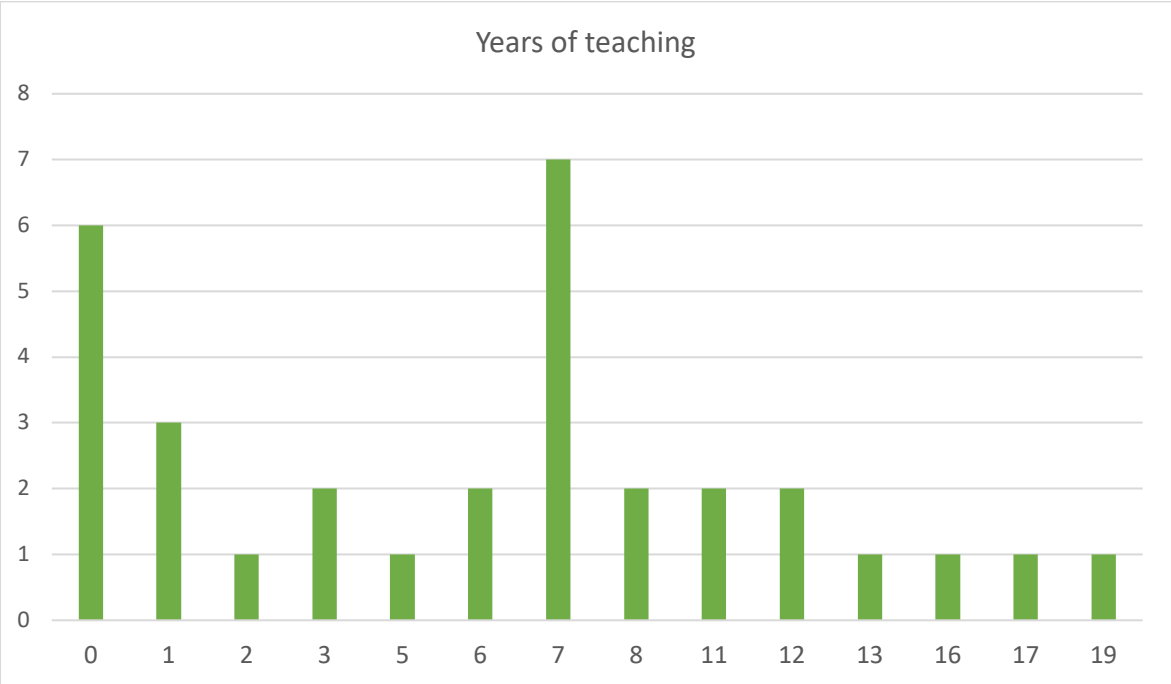
School E preferred to participate in a quasi-experiment and nominated 32 teachers to join the treatment group. 32 teachers in School E participated in the quasi-experiment. Their background information and teaching experience are illustrated in Appendix table N-4, Appendix table N-5, and Appendix figure N-1.

Most of the teachers, as teachers in the RCT, were females, of Han ethnicity, and in permanent employment status. However, a higher proportion of these teachers had a Master's degree or higher compared to the overall RCT teachers. None of the 32 teachers had a junior college diploma.

Appendix table N-4 Teacher background information in QE

		%
Sex	Female	84.4
	Male	15.6
Ethnicity	Han	100
Highest degree qualification	Bachelor	75
	Master or higher	25
Permanent employment status		100
Research work		12.5
Administrative role		3.1
Total N		32

The QE teachers were mainly young teachers with an average of 6.38 years of teaching experience and a median of 7 years. Appendix figure N-1 illustrates the number of teachers by years of teaching.



Appendix figure N-1 Years of teaching in QE

11 teachers provided supplementary information about their major subjects for their highest degree qualification as shown in Appendix table N-5.

Appendix table N-5 Teacher major subjects for their highest degree qualification in QE

	N
Teacher Education - Chinese	4
General education	1
Chinese Language and Literature	1

Others	6
Missing	21

The proportion of teachers who majored in other subjects such as animation, management, and software engineering were higher than other major subjects. However, due to the large amount of missing, information on teachers' major subjects at School E is limited. In contrast to teachers in the RCT, the QE group consisted of early-career teachers with stronger educational backgrounds and greater diversity in their major subjects.

2) Impact on teachers' evidence-use, QE school included

Including teachers at the QE school made little change in the main analysis results of impact on teachers' perceptions on research evidence (Appendix table N-6).

Appendix table N-6 Effect sizes of teachers' perceptions on research evidence, QE school included

Items	post	gain
Promising projects and materials from your school, or the local system	-0.14	-0.06
Colleagues e.g. informal discussion, classroom observation feedback	+0.11	+0.12
High profile education figures e.g. authors of edu-books, education consultants	+0.08	+0.13
Academic researchers	-0.14	-0.02
China Education Guidance from Department of Education	-0.01	-0.02
Educational news reports/ articles e.g. Teachers in primary schools in China	-0.18	-0.03
None of these sources	-0.33	-0.13
Other	-0.44	-0.06
Search engines e.g. Baidu	-0.22	-0.20
Academic Databases e.g. China National Knowledge Infrastructure	+0.06	+0.02
Reports, books, and articles published by the government, schools or teachers	+0.12	+0.09
Research articles or summaries published by researchers	-0.06	+0.01
Professional development conferences	-0.05	+0.05
I can understand the findings of research	-0.10	-0.11
I can decide whether the findings of research are trustworthy	-0.13	-0.19
I can select appropriate research evidence to support my teaching practice	+0.18	+0.07
I can adapt research for my teaching practice	+0.09	+0.11
I can conduct my own robust research	+0.21	+0.18
Research findings is useful to improve my teaching	+0.15	+0.26
Social media e.g. TED, WeChat blogs, Zhihu	+0.18	+0.19
I can understand the findings of research	-0.10	-0.11
I can decide whether the findings of research are trustworthy	-0.13	-0.19
I can select appropriate research evidence to support my teaching practice	+0.18	+0.07
I can adapt research for my teaching practice	+0.09	+0.11
I can conduct my own robust research	+0.21	+0.18
Research findings are useful to improve my teaching	+0.15	+0.26
School provides a more evidence-based syllabi	-0.15	-0.03
I have less managerial or teaching workload	+0.03	-0.02

I get free access to the latest education research evidence	-0.12	-0.16
Researchers give us specific training and instructions	+0.05	-0.09
I observe other teachers' class using evidence-based practice	+0.15	+0.05

The intervention improved teachers' basic knowledge of research evidence (Appendix table N-7). Treatment teachers had higher correct rate of describing research evidence (+0.20) and understanding impact (+0.08). However, their ability of judging research quality was lower than the comparison group.

Appendix table N-7 Knowledge of research evidence, QE school included

	Effect size	
	post	gain
Correct description of research evidence	+0.01	+0.20
Understanding impact	+0.11	+0.08
Correct identification of research quality	-0.12	-0.17
Research quality difference	-0.03	+0.07
Evidence-led decision-making	+0.45	+0.02

Appendix O. Truth table

Experienced	Teacher education	Workshop		Usage frequency	Demonstration lessons		Ability				N	Outcome
		1st	2nd		Attend	Provide	Understand	Distinguish	Describe	Decide		
0	1	1	0	1	0	0	1	0	0	0	1	1
0	1	0	1	0	0	0	0	1	0	0	1	1
1	1	0	0	1	0	0	0	1	0	0	1	1
0	1	1	0	0	1	0	0	1	0	0	1	1
0	1	1	0	1	1	0	0	1	0	0	1	1
0	1	1	1	1	0	1	0	1	0	0	1	1
1	0	1	1	1	1	1	1	1	0	0	1	1
0	0	0	1	0	1	0	0	1	0	0	1	1
0	1	1	1	1	0	1	0	1	0	0	1	1
0	1	1	1	1	1	1	0	0	0	1	1	1
1	1	1	1	1	0	0	1	0	0	1	1	1
1	1	1	1	1	1	1	1	0	0	1	1	1
0	1	0	1	0	1	0	0	1	0	1	1	1
0	0	1	1	1	0	1	0	1	0	1	1	1
0	0	0	1	1	1	1	0	1	0	1	1	1
0	0	1	1	1	1	1	0	1	0	1	1	1
1	1	0	0	0	1	0	1	1	0	1	1	1
0	1	1	1	0	0	1	1	1	0	1	1	1
0	0	0	0	1	0	1	1	1	0	1	1	1
0	0	1	0	1	1	1	1	1	0	1	1	1
1	1	1	1	1	1	0	0	1	0	1	1	1
0	0	1	1	1	0	1	1	1	0	1	1	1

Experienced	Teacher education	Workshop		Usage frequency	Demonstration lessons		Ability				N	Outcome
		1st	2nd		Attend	Provide	Understand	Distinguish	Describe	Decide		
0	1	1	0	1	0	0	1	0	1	1	1	1
0	0	1	0	1	1	0	0	1	1	1	1	1
0	0	0	1	0	1	0	1	1	1	1	1	1
1	1	1	1	1	0	0	0	1	0	0	1	1
0	1	1	0	1	1	0	0	1	0	0	1	1
1	0	1	1	1	1	1	0	1	0	0	1	1
0	0	0	1	0	0	0	1	1	0	0	1	1
1	0	1	1	1	0	1	1	1	0	0	1	1
0	0	1	1	1	1	1	0	1	0	0	1	1
0	0	0	1	0	1	0	0	1	0	1	1	1
1	1	0	1	0	1	0	0	1	0	1	1	1
1	1	1	1	1	1	1	0	1	0	1	1	1
0	0	1	1	1	0	1	1	1	0	1	1	1
1	0	1	1	1	1	1	1	1	1	1	1	1
0	1	1	0	0	1	0	0	1	0	1	2	0
1	1	0	0	1	0	0	0	0	0	0	1	0
0	1	1	0	1	0	0	0	1	0	0	1	0
0	0	1	0	0	1	0	0	1	0	0	1	0
0	1	0	1	0	1	0	0	1	0	0	1	0
1	0	1	0	0	0	1	0	1	0	0	1	0
1	0	1	1	1	1	1	0	1	0	0	1	0
1	0	1	0	0	1	0	1	0	0	1	1	0
0	0	0	0	0	1	0	0	1	0	1	1	0

Experienced	Teacher education	Workshop		Usage frequency	Demonstration lessons		Ability				N	Outcome
		1st	2nd		Attend	Provide	Understand	Distinguish	Describe	Decide		
0	1	1	0	0	0	1	0	1	0	1	1	0
1	0	0	1	1	1	1	0	0	1	1	1	0
0	0	0	0	1	0	0	1	0	0	0	1	0
0	1	1	1	1	1	1	1	0	0	0	1	0
0	0	0	1	1	0	0	0	1	0	0	1	0
1	1	1	0	1	1	1	0	1	0	0	1	0
0	0	1	0	0	1	1	1	1	0	0	1	0
0	1	0	0	1	0	0	0	1	1	0	1	0
0	0	1	1	1	1	1	0	0	0	1	1	0
1	0	1	1	1	0	0	1	0	0	1	1	0
0	1	1	0	0	0	0	0	1	0	1	1	0
0	1	1	1	1	1	1	0	1	0	1	1	0
1	0	0	0	1	0	0	1	1	0	1	1	0
1	1	0	0	1	0	0	1	1	0	1	1	0
1	0	1	0	1	1	1	0	1	0	1	1	0
0	1	1	1	1	1	1	0	0	1	1	1	0
1	0	1	1	1	1	1	0	1	1	1	1	0
1	1	1	1	0	1	1	1	1	1	1	1	0

Appendix P. Example classroom observation record

Date	19th March, 2024
Class	Class 501, Huashanlu Primary School
Topic	牧场之国 (A nation of ranches)
Applied strategy	Peer tutoring
Record & summary	Peer reading: the teacher illustrated key sentences for students to read in peers. Peer tutoring for two paragraphs: Read the paragraphs fluently, Highlight the sentences that you find them interesting, How do you imagine the scenario? What does the "real Netherlands" look like, based on the authors' expression? Read again the paragraphs.
Evaluation	The teacher used the strategy fluently. Activities/ techniques were applied correctly. Did the teacher organise the time properly? Need improvement. The teacher only used peer tutoring for less than 10 minutes, which did not meet the requirement of peer tutoring. The teacher managed the classroom well. Need improvement. The teacher participated in group discussion randomly, without a focus list (so students need more support may not get enough support). How were students' reaction? Were they engaged in the activities? Need improvement. The tasks for peer tutoring might be too easy for Year 5 students so they just participated in the activity and finished the task without much effort. Students may not need this kind of "tutoring".
Comments	Overall, the teacher managed the classroom well. Some suggestions were made: Peer tutoring should reach at least 15 minutes during a lesson. A focusing list of students may be useful for teachers to pay more attention to students who need extra support, especially for large class size. More difficult tasks could be proposed to engage students in the lesson.
Teacher feedback	Students could reach the teaching aim in the first peer reading task. However, in the second task, students found out the response to the proposed questions but could not reach the aim of reading fluently. "I do not think students reached the second teaching aim."

Date	22nd March, 2024
Class	Class 502, Yangguang Primary School
Topic	Growth mindset workshop 1
Applied strategy	Growth mindset
Record & summary	Introduction: based on the books you read recently, which character do you think is the cleverest? What are the characteristics of "clever"? Question and discussion 1: What makes you clever? Use the notes. Students' answers: Learning, thinking, talent, environment, frustration, practice, reading… Activity 1: Read the Growth mindset reading material. Again, answer the question, what makes you clever? + learning enthusiasm, choice Question and discussion 2: Do you think you are clever? When do you think you are clever? Situation: your friend A got 100 in a recent quiz, while you only got 85. How do you think and what should you do? Activity 2: Watch Dweck Speech: Growth mindset.
Evaluation	The teacher used the strategy fluently? Activities/ techniques were applied correctly. Did the teacher organise the time properly? Activity 2 took up too much time and students were not patient and had difficulties to understand. Could the teacher manage the classroom? Students were well organised but some questions lacked logic. The teacher could not respond to students' reaction and catch the point to move on. How were students' reaction? Were they engaged in the activities? Students performed well in the discussion session but lost patience in Activity 2.
Comments	Overall, the teacher passed on basic ideas of Growth mindset to students. All students read the Growth mindset reading material which was the same intervention developed by Dweck. To better organise the lesson, some suggestions were posed: The intervention material should be delivered, when necessary. Or, students could easily guess what kind of answer the teacher expected. Questions should be clear and closely relevant to Growth mindset (we discussed "be clever" or "become cleverer"). Activities could be simplified to better disseminate Growth mindset.
Teacher feedback	It is a challenge to use "Growth mindset". How to point out "Growth mindset" is still a problem. The idea is not fully developed and could be dig deeper. But this topic inspired teachers and students a lot.