Student Evaluation of Teaching:

Can it raise attainment in secondary schools? A cluster randomised controlled trial.

Stuart James Martin Kime

Submitted in partial fulfilment of the qualification of Doctorate of Philosophy in Education

School of Education, Durham University

2017

Abstract

The effectiveness of teaching is cited as one of the most significant actors on students' learning (Centra, 1993; Creemers, 1994; Marsh, Nagengast, Fletcher, & Televantou, 2011), yet the measurement of it is an imperfect and problematic activity, not least due to the lack of consensus over a definition of effective teaching and, consequently, no consistently-used measurement instrument(s) designed for the purpose (Chingos & Peterson, 2011; Goldhaber & Anthony, 2004; Kane, McCaffrey, Miller, & Staiger, 2013).

This thesis describes a study designed to validate an instrument intended to capture secondary school students' evaluation of teaching (SET) ratings in 36 UK schools during the 2012-13 academic year, and a randomised controlled trial (RCT) to evaluate the impact on student progress of a peer-coached consultative feedback intervention for teachers.

The thesis concludes that the SecondarySEEQ instrument is valid and reliable for the purpose of capturing AS and A Level students' perceptions of the quality of teaching they receive from their teachers, a finding which adds weight to the case for SETs as a useful component of a diagnostic feedback system for teachers. I also conclude that the peer-coached intervention had a no detectable effect on student outcomes.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ALIS	Advanced Level Information System
CACE	Complier Average Causal Effect
CEM	Centre for Evaluation and Monitoring
EEF	Education Endowment Foundation
ITT	Intention To Treat
GCSE	General Certificate of Secondary Education
RCT	Randomised Controlled Trial
SEEQ	Student Evaluation of Educational Quality
SecondarySEEQ	Secondary Student Evaluation of Educational Quality
SfBT	Solution-focused Brief Therapy

Acknowledgements

Studying for a Doctorate in Education has been a labour of love. While this thesis bears my name on its title page, it is the work of thousands: to the students, teachers and school leaders who volunteered their time, enthusiasm and advice, I am hugely grateful.

Both this thesis and my own identity as a researcher have been shaped in large part by the guidance of my two supervisors, Professor Rob Coe and Professor Steve Higgins. Both have been inspirational beyond measure: their dedication to meaningful, sustained improvement in education has left an indelible imprint on me and my work; it gives me huge pride to say that I was supervised by them. Rob and Steve, I dedicate this thesis to you.

I am also indebted to my friend, business partner and long-suffering proofreader, Jack Deverson. I have frustrated you to the point of implosion, but you have helped me complete this thesis, so I dedicate it to you, too.

Also, to everyone at the Centre for Evaluation and Monitoring at Durham University who welcomed me, supported me, guided me and encouraged me to 'get on with it', I thank you. Likewise, I offer my sincere thanks to the team at the LEAD Graduate School and Research Network at the Eberhard Karls Universität, Tübingen; their support – especially that of Professor Doctor Augustin Kelava – was excellent, timely, and uplifting.

Finally, without the significant and unwavering support of my family, I would never have embarked upon on a doctorate, let alone completed one. The ways you have helped me are innumerable and would take me well over the word count, so know this: John, Sue and Andy, little would have happened without your support, encouragement and gentle reminders that it would be a good idea to get it finished. To you, I dedicate this thesis.

Chapter 1: Background to the study

1.1 Rationale

The rationale and design of this study were prompted in part by ten years' experience as a classroom teacher of English in UK secondary schools. During my years as a Newly Qualified Teacher (NQT), classroom teacher, Head of Department and Director of Studies, I became increasingly aware of the need for high-quality feedback in the areas of education with which I had intimate contact: students' learning and teachers' continuing professional development (CPD). On a day-to-day basis, as teachers deal with pastoral issues, safeguarding matters, the teaching of British values, friendship disputes, the emotional consequences of medical conditions, academic progress monitoring and planning sequences of effective lessons - of which they can be proud as professionals, which promote learning, and with which an inspector would be satisfied - critical evaluation of the methodological quality of the techniques at their disposal is not a high priority, nor one to which a great deal of time is dedicated. Ending my time as a secondary school teacher and entering academia full-time, I found the experience of recalibrating my view of the landscapes of education research and teaching a difficult but enlivening one.

Interestingly, my departure from full-time employment as a teacher appears to have coincided with a number of other events which, currently, are proving instrumental to some kind of shift in the view of research held collectively among teachers[1]. The creation

[1] There should be no confusion at this point: this is an example of correlation rather than causation.

of ResearchED[2] - a grassroots movement led by Tom Bennett - offers a view of this shift in thinking, as demonstrated by the quotation below, taken directly from www.researched2013.co.uk:

> This conference seeks to bring teacher and researcher together. To draw out the best of the practitioner's experience and the analyst's theory, and produce something better: evidence based education that actually fits the classroom, rather than attempts to make the classroom fit it.

The last line of this statement is, if any line could be, a fine encapsulation of the motivation for my research; the design and implementation of the peer-coached feedback consultation is my version of the statement embodied.

1.2 The Centre for Evaluation and Monitoring, Durham University

During much of the time I spent on my doctoral work, I was fortunate to be associated with the Centre for Evaluation and Monitoring (CEM) at Durham University. CEM is the largest educational research unit in a UK university and has been in existence since 1983; it is the organisation in which I was based while conducting this study. It is a non-profit organisation whose functions are manifold; they can best be summarised as those of an external evaluator for educational interventions, and a provider of curriculum-based and ability tests to primary and secondary schools around the world.

CEM's mission is clear. It is one both I and this thesis share: the basis of work conducted there is to improve teaching and learning for young people[3]. As a research organisation,

[2] http://www.researched2013.co.uk/about/about-researched-2013/ accessed 9/5/14. researchED came to life following a conversation between Bennett, Sam Freedman (former Advisor to Michael Gove) and the hospital doctor and author, Ben Goldacre. It is a fine example of stakeholders in an education system leading the debate on a topic of relevance and importance to the system.
[3] http://www.cem.org/mission-statement accessed 5/11/13.

the Centre has an interest in developing reliable, valid instruments which measure important features in the teaching and learning process, such as student progress (as measured by value-added) and a student's ability to learn (as measured by various tests of developed ability, such as the Advanced Level Information System (ALIS), the tool used in this project to provide baseline and post-test data, detail about which comes later in this chapter). This mission provided the perfect foundation for a study concerned with evaluating teacher effectiveness by means of student perceptions of their teachers' performance.

The completion of a CEM contract to provide testing services to a school or group of schools includes a clause which indicates the institution's (or institutions') consent to allow their anonymised student test data to be used for research purposes. This facet of the agreement was of significant benefit to this project as it gave an entry point to schools already permitting such uses of their data, but also provided ground that could perhaps be considered more fertile for research purposes: schools using CEM data may well, on average, be more disposed to engaging in research activities more broadly than their non-CEM counterparts. Greater discussion of the potential bias lent by such potentially systematic differences is given in chapter 4.

1.3 Advanced Level Information System (ALIS)

The Advanced Level Information System[4] (ALIS) is an online tool used in post-16 education for diagnostic purposes[5] as a performance indicator. A voluntary assessment for schools, it

[4] http://www.cem.org/alis/introduction accessed 28/7/14.
[5] ALIS provides measurements of a student's abilities in vocabulary, mathematics and non-verbal skills; these scores are standardised.

provides predictions of Advanced Subsidiary Level (AS Level) and Advanced Level (A Level) outcomes (amongst others) and value-added analyses for those students who take it in each of the subjects they study[6]; it has a long history (it was launched in 1983) of collecting high-quality, valid and reliable data. ALIS reports the correlation between average GCSE score and outcomes at AS Level and A Level. In 2013 (the year in which this PhD study was conducted), the average correlation across all subjects for which ALIS predicted outcomes was 0.58, with correlations ranging from 0.36 (ICT Applied, Double Award) to 0.73 (DT Systems and Control).

The ALIS baseline assessment is taken by individual students on a computer; it takes around one hour to complete. The baseline feedback (comprising information about a student's vocabulary ability, mathematical ability and their non-verbal ability) sent to schools is often used for target-setting and progress-monitoring purposes. It should be noted, however, that while this baseline assessment provides its own prediction to AS Level and A' Level outcomes, it is the prediction from a student's average GCSE score (as mentioned above) which is used in the analysis of value-added. For this reason, school-level average GCSE scores are used in this project as the baseline data for randomisation purposes, and the predictions from average GCSE scores to AS Level and A Level outcomes used in the analysis. In the analysis, the standardised value-added scores for each student in the dataset are used as a measure of student progress; the assumption made is that this measurement offers a more accurate representation of learning than would raw student outcomes (grades achieved).

[6] Such students are, most often, aged 17 – 18 years old.

The analysis in this study (presented in Chapter 8) uses a multilevel model in a confirmatory factor analysis with cross loadings to assess both the degree to which ratings on the SecondarySEEQ correlate with standardised value-added scores, and the impact of the peer-coached feedback intervention on the principal outcome measure (AS Level and A Level standardised value-added scores).

1.4 Value-added

In the context of ALIS data, value-added data are those which express measurements of student progress over time, when compared to similar pupils[7]. Essentially, the predictive data for each subject taken by a student (based on that student's average GCSE score) are compared with the achieved outcome; the residual is known as value-added, and can be positive or negative (because it refers to an average). For the purpose of this study, value-added information provides an excellent way of controlling for baseline characteristics of each student, something that is important to the strength of the claims made following the analysis.

It should be noted that value-added is conceptualised in this study as a measure of student progress over time; it is not in itself a measure of teaching quality or school quality. Examples of ways in which value-added data are used in other education systems give great cause for concern; in the Los Angeles Unified School District (LAUSD), for instance, value-added data are published by the Los Angeles Times in an online, searchable database[8]. This database allows anyone visiting the website to view value-added data at the school-level

[7] These similarities are age and stage of education.
[8] http://projects.latimes.com/value-added/ accessed 29/04/15.

and at the teacher level[9] (with both schools and teachers being identified by name). This application of value-added data is presented here as a point of contrast: this thesis takes issue with the ascription of student progress so directly to individual teachers, a topic that will be discussed in greater depth in subsequent chapters.

1.5 Feedback

In designing an intervention and evaluating it, the general area of greatest interest to me both professionally and conceptually was feedback. From a pragmatic point of view, high-quality feedback seemed to be a plausible means by which teachers' professional learning could be encouraged; early reading for this project highlighted that much has been written about the power of feedback given to both students and teachers, (Dominick, Reilly, & McGourty, 1997; Gregory, Beck, & Carr, 2011; Hattie & Timperley, 2007; L'Hommedieu, Menges, & Brinko, 1990). Feedback, both in its conceptual form and in the practical application of it, is explored fully in Chapter 3.

Garnering feedback from students has a history which does not always make for encouraging reading (as is explored later). Nonetheless, I was (and still am) convinced of the great opportunities to gain insights into teaching by asking students to share their opinions on the teaching they receive. It was from this standpoint that Student Evaluation of Teaching (SET) became an area of interest and, apparently, a suitable vehicle with which to capture the views of students. The literature on the subject is explored fully in Chapter 3; initial reviews, however, clearly indicated that, in higher education contexts, SETs are considered (depending on the ratings instrument used) reliable, valid and useful in teaching

[9] In 2010, there was great concern raised in Los Angeles that a value-added rating of 'less effective than average' ascribed to one teacher (Rigoberto Ruelas) had contributed to his suicide.

(Marsh, 1982). One aspect that I considered crucial seemed, in many instances, to be missing, however: consultative feedback to teachers for the purpose of improving their practice. Synthesising the literature on SETs, feedback, goal-setting and intention implementation generated the structure and content of the intervention under evaluation in this study.

Out of this practical interest and theoretical background came the research questions the study set out to answer:

1. Is the SecondarySEEQ instrument used in this PhD study a valid means with which to capture measurements of teaching effectiveness?
2. Does the presence of a termly, one-hour, peer-coached teacher consultation meeting based on SecondarySEEQ student evaluation of teaching data (treatment) create a larger effect on attainment than feedback alone (control)?

The project set out to validate an existing SET instrument (the Student Evaluation of Educational Quality instrument (Marsh, 1982)) in a novel context (secondary school classes studying for A Level examinations[10]) and evaluate the impact on Advanced Subsidiary Level (AS Level) and Advanced Level (A Level) attainment of a feedback consultation for teachers.

1.6 Measuring the effectiveness of teaching: What is meant by 'effective'?

Any discussion of effectiveness of teaching must, as Biesta (2007) points out, address the question 'Effective for what?'. Being precise about the intended outcomes of teaching is

[10] A (Advanced) Levels are academic qualifications awarded to students who have satisfied the requirements set down by examination bodies. Students taking A Levels are, generally, aged 17-18 years old. At the time the study was conducted, students in England were required by law to enrol in "some form of education or training" until the age of 16; since this time, the law has changed and, as of September 2013, the age until which young people must be enrolled in education or training is 18 (https://www.gov.uk/know-when-you-can-leave-school accessed 9/5/14).

the only way truly to ascertain what constitutes effective teaching; this is a highly problematic area for this study.

One of the central challenges of this PhD thesis comes in the form of a lack of agreement in the literature about a definition of 'effective teaching'. This is a theme which is picked up later in the thesis, but it should be stated for the purposes of foregrounding this study that such a unified, agreed definition is absent from the debate. One of the most recent reviews of the evidence on teaching effectiveness (Coe, Aloisi, Higgins, & Elliot Major, 2014) indicates that more work is required to determine a better definition (a seam which has run through this PhD project from its inception), but suggests that there may be 'quick wins' which could swiftly augment understanding of the area.

1.7 The Student Evaluation of Educational Quality (SEEQ) instrument

The Student Evaluation of Educational Quality (SEEQ) instrument (Marsh, 1982) has been demonstrated to provide reliable, stable evaluation data which are chiefly a product of the teacher, rather than of the course they teach (Marsh, 2007c). The SEEQ has most commonly been used in the United States and Canada (where SETs are most commonly collected (Centra, 2003)), though its use in Australia and other countries is also significant (Marsh & Roche, 1997). In light of these factors, and the support I received from Professor Herb Marsh, the SEEQ became the ideal candidate instrument for this trial. The specific aspects of teaching that the SEEQ purports to measure are detailed in Chapter 2.

1.8 Continuing Professional Development (CPD) for teachers

As a teacher and manager in schools, my experience of the quality of teachers' continuing professional development (CPD) was varied at best. Moreover, I was often concerned by

the significant sums of money being paid to training providers for their services[11]; equally, the quality of the information imparted seemed questionable on numerous occasions, as a one-size-fits-all model was applied to teachers from a variety of different schools. Even CPD events provided by examination boards came to the public's attention for all the wrong reasons in December 2011 (BBC, 2011), though perhaps this shouldn't come as too great a surprise, given the pressures on teachers to enable their students to achieve the highest grades possible, and the competitive market in which examination boards operate.

Many events purporting to be focused on the development of teachers' skills served more of a box-ticking function, setting up arguments and potential strategies to improve aspects of teaching, but lacking long-term follow-up and sustained support. Too often, my own experience of CPD events tallied with those of colleagues: lip-service was paid to the event's mission and content, but serious positive change was often lacking as an end product. A one-day CPD event, taking a teacher out of school for a day, provides some thinking time for the teacher, but this is ineffective if time and resources are not allocated immediately after the training event for reflection and action-planning. Overarching all of these criticisms is the complaint that CPD rarely addresses the underlying pedagogical needs of the approaches it seeks to promote: it is far easier to *say* that a specific strategy or approach is being applied than to *understand* the pedagogy supporting it.

There is evidence in the literature that this kind of experience is not unusual amongst teachers more broadly (Vescio, Ross, & Adams, 2008; Wei, Darling-Hammond, Andree,

[11] The commercial aspect of teacher training is an interesting area in itself. Many providers claim their products and services have an 'impact' on student and teacher outcomes, often with little or no evidence of such impact. In education systems which apply pressure to teachers and school leaders to demonstrate improvement, the demand for professional development courses and tools is met with a ready supply, and often at relatively high financial cost.

Richardson, & Orphanos, 2009; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007), pointing to a risible irony for the teaching profession: a lot of CPD events (lessons for teachers) would not compare favourably with their student-focused counterparts in schools (lessons for students). Were we to look critically at the quality of much of the professional development delivered to teachers (and the associated reflection and action that ensues) through the same lens with which we view lessons, the scene would be a disappointing one.

Such assertions about the effectiveness of CPD for teachers have prompted interesting reports and recommendations for policy and practice (Adey, Hewitt, Hewitt, & Landau, 2004; Borko, 2004; Cordingley & Bell, 2012; Darling-Hammond & McLaughlin, 1995; Little, 1993; S. Wilson & Berne, 1999). Most recently, Cordingley (2012) noted the following as facets of CPD that appeared likely to be effective both for teachers' learning as well as students' benefit, though even these are not the most robust:

- Collaboration, particularly coaching or mentoring;
- Support from specialists in the area;
- Focus on student aspirations;
- Sustained over time;
- Exploration of new ideas through practice;
- The presence of networks of teachers;
- Structured dialogue.

The 'Developing Great Teaching' report commissioned by the Teacher Development Trust in the UK (S. E. Higgins, Cordingley, Greany, & Coe, 2015) draws on the best available evidence on professional development to make practical recommendations for both schools and training providers, many of which echo the Cordingley (2012) report. Such

guidance is important and useful, but only regulation and market forces will, ultimately, alter the way in which teacher professional development is engaged with and delivered. Mandates from regulatory bodies (or incentives provided by them) would be needed at the system level for schools to change the structures and processes they have in place for teachers' professional learning, and to align it with the best available evidence. Similarly, the market needs to demand a different product from training providers if there is to be any change from the status quo (chiefly one-day courses with little or no follow-up).

While support exists for the inclusion of the facets of teacher development interventions suggested by Cordingley (2012) and Higgins et al. (2015), there are still gaps in the evidence, as made clear by Timperley, Wilson, Barrar, and Fung (2007). The authors identify a paucity of the theoretical base pertaining to the professional learning of teachers, weak empirical evidence regarding the composition and delivery of effective CPD by providers, and a lack of research on the mentoring of early-career teachers. Thus, the reviews conducted by those authors previously mentioned offer a step in the right direction, but it must be recognised that until experimental designs are used in the evaluation of teacher professional learning interventions, the field will remain lacking in the required evidence.

In the UK, the foundation of the Teacher Development Trust[12] (TDT) and its associated groups (the National Teacher Enquiry Network[13] (NTEN) and the Good CPD Guide[14] have brought to the fore the debate regarding the quality and composition of effective CPD. In a similar fashion to researchED, TDT and NTEN engage with researchers, policy-makers and practitioners in the shared pursuit of a better understanding of what effective CPD is and

[12] http://tdtrust.org/ accessed 28/7/14
[13] http://tdtrust.org/nten/home/ accessed 28/7/14
[14] http://www.goodcpdguide.com/ accessed 28/7/14

how it might be delivered. One of the key facets of their work is the support of evaluations of CPD interventions for teachers, particularly at the intersection between CPD providers and teachers. Acknowledging the guidance offered by the best available research on CPD for teachers, providers are actively encouraged to put in place processes for evaluating the impact of their work.

The notion of CPD as a significant factor in teacher learning - and, subsequently, student performance - is an interesting one when one considers the commercial element that is constantly at play. I detail the literature on CPD in greater detail in Chapter 3, but wish here to set out a dichotomy that presents a challenge.

By engaging the services of CPD providers external to their school, teachers and school leaders seek a different perspective, perhaps from trainers or consultants with greater experience and understanding of what others have done and currently do. Thereby, for all but the internal, home-grown CPD that goes on, a fee is levied and associated costs (travel, subsistence) charged. The question, therefore, that is uppermost is simply: is what is being paid for effective? Ultimately, the only way truly to generate and answer to this question is to conduct an evaluation and this, I contend, is a difficult concept for intervention developers and training providers. Why do I say that it is 'difficult', then? A well-designed and implemented evaluation will report its findings accurately, and these may well demonstrate that the impact of the intervention is zero, or even negative. From a commercial standpoint, I would argue that having a robust evaluation of an intervention conducted brings with it the potential of (in equal measure, if we remain in a state of equipoise) evidence with which no advertising campaign could ever compete ('our intervention has a positive effect and adds six months' progress, on average'), and evidence

with which no advertising campaign could ever remedy ('our intervention doesn't work, but we'd like you to buy it anyway').

It is in light of the literature on teacher learning and development that the intervention at the heart of this project was designed; this is detailed in Chapter 4. Moreover, much of the conceptual framework for the intervention was drawn from Thompson and Wiliam's (2008) work on the 'tight but loose' model. This model, also detailed in Chapter 4, incorporates a concentration on and adherence to the key tenets of an intervention while allowing for the many accommodations that must be made in classrooms. To expect that every teacher involved in this project would diligently follow a rigid and precise set of intervention implementation guidelines would be naïve; the tight but loose approach offers a pragmatic model that allows for control and variation simultaneously.

1.9 The evaluation imperative

In the recent history of noteworthy educational interventions, Brain Gym®[15] stands out as one which has garnered much attention, but which also provides a salient lesson: not evaluating the effect of an intervention (when techniques and skills are available to do so) is tantamount to systematic professional misconduct on the part of those policy-makers, intervention developers and advertisers who propound the (untested, therefore unknown) 'benefits' of a product. As Willingham (2012) points out, readers or viewers often respond to the use of scientific terminology in relation to products or processes designed to have a positive impact (medicines, educational interventions, therapies) with uncritical acceptance borne of reverence and the fear of looking foolish. The jargon-based heuristics

[15] http://www.braingym.org/ accessed 9/5/14

which are present in all walks of life (from shampoo adverts to education policy-making) become a shared (but meaningless) language; if one doesn't share it and at least pretend to understand it, the perception of being ill-informed can grow. So, when an intervention such as Brain Gym® is touted as having profound positive effects on intellectual and physical development, it becomes difficult for what Kahneman (2011) calls 'the System 2 process' to cancel out the noise created by an excited System 1. Consequently, even an intervention based on theoretical foundations long ago demonstrated as invalid and lacking sound empirical research to validate its claims can become popular in more than 80 countries (Spaulding, Mostert, & Beam, 2010).

1.10 Intention implementation

The methodological weaknesses in studies of teacher professional learning are compounded by the particular scarcity of research into the impact of interventions based on the theory of planned behaviour (Ajzen, 1991), and built upon by the work of Gollwitzer (1999) into the effects of implementation intentions. Such theories speak to the problematic area addressed by this PhD study: difficulties exist in 'translating' goals into actions. Gollwitzer (1993) offered a summary of the way in which implementation intentions function by describing the "passing of the control of one's behaviour on to the environment" (p.173), intimating a deliberate disassociation of the self from the actions and behaviour of the self. This notion is one which runs through the core of the intervention under evaluation in this study.

1.11 Contexts and relevance of the study

Given the growing interest in evidence-based teaching in the UK[16], there is a certain prescience to a study that attempts to validate a student evaluation of teaching instrument for use in UK secondary schools, and to trial an intervention focused on teacher professional development which uses evidence of student perceptions as the medium for the improvement of student attainment. There are three key contexts which it is important to consider in order to understand the broader background to this study: the political, the economic and the pedagogical.

1.12 The political context

The landscape of formal education across the primary, secondary and tertiary phases has undergone dramatic reshaping over the last 20 years in the United Kingdom[17]: the increased number and type of qualifications brought in by the Blair Labour government, the widening of access to higher education, shifting school inspection priorities and schedules, to name a few. In the UK, these changes have been driven by political decisions which have, for the most part, been taken for economic reasons (more on this in the next section), or those designed to serve either a societal function (such as greater equality and diversity) or political aim (re-election, for instance). High-quality, evidence-informed educational rationales have not often been the chief actors in the decision-making process concerned with teaching and learning in the UK.

[16] On 9th December 2014, the Department for Education announced its support for a College of Teaching, a key intended feature of which would be to place evidence at the heart of its work. https://www.gov.uk/government/news/new-measures-to-support-a-world-class-teaching-profession accessed 10/12/14. Much of the Department's focus currently is on the embedding of high-quality evidence into the practices of the organisation across the piece, as evidenced by such initiatives as the Policy Fellowship (https://quarterly.blog.gov.uk/2015/01/27/raising-the-bar/ accessed 15/05/15) to which I was appointed with the remit of helping the Department use research evidence in policy-making processes and decision.

[17] Such changes are not exclusive to the UK, but are of most relevance here.

The most recent significant political shift came about with the advent of the coalition government, formed of the Conservative and Liberal Democrat parties on 11th May 2010. Historically, coalition governments in the UK have some interesting features: notably, they often form during or soon after some kind of significant global event, the impacts of which are severe and systemic (the Asquith coalition of World War 1; the Lloyd George Ministry of 1916 - 1922; the Churchill war ministry of World War 2). It is the very act of coming together to share power with a party whose political bent is different that is perceived to lend the legitimacy required to govern in a state of crisis or perceived crisis. The event which, almost inevitably, led to the formation of the current coalition was the economic decline that began in December 2007 but which made its first significant nosedive in September 2008.

It is possible to argue that the kind of political division created by the lack of a majority vote being awarded has, directly or indirectly, had an impact on education policy in the UK. With a lack of consensus among voters regarding which party should be charged with the running of the state, the potential for an adjunct lack of unity in education policy-making may exist. Indeed, it is interesting to note how an initiative such as the Pupil Premium (Chowdry, Greaves, & Sibieta, 2010) - an idea with the education of the poorest children in England at its core - has become a party political tool for the Liberal Democrats, with the Rt. Hon. Nick Clegg MP championing it and securing its inclusion in the coalition government's plans. My own experience of working in government (as a Civil Servant) clarified the importance of taking into consideration the political (party political, for the most part) deliberations that go on over association with one policy initiative or another. This aspect of policy development and implementation should not be ignored; it is the ministerial preference,

the party line or the media hobby-horse which will often sway the ultimate decision[18]. While there is a growing understanding and use of evidence in the education policy-making process in the UK, it remains a process which is unhelpfully coloured by politics and parochial interest: anyone whose job relied on an electoral majority would keep this in mind when making decisions which could adversely affect their public appeal.

1.13 The economic context

One of the interesting consequences of a recession is the behavioural reaction witnessed among leaders of organisations of all types. Devoid of seemingly unlimited lines of credit or vast assets to support every new idea, greater scrutiny is given to proposals and the economic imperative begins to trump others such as the social, or the educational. Austerity becomes a buzz word. It is broadly in such an environment that this study was designed and, as such, attention was given to the creation of a scalable, cost-effective intervention that had pragmatism as a driving force. I felt strongly from the project's inception that to develop, trial and evaluate an intervention based on a student evaluation teaching instrument and process that would, ultimately, be unfeasible at scale due to cost or other impediments would be an exercise in academic vanity. In light of this, I drew on the work collected together by the American education measurement and research organisation, ETS (Wylie, 2008), adopting a 'tight but loose' model. Just as businesses and other organisations must reconsider how limited resources are spent in straitened times, so developers of educational interventions should, I believe, have one eye firmly on the

[18] I have seen the results of Ministerial reviews of education policy proposals. One consequence of the redaction of sections of a proposal is that what began as a unified policy becomes a fragmented assembly of parts.

bottom line and the pragmatism of their proposed work. The development of the intervention designed for this PhD study is detailed in Chapter 6.

1.14 The pedagogical context

In an attempt to reflect the societal importance of democracy (Flutter & Rudduck, 2004), pupil voice is an important - if contentious - real-world application of the UN's Convention on the Rights of the Child (CRC) (UN, 1989). Its importance is grounded in the notion that children should have their voices heard in a meaningful, non-patronising manner; the purpose of such an action should be for the betterment of the contexts in which children find themselves.

The development of pupil voice initiatives in UK schools can be traced back to the work of Giroux (1986) and through, importantly, to the United Nations' Convention on the Rights of the Child (UN, 1989). This treaty formalised much of the debate that had grown in academic discussions and publications, placing a legal bent on the exposition of students' opinions and thoughts.

Noyes (2005) speaks of 'pupil voice' as something related to engaging pupils with the purpose of improving the environment in which they 'live and learn'. It may be understood - in language used by the UK Government - as a data-gathering activity designed to increase the sway held by students in those provisions made for their education by ensuring that pupils' views and opinions are taken into account when schools make significant decisions[19].

[19] http://www.education.gov.uk/vocabularies/educationtermsandtags/7069 accessed 5/11/13

Pupil voice data for the aforementioned purpose might be considered 'low-stakes' in essence: data whose function is to inform and generate discussion, rather than to provide the basis for a high-stakes decision, such as the recruitment of a teacher at a school. Yet, as Campbell's Law states:

> "The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures, and the more apt it will be to distort and corrupt the social processes it is intended to monitor." (D. Campbell, 1975)

So it is unsurprising to find examples of poorly-managed pupil voice exercises, such as one in a UK school which was described publicly in 2010 by one union leader as a "demeaning, embarrassing and humiliating practice"[20] (students were asked to take part in the interview process for a teacher; one student on the interview panel labelled the teacher in question 'Humpty Dumpty'[21]). Clearly, poor execution of an exercise whose focus is on encouraging students to share their opinions on matters related to their education has the potential to be highly problematic, particularly in the UK context in the 21st century[22]. The heart of this study is in engaging students in sharing their opinions - not simply of schooling, but specifically about features of individual teachers' practice. As will be seen in subsequent chapters, the design of the peer-coached consultation feedback intervention created for this study acknowledged the potential for the kind of problem highlighted by Campbell's Law, and took steps to mitigate the potential for such manipulations occurring.

[20] http://news.bbc.co.uk/2/hi/uk_news/education/8599485.stm accessed 5/11/13

[21] One rationale for the inclusion of students on the interview panel was to raise their self-esteem and motivation. Clearly, here is seen Campbell's Law in action.

[22] It is interesting to compare the context of a UK pupil voice exercise with, for instance, a Thai context. Often, Thai children are reluctant to share anything other than a positive or simply benign view of their school experience. Culturally, to otherwise would be inappropriate.

1.15 Summary

The motivation for this study, its purpose, design and implementation are all grounded in the practical context of my ten years spent teaching. It should not be surprising, therefore, that a pragmatic stance is taken throughout, particularly with reference to the implementation of the intervention and its potential future scalability. An underlying philosophical position taken in the construction of this study is that research in education should, as far as possible, be dedicated first and foremost to answering fundamental questions about the impact of what we do in our schools, while acknowledging the debate about the 'non-causal' nature of many aspects of education (Biesta, 2007). While this somewhat 'functionalist' approach to research does not often resonate with the classical view of the researcher setting the agenda and following his/her own interests, it is one which sees the needs of the system - and the society to which the system belongs - as having greater importance than the individual researcher.

Consequently, the following chapters will describe a low-cost (both in time and money terms), consultative feedback intervention which is both replicable and wholly scalable to the system level, and the evaluation designed to measure its impact. Thereafter, analysis of the impact is presented, along with data to support the validity of the instrument in its novel context.

Before addressing the specific details of intervention and evaluation design, it is important to underpin this thesis' argumentation by reviewing and assessing the literature relevant to the study, a task undertaken in Chapter 2.

Chapter 2: A review of the literature

Acknowledging that - theoretically, at least - Student Evaluation of Teaching (SET) may have applications in secondary school contexts in the United Kingdom, it is important to note at the beginning of this literature review that the evidence of its application here in the UK is very limited. This dearth of evidence speaks very much to one of the broad themes that this PhD study touches on: literature reviews of even the most robust and highest quality research can only indicate the degree to which some*thing* worked some*where*. The assumption that, because the results of multiple randomised controlled trials of a particular intervention were positive, the intervention will work in a new context is one challenged both by this thesis and, more thoroughly by Cartwright and Hardie (2012). Moreover, only the literature which has been published can be reviewed: publication bias must be acknowledged as having a potentially adverse effect on any literature review, this one included (Dickersin, 1990).

A review of the literature associated with this PhD study should make reference to the substantive and theoretical evidence found in the literature which relates to the key elements of the study: validity; Student Evaluation of Teaching (SET); feedback; teacher professional development; goal-setting.

In assessing the contributions of the evidence relating to interventions associated with the substantive areas mentioned above, the most robust approach to take would be to conduct systematic reviews, in line with the guidance offered by The Cochrane Collaboration (J. Higgins & Green, 2008), reducing the amount of bias at play (Oxman & Guyatt, 1993). No

systematic review of the literature was carried out for this PhD project, and this is here noted as a weakness of the review presented in this chapter. Acknowledging this potential source of bias and consequent weakness of the survey, it is hoped that the argument developed here demonstrates a logical structure and appropriate range of views as to be considered worthwhile and accurate.

This chapter reviews the evidence found in the literature in the following areas which were identified as critical to this study:

- What is effective teaching?
- Validity;
- Feedback;
- Student Evaluation of Teaching (SET);
- Goal-setting;
- Continuing Professional Development (CPD).

In doing so, the context in which this PhD study sits is described and the rationale for its existence illustrated. An argument is developed out of an alliance between the existing research and the current study, which seeks to show a logical association of that which has gone before to contemporary work, but which also seeks to highlight areas in the literature which cause concern and prompt further work.

2.1 What is effective teaching?

There is no consensus over a definition of effective teaching (Coe et al., 2014). The discussion of defining a set of desirable aspects of teacher performance (what we actually want them to be good at) is one well-presented by Wiggins (1989): while it is a critical aspect of the construction of a performance assessment, an agreed stance on the constituent parts or multiple dimensions (assuming acceptance of the notion that teaching is an activity which is composed of a multiplicity of behaviours and actions) of effective teaching is hard to achieve. The array of definitions of teaching quality is presented neatly by Cochran-Smith (2003) and Fenstermacher and Richardson (2007). The latter indicate that 'good teaching' is one of a four-part constitution required for successful student learning to occur (the others being: the willingness to learn and effort applied by the student; a supportive teaching and learning environment, not simply in school, but at home and within the student's community; the opportunity to teach and to learn - i.e., that resources and time are devoted to learning), and that all aspects of the constitution must be addressed fully. Somewhat provocatively, the authors suggest that the oft-used phrase 'quality teaching' is most likely a product of the successful enactment of all four parts of their suggested constitution, not simply a set of behaviours and actions performed by a teacher. I tend toward the notion that effective teaching is composed of learned, improvable behaviours, and that there are certain observable latent variables which may be used as indicators of effective teaching. It is, however, useful to note that this dissertation does not set out to propose a new definition of effective teaching; rather its focus is on the validation of an existing instrument in a novel context, and the evaluation of an intervention.

In light of this notion that the quality of teaching depends on more than simply the activities of the teacher, the validation of a multidimensional instrument which seeks students' perceptions about the quality of the teaching environment (in respect of a context in which students feel able to learn), the resources offered to support learning, as well as the actions and behaviours of their teacher becomes interesting. While the SecondarySEEQ does not explicitly probe students' views on their own effort levels, nor on the supportive nature of their home and broader communities, it does engage with the notion of the broad composition of effective teaching suggested by Fenstermacher and Richardson (2007). Most importantly for this validation process, the authors separate 'good' teaching from 'successful' teaching, stating that the former is teaching which adheres to professional, moral and pedagogical standards for the delivery of content, whereas the latter is seen as teaching which, as a causal mechanism, leads to student learning. It is this latter definition which this thesis holds at its core.

2.2 Validity

Wherever one looks in the domains of educational and psychological measurement, the term 'validity' is used with high frequency and importance; it is considered in the APA Standards as "the most fundamental consideration in developing and evaluating tests" (AERA, APA, & NCME, 1999). The accuracy of its contemporary use is, however, an arguable point, as has been noted clearly by several prominent researchers (Arthur, Waring, Coe, & Hedges, 2012; Newton, 2013; Newton & Shaw, 2014). Just as the history of semantics is strewn with words which have changed in meaning over time (for instance 'awful', 'myriad' and 'senile'), creating misunderstandings as a result, measurement research finds itself grappling with the multiplicity of meanings applied to a term very much at its heart: validity. In drawing conclusions about the validity of the instrument used in this PhD study, it is of crucial importance to set out clearly what is meant here by the word, and what is not. Without a clearly-explained working definition of validity, any claims made about it in this thesis will be hollow.

Validity is a term coined by Kelley (1927); it indicated the degree to which an instrument measured that which it purported to measure, and was divided into three types: content validity, criterion validity, and construct validity. This definition was elaborated upon by Messick (1989, 1990, 1994) to incorporate six distinct aspects of construct validity: content; substantive; structural; generalisability; external; and consequential.

The importance of validity as a concept and indicator became apparent with its appearance in the AERA et al. (1999) *Standards*: "Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests". This augmentation of the common classic definition - that validity is 'the degree to which a test

measures that which it is intended to measure' - places interpretation (or claim) as a central feature of the concept. More broadly, the literature suggests that validity is a property of something (though *of what* is where the argument becomes fractured), and that validity is something concerned with strength (though *of what* is, again, the problematic area) (Newton, 2013). For the purposes of this study, the notion propounded by Borsboom, Mellenbergh, and van Heerden (2004) that there is in a measurement process an attribute which acts in a causal manner to determine the values of the measurement appears useful, though it seems somewhat reductive; in this context, validity is viewed as a concept which, with echoes of both Messick (1989) and Newton and Shaw (2014), seeks to affirm measurement by answering directly the questions: a) does the attribute 'effective teaching' (defined as teaching which has an effect on students' learning, as measured by attainment measures) exist? and b) are variations in measurements caused by the attribute? While a more clearly-stated definition of validity is offered in this way, it must also be recognised that it adds yet another stratum to an already tightly-packed and multi-layered understanding(s) of the term.

So where does this PhD study stand in its understanding and use of the term validity? Firstly, I recognise that, conceptually, validity is a term which cannot be ignored, despite the fact that its multiplicity of meanings renders it more useful for obfuscation than it does for clarification. Recognising its importance, however, does not automatically mean that it needs to be considered a unified concept (which, it is contended, it is not); preferred in this thesis is a model proposed by Newton and Shaw (2014). This 'neo-Messickian' (ibid) model suggests avoiding a discussion of the global term 'validity', focusing instead on the *quality* of the measurement (its capacity to measure the desired attribute (thereby incorporating

the definition proposed by Kelley (1927)), and its reliability as a tool for making more accurate decisions, and the *value* of the measurement (in terms of economic costs, decision-making costs, as well as intended and unintended consequences) to policy[23]. By adopting this approach as part of the conceptual framework for this study, I seek to separate the key components of the SET instrument (its quality and its usefulness) and to assess each independently. This is taken up further in Chapter 7.

2.3 Students' Evaluation of Teaching

> "Probably, students' evaluations of teaching effectiveness are the most thoroughly studied of all forms of personnel evaluation, and one of the best in terms of being supported by empirical research." (Marsh, 1987)

As a contextual setting for a review of the literature on Students' Evaluation of Teaching (SETs), this sentence (p.369) provides a firm foundation. While there is a wealth of information available, however, the field is one in which disagreement over key issues of reliability, validity and usefulness of SETs continues unabated. With this in mind, this review of the available evidence seeks to present an historical overview of SETs, an engagement in the debate surrounding validity, reliability and usefulness of SETs, and a description of how this PhD study builds upon and contributes to the existing evidence. Again, it is important to keep in mind that I here review the available literature, conscious that publication bias may affect the conclusions I draw.

[23] The term 'policy' refers here to decisions made at any level (individual, organisation, system) which are based on the claims drawn from interpretations of the measurements. Discussion of validity as a concept concerned with the claims made on their basis is well-made in Arthur et al. (2012).

2.4 The background to SET research

While there is a history of SETs dating back to the early twentieth century[24], things began to gain greater traction in the 1970s and 1980s. Nonetheless, it is worth citing H. H. Remmers' work (Remmers & Brandenburg, 1927) at Purdue University as crucial to the development of SETs. While Remmers initially said that the student ratings he collected were not to be construed as measurements of teaching effectiveness and was, initially, concerned to ensure that more research was undertaken before the validity of the data for use in promotion and similar decisions was considered appropriate. Subsequent revisions of his position suggest growing confidence in SETs. After several decades of work, Remmers concluded that SETs were reliable and valid (Remmers, Martin, & Elliott, 1949) and that their function as a measure of effective teaching must not be ignored (Drucker & Remmers, 1950).

2.5 From Remmers to Marsh

Remmers' work must be mentioned here because its conclusions about the reliability, validity and usefulness of SETs have been widely disputed (Marsh, 1987). Despite the objections of some researchers, Remmers' conclusions have very much been echoed by the work of Marsh (for example: Marsh (1982, 2007b, 2007c)), whose research is cited extensively in this thesis. It should be stated clearly, then, that while Marsh may view Remmers as 'The Father of Student Evaluation Research' (Marsh, 1987), Marsh himself has become its 'favoured son'; this thesis speaks very much from a stance of scientific objectivity, but acknowledges the work of Remmers and Marsh in respect of the reliability,

[24] Several North American universities such as Harvard, Purdue and the University of Texas were among the small number in the 1920s evaluating teaching by means of student feedback.

validity and usefulness of SETs; this is a point that will be developed further in Chapter 7, in which the data collected from the main trial are analysed. That said, issue is taken with what Apodaca and Grad (2005) consider an over-reliance on statistics at the expense of substantive findings: without such detailed and careful statistical analyses of SET data, there would be little foundation on which to base substantive research. While one could enter into a 'chicken and egg' argument, it is a redundant activity and one which can be concluded always with agreement that a broad-based substantive, empirical and methodologically-appropriate approach is needed to understand SETs fully. This is the approach I adopt in subsequent chapters, particularly in Chapters 7 and 8 in which analyses and discussions are presented.

Having stated a personal view (with firm foundations in high-quality evidence) that SETs in higher education contexts can be reliable and valid measures of teaching quality, it is interesting to consider the assertion made by Aleamoni (1981) that SETs have - across the many reviews of them that appeared in education research publications in the 1970s and 1980s in particular - been deemed by some "reliable, valid and useful" at the same time as being deemed by others "unreliable, invalid and useless" (ibid). Marsh (1987) points out in his seminal monograph that this huge variation in opinion can be ascribed to four sources:

1. The biased views brought to SET research by those who investigate it;
2. The naïve and unrealistic perceptions of the functions SETs can fulfil;
3. The myriad poorly-constructed, invalid instruments that are available to capture SETs;
4. The lack of a unified, robust approach to the construction of SET instruments and the research associated with them.

While Marsh was writing almost three decades prior to this thesis being penned, the four sources of discrepancy on views of SETs exist to this day, including in the UK. The interest in and contentions surrounding student voice (BBC, 2010; Robinson & Taylor, 2013)

indicate that the use of students' views of teaching is often grounded in a poorly-constructed representation of robust evidence, implementing untested concepts and instruments, then using these as tools for decision-making. But this is an unsurprising conclusion when one looks more broadly to the system of primary and secondary education in the UK. Only now - with the high-quality research reports commissioned by the Education Endowment Foundation[25], the database of well-designed and faithfully-implemented studies gathered by the Institute of Education Sciences[26] in their What Works Clearinghouse[27] and the steadily-building interest in evidence-informed policy among both policy-makers and practitioners - does the history of poorly-evidenced education policies and practices have shed upon it the light (of hindsight[28]) required to recognise the failings that have led to the use of poorly-constructed, invalid instruments for inappropriate purposes. Without an education system that routinely trains its professionals to recognise, understand, interpret and use appropriately high-quality research (as the UK currently fails to do), how can it be expected that teachers and school leaders would not use invalid, unreliable measures of teaching effectiveness for a range of purposes? While ignorance should not be an excuse, it is certainly a reason. Unintended ignorance sits at the heart of the matter, accompanied closely by the pressing necessities of policy initiatives (most of which, in recent history, have been launched in Whitehall and enacted in various forms at the far-flung ends of its reach) and the need to be seen to respond appropriately to regulatory mandates and trends.

[25] http://educationendowmentfoundation.org.uk/projects/projects-a-z/?tile=3&ids=0|531|595|529|530 accessed 10/6/14
[26] http://ies.ed.gov/ accessed 10/6/14
[27] http://ies.ed.gov/ncee/wwc/ accessed 10/6/14
[28] It is easy to judge when one has the luxury of hindsight; this statement is not meant as a judgement, rather as an observation which recognises the luxury.

Without adequate, sustained training and the support to access, understand and use SET (and other) research, the future for SET use in the UK is quite bleak. It is part of the mission of this thesis, therefore, not only to present the conclusions of a validation process and an impact evaluation of an intervention conducted in secondary schools, but also to restate the case for the reliability and validity of SETs both in higher and secondary education contexts. In doing so, it is hoped that an extra voice is added to the call for only properly validated and well-researched instruments to be used (in all settings). This is made more pertinent by the growing interest in evidence-based teaching and in more rigorous approaches to the measurement of teacher effectiveness.

2.6 Purposes of SETs

The proliferation of SETs in universities in North America is well-documented (d'Apollonia & Abrami, 1997; Dommeyer, Baum, Chapman, & Hanna, 2002; Richardson, 2005). In this setting, four key purposes of SETs were identified by Marsh and Dunkin (1992):

- Diagnostic feedback provided to teachers on their teaching effectiveness;
- Administrative decision-making information;
- Information for student use in the selection of courses and instructors;
- An outcome or process description in research on teaching.

One of the commonly-quoted observations made by the authors is about the ubiquity of the first purpose, particularly in respect of the variability of usage of the subsequent three. It is, perhaps, understandable that diagnostic feedback derived from SET data would be a common feature of its use; certainly in the context of this PhD study, this diagnostic function is performed by SET data assuming a formative function. Moreover, it is contended

that SET data alone used for high-stakes purposes (as, certainly the second and third purposes listed above may be termed) are not sufficiently reliable a measure, offering, perhaps, an indication of why their usage in these contexts is variable.

It is interesting to consider the significant variation in the uses to which SETs are put. Variously, they have informed promotion decisions, been sold to students in university bookstores as a means to assist with course selection decisions, and, in other settings, been considered data to be collected in the strictest confidence. In consideration of this multiplicity of functions, it is enlightening to refer to Newton (2007):

> "Once a system has been designed with an explicit prioritization of purposes in mind, the operational problem will then become how to ensure that results are not used for inappropriate purposes." (p.168)

Defining the purpose of an assessment before the assessment takes place is critical, but is only half of the ultimate battle. One may have a very clear statement of intent for assessment data, but, as Newton points out, the key real-world problem is the prevention of inappropriate uses of them (or simply uses for which they were not intended). This is a notion that has had a significant impact on the design of my PhD study, especially given the potentially sensitive nature of the data generated, and the purpose to which they were to be put: teacher professional development. To this end, a decision was taken in the early part of the design process to adhere to two clear rules:

1. Class-average SET data (the recommended level-for analysis of the data) will be shared only with the teacher to whom they refer;

2. The only stated use of the data is as a basis for a set of teacher-led professional development decisions and actions (defined by the teacher to whom the data refer).

For the purposes of clarity and transparency, these messages were communicated both to teachers interested in participating, and to the leadership of the schools in which these teachers were based. Consequently, an environment was created where all involved knew clearly what the data could be used for, and what they should not be used for. Further detail on this is found in Chapter 6.

2.7 Unidimensionality or multidimensionality?

The idea that teaching effectiveness is a unidimensional construct appears, at face value, to be somewhat implausible as well as quite dispiriting. That there might be a singular entity called 'effective teaching' or "general instructional skill" (Abrami & d'Apollonia, 1991) that cannot be understood in terms of its differential constituent parts is anathema - even unrecognisable - to teachers. Indeed, it may be argued that it is precisely the delicate balancing-act that goes on daily in classrooms that makes for teaching being a profession, rather than simply an occupation. The nuanced and judicious use of humour, recapitulation, feedback, interaction, enthusiasm and other tools in the teacher's 'toolkit' point towards an endeavour that is multi-faceted, complex and realised through behaviours (initially those of the teacher and, subsequently, those of their students).

Equally, the unidimensional construct argument is a depressing one. Assuming that the multiple dimensions of teaching effectiveness that are seen in instruments from Feldman (1976), through Marsh (1982), Ferguson (2008) and Fauth, Decristan, Rieser, Klieme, and

Büttner (2014) are realised through the behaviours of teachers, one can legitimately conclude that this multifaceted undertaking is one that, through measurement, feedback and consultation, might be improved.

During my time as a teacher, I found that instructing a student to 'work harder' or 'do better' was a lazy and generally fruitless way of encouraging improvement, chiefly because it lacks the kind of specificity and staged approach needed. Similarly, asking a teacher to 'be better' in the classroom represents a recognition of teaching as a profession whose actions are built upon a set of behaviours. Consequently, differentiating between the purposes of SET ratings is critical to the dimensionality argument, given that diagnostic, formative purposes would be better served by multidimensional information, while summative decisions would more profitably be served by a unidimensional measurement.

Perhaps a more productive way to view teaching (and the evaluation of it by students) is as both a unidimensional construct *and* a multidimensional one. The dependency here is placed upon the purpose of the assessment: if one is interested in evaluation for an administrative purpose (retention of teachers or pay decisions, for instance), a multidimensional construct becomes so unwieldy as to be rendered useless. Likewise, assessing the effectiveness of teaching for the purpose of furnishing an individual teacher with diagnostic information about specific aspects of their practice (which, individually, are more likely to be improvable than simply 'teaching' in general terms) is a purpose better served by an approach with multidimensionality at its core. Indeed, Apodaca and Grad (2005) suggest that the SEEQ (Marsh, 1982) (and other instruments discussed by the authors) are disposed to a general factor, asserting the potential unidimensionality of the rating form itself. Furthering this argument, Abrami, d'Apollonia and colleagues (Abrami &

d'Apollonia, 1991; Abrami, d'Apollonia, & Cohen, 1990; Abrami, d'Apollonia, & Rosenfield, 1997) make a case for teaching to be considered as a multidimensional construct, but students' perceptions as captured by SET ratings to be unidimensional in essence. This thesis will present a case in subsequent chapters that teaching is a multidimensional construct, and that the SecondarySEEQ instrument captures multidimensional information about teaching which may be used as a diagnostic tool to help teachers identify areas of practice for improvement. In summary, this project is, in this respect at least, driven by the concept that the purpose of an assessment should be clear at the design stage (Newton, 2007), and that this purpose (summative or formative, for instance) will dictate whether or not a unidimensional model, a multidimensional model, or a mixed (both general factor and distinct unique factors) model is appropriate. By presenting a structural model almost identical to that proposed by Marsh (1982), I will conclude that a multidimensional model fits the SecondarySEEQ data (from the current investigation) best, and that this model is the most parsimonious and useful approach in supporting a consultative feedback intervention for teachers based on SET data.

2.8 Specification of dimensions

In accepting that teaching is a multidimensional construct (as is supported in the literature as a theory of effective teaching), an exploration of precisely which dimensions should be measured must be undertaken. A key moment in the history of well-validated instruments came with the 19 ratings factors described by Feldman (1976), as denoted in Table 2.1 below.

19 Instructional Rating Dimensions Adapted from Feldman (1976) (source: Marsh, 1987)	
1	Teacher's stimulation of interest in the course and subject matter.
2	Teacher's enthusiasm for subject or for teaching.
3	Teacher's knowledge of the subject.
4	Teacher's intellectual expansiveness and breadth of coverage.
5	Teacher's preparation and organisation of the course.
6	Clarity and understandableness (sic) of presentations and explanations.
7	Teacher's elocutionary skills.
8	Teacher's sensitivity to, and concern with, class level and progress.
9	Clarity of course objectives and requirements.
10	Nature and value of the course material including its usefulness and relevance.
11	Nature and usefulness of supplementary materials and teaching aids.
12	Difficulty and workload of the course.
13	Teacher's fairness and impartiality of evaluation of students; quality of exams.
14	Classroom management.
15	Nature, quality and frequency of feedback from teacher to students.
16	Teacher's encouragement of questions and discussion, and openness to others' opinions.
17	Intellectual challenge and encouragement of independent thought.
18	Teacher's concern and respect for students; friendliness of the teacher.
19	Teacher's availability and helpfulness.

Table 2.1: 19 factors identified by Feldman (1976)

Feldman later amended the instrument by including two extra dimensions relating to the outcome of teaching and the personality of the teacher, while dimensions 12 and 14 from the table above were removed. Interestingly, dimension 14 (classroom management) is the factor noted by Ferguson (2008) as the one most highly-predictive of student progress gains, and one echoed in the instrument adapted for this study in the Organisation and Clarity factor. Fauth et al. (2014) also draw similar conclusions about classroom management; the aforementioned factor presented in their work with elementary school children in German consisted mainly of items referring to children not talking when others were, being quiet when the teacher spoke, and following instructions.

Marsh (1987) cites four other instruments which, in his opinion, represent the best the literature has to offer in terms of well-defined factor structures, the measurements of which are indicative of distinct aspects of teaching:

- The Endeavor instrument (Frey, Leonard, & Beatty, 1975): Presentation Clarity; Workload; Personal Attention; Class Discussion; Organisation / Planning; Grading; Student Accomplishments.
- The Student Description of Teaching (SDT) instrument (Hildebrand, Wilson, & Dienst, 1971): Analytic / Synthetic Approach; Organisation / Clarity; Instructor Group Interaction; Instructor Individual Interaction; Dynamism / Enthusiasm.
- The Student Evaluation of Educational Quality (SEEQ) instrument (Marsh, 1982): Learning / Value; Instructor Enthusiasm; Organisation; Individual Rapport; Group Interaction; Breadth of Coverage; Examinations / Grading; Assignments / Readings; Workload / Difficulty.

- The Michigan State SIRS instrument (Warrington, 1973): Instructor Involvement; Student Interest and Performance; Student-Instructor Interaction; Course Demands; Course Organisation.

These instruments, by virtue of their systematic approach to development and similarity of aspects of teaching measured (as noted by Marsh (1987), make well-supported construct validity claims. Factor analyses of each instrument show clearly that SETs do, indeed, measure distinct aspects of teaching, a finding supported byFeldman (1976).

2.9 Bias

It is important to recognise, however, that students rating their teachers using even the best-developed, most roundly-validated instrument are only able to offer evaluations of those features of teaching which are observable by them. Consequently, certain items will only indicate a partial view which, ultimately, must be considered biased (Abrami et al., 1997). For instance, as Apodaca and Grad (2005) point out, students' judgements of the observable aspects of planning may be quite reliable (structured lessons using appropriate resources), whereas their ability to determine whether or not a teacher has delivered an appropriate breadth of course material may be less so. In later chapters, my analysis of the instrument validated in this project will address bias directly.

2.10 The validity and impact of SETs

The notion of the validity of SETs is one which, itself, requires clarification before the following discussion. The position adopted by Abrami et al. (1990) points to two perspectives on the validity of SETs: the accuracy of SETs as a representation of the views students hold about the quality of teaching (irrespective of any impact on learning

outcomes). This is very much a value proposition which places the student in the role of consumer; the authors note that SETs rarely come in for criticism as measures of satisfaction with teaching. The second perspective put forward by Abrami et al. (1990) is that SETs may be considered valid if they reflect accurately the effectiveness of the rated teacher's teaching; it is from this second perspective, the authors argue, that criticism of SETs' validity often emerges. Once again, it is useful to rehearse the argument that the measurement of teaching quality is a challenging endeavour about which there is no consensus. In this study, the standardised value-added metric generated by analysis of pre-test and post-test data will be used as a measure of pupil progress over time and, thereby, the measurement will be conceived as a proxy for learning.

In the current PhD study, investigating the validity of SETs is operationalised with a very narrow focus. That focus is on an instrument based on the Student Evaluation of Educational Quality instrument (Marsh, 1982), known as the SecondarySEEQ, and its ability to reflect.

Consideration of the validity of SETs should begin with a discussion of the proposed purposes of them. Spooren, Brockx, and Mortelmans (2013) address this matter directly (with reference to Kember, Leung, and Kwan (2002)) and indicate the following three purposes:

1. improving teaching quality;
2. informing appraisal exercises;
3. informing processes of institutional accountability.

Different stakeholders in a school or university place differing demands on SETs: teachers' interest lies mainly with using the information gleaned from SETs for the purpose of improving their practice; managers and leaders share this purpose, but often add the need to provide information in support of appraisal processes.[29]

2.11 The Student Evaluation of Educational Quality instrument

The instrument with the broadest documented use, and the one most studied, is Marsh's (1982) Student Evaluation of Educational Quality (SEEQ) instrument. Students invited to complete this survey respond to 37 items (with additional items added at the discretion of the instrument administrator) which form the basis of a nine-factor scale, seen below:

- Learning / value
- Enthusiasm
- Organisation
- Group interaction
- Individual rapport
- Breadth of coverage
- Exams
- Assignments
- Workload

Using a Likert scale, the SEEQ collects individual student ratings of teaching quality which are subsequently aggregated at the class level across the nine factors. Analyses by Marsh and Hocevar (1991a) indicate both the reliability of student evaluations collected using the instrument, as well as evidence suggesting that these evaluations are a function of the

[29] Interestingly, Clipa (2011) points to a book by Jinga and Negreţ-Dobridor (2004) in which the purposes of evaluating teaching quality were surveyed in different countries. The latter found that English uses include the measurement of institution-level quality; teaching and learning process improvement; continuing professional development (CPD). In Scotland, the improvement of teaching practise during teacher training featured, as did enhancement of professionalism and teaching standards more generally.

teacher teaching the course, rather than of the course they teach. Stability over time is also well-documented (Marsh & Hocevar, 1991b), as are the multiple dimensions of teaching quality the instrument is designed to capture (Marsh, 1991) - though this latter point has been disputed, with Abrami and d'Apollonia (1991) asserting that a single, global factor (which they termed 'general instructional skill') could be derived from the SEEQ. Subsequent work has, however, illustrated the multidimensional structure described by the SEEQ (Marsh, 2007b). Further support for the multidimensionality of teaching effectiveness is also found in Burdsal and Harrison (2008) and in Chapter 7 of this thesis. An example of the SEEQ is found in Appendix 2a.

Ultimately, this study will draw on the recent work of Newton and Shaw (2014) to assess the validity of the SecondarySEEQ instrument, investigating the technical quality of the measurement, as well as the value to education policy of data gathered by the instrument.

2.12 Feedback for goal-setting

SET instruments provide feedback on student perceptions of teaching, but this alone is unlikely to have an impact on teachers' actions and behaviour and, subsequently, on student achievement. Therefore, it is important to understand how such an impact may be effected. If it is assumed that a SET instrument such as that trialled in this PhD study can provide valid measurements of teaching effectiveness, work done to understand how these data can be used to support teacher development is valuable.

The notion that the conscious creation of goals affects actions is one developed in the early 1970s by the American psychologist Ryan (1970). Somewhat presciently (given the apparent dearth of research on motivation and goal-setting), the argument was presented by Ryan (1970) that "it seems a simple fact that human behaviour is affected by conscious purposes, plans, intentions, tasks and the like". These key causal components of life's actions were referred to, as Locke and Latham (2002) explain, as 'first-level explanatory concepts'.

For the purposes of this study, I adopt the definition of a goal used by Locke and Latham (2002):

> "A goal is the object or aim of an action, for example, to attain a specific standard of proficiency, usually within a specified time limit." (p.705)

Such a 'first-level explanatory concept' forms a critical component of this study, but is usefully contrasted in light of the oft-used, much vaunted exhortation to 'do your best'.

2.13 Doing one's best

To urge someone simply to do their 'best' in the pursuit of the improvement of performance is an ineffective strategy, chiefly due to the subjectivity and lack of concrete reference points with which to compare subsequent performance. Meta-analyses have found effect sizes ranging from 0.42 to 0.80 derived from instances when challenging and precise goals were set (Locke & Latham, 1990), providing support to the intuitive notion that clear, concise aims lead to more fruitful outcomes.

Precision and clarity are features of goal-setting echoed in the elegant model proposed by Locke and Latham (2002), the core of which is the specificity and difficulty of the goal set: it should be sufficiently challenging to warrant effort, and adequately precise in description as to be recognisable if achieved. The goal is then moderated, the authors propose, by an individual's commitment to the goal, the importance placed upon it, along with the self-efficacy perceived by the individual to achieve it, the feedback provided to support its achievement, and the level of complexity associated with achievement. The choices made by the individual seeking the goal, their effort, persistence and decisions regarding appropriate strategies to help move towards successful achievement of the goal are all mechanisms in the high-performance cycle; these, along with the moderators mentioned above have an impact on the performance of the individual in their attempt to achieve their goal. Finally, the authors propose that satisfaction with performance and the rewards associated with it lead to a willingness to commit to new challenges, and thereby feed into subsequent cycles of goal-setting and performance. As will be seen in subsequent chapters, this cycle is a core component of this study's intervention design.

Marsh (2011) refers to university instructors' needs for feedback on the specific task of improving student ratings (using the SEEQ instrument) by improving specific areas of their teaching. Aligning closely with the conclusions about goal-setting and high performance of Locke, Latham and others, Marsh (2011) cites both Marsh (2007a) and Marsh and Overall (1980) in stating that SET feedback given to teachers has a modest positive effect compared to control group teachers receiving no feedback (0.33SD higher than the control group on average). He also notes, however, that the implementation of a consultative feedback intervention had "substantially" larger positive effects, a finding echoed by Penny and Coe (2004), and one used to inform this study.

But the provision of feedback for the purpose of goal-setting is best viewed in the context of other forms of goal-setting, with specific reference being made to their relative effectiveness. Hattie and Timperley (2007) summarised work done by Kluger and DeNisi (1996) to present findings in the latter's significant meta-analysis which contextualise feedback with goal-setting as the moderating feature, as seen in Table 2.2 below.

Moderator	Number of effects	Effect size
Correct feedback		
'Tis correct	114	0.43
'Tis incorrect	197	0.25
Task feedback about changes from previous trials		
Yes	50	0.55
No	380	0.28
Task feedback designed to discourage the student		
Yes	49	-0.14
No	388	0.33
Praise feedback about the task		
Yes	80	0.09
No	358	0.34
Feedback provided from a computer		
Yes	87	0.41
No	337	0.23
Number of times feedback was provided		
Lots	97	0.32
Little	171	0.39
Task complexity		
Very complex	107	0.03
Not complex	114	0.55
Goal setting		
Difficult goals	37	0.51
Easy, do your best goals	373	0.30
Threat to self-esteem		
Much threat	102	0.08
Little threat	170	0.47

Table 2.2: Summary of effect sizes relating to types of feedback

Certainly of interest here is the difference in effect size found between very complex and non-complex tasks. Recent evidence on behavioural change summarised by the Behavioural Insights Team (BIT) in their EAST pamphlet (Service et al., 2014) points to four core traits of interventions which are more likely to succeed in changing people's behaviours. BIT suggests in the publication that making it easy, attractive, social and timely

for people to change their behaviour increases the likelihood of them doing so; the first of these elements may be interpreted as 'non-complex'. Acknowledging that human behaviour is often not rational, the conclusions drawn by authors from a variety of fields agree: reducing the complexity of a task increases the likelihood of increasing its effect. The relevant extension from this is to suggest that a reduction in the complexity of the task undertaken by a teacher - in light of feedback on their teaching - increases the likelihood of the feedback provided having a positive effect. This, however, is a problematic assumption.

To assume that teaching is a task whose complexity can be reduced or simplified is bold, possibly even foolish. It may not be possible to simplify teaching. Nonetheless, the approach adopted in this study (as shown in subsequent chapters) is to focus a teacher's mind on one or two specific strategies designed to improve their teaching. Though this does not necessarily reduce task complexity, it specifies and limits the tasks undertaken, hopefully generating a sense that they are at least manageable, if not non-complex.

While the lesson taught by Cartwright and Hardie (2012) to ensure that simple, unmediated translation of findings from one context to another does *not* occur is brought to mind here - the results presented above do not directly and exclusively relate to the population of interest in this study as many of the studies involved were classroom-based, involving students - this is one of the more robust meta-analytic reports of effects sizes from feedback generally, and should not be ignored. A discussion of the conclusions drawn by Hattie and Timperley (2007), in light of the illuminating work by Kluger and DeNisi (1996), which relates to how the findings around effective feedback impact the design of the intervention in this study is taken up in Chapter 6, but at this juncture it is useful to draw

attention to the positive effects of feedback regarding low-complexity but high-difficulty goals, in which a computer is used to provide the feedback. (The features of this are also used in the design of the intervention under evaluation in this study.) Moreover, the effect of diminished threats to self-esteem is a thread that will be picked up again when discussing the construction of the peer-coached feedback intervention with a single, self-chosen, trusted colleague.

One aspect of the provision of feedback that was extended in this study was the use of solution-focused brief therapy (SfBT) to inform the online meeting guide used by participants when holding their peer-coached feedback meetings.

2.14 Solution-focused Brief Therapy

Solution-focused Brief Therapy (SfBT) is a form of cognitive behavioural therapy which seeks to play on the strengths of an individual patient, with a view to using these in some kind of change process (Corcoran & Pillai, 2007). Rather than focusing on a retrospective view - one which highlights the history of a problem - SfBT looks to a future that is perceived as better than the present by means of visualising an idealised state. This is achieved by identifying instances from the individual's past when a particular solution has helped them to overcome difficulties or improve situations, then generating an understanding of how, under current circumstances, this solution could be used again for positive change (de Shazer, 1994).

Key assumptions which underpin SfBT and which are desirable for therapists to acknowledge are emphasised in George, Iveson, and Ratner (2000) as follows:

- in resolving a problem, it is unnecessary to understand its cause;

- the successful outcome of therapy depends on knowing a client's goals;
- irrespective of the apparent permanence of a problem, there are always times when a client is already engaged in the creation and development of solutions;
- problems are not indicators of underlying pathologies or deficits;
- even the smallest of changes is needed to begin generating a solution to a problem.

The rationale for adopting a framework based on Solution-focused Brief Therapy for the peer-coached intervention refers back to the conclusions regarding task complexity reported by Kluger and DeNisi (1996) and echoed more recently by the Behavioural Insights Team (Service et al., 2014). Acknowledging that reducing the complexity of the task would likely increase the potential effect of the intervention (Kluger & DeNisi, 1996), the relative simplicity of SfBT and the brevity of its duration provided a framework which was both overt to teachers, but also which was clearly delineated in terms of content and time. Ostensibly, the peer-coached meetings in which teachers in the intervention arm were expected to engage were created to appear easy to manage, feasible and valuable to those involved.

A description of how SfBT is used in the design of the intervention is given in Chapter 6. It is important to say here, however, that in part, this form of therapy influenced the design of the intervention by way of replicating some aspects of the professional learning community (albeit in miniature, in this instance).

2.15 Professional Learning Communities

An interesting finding published by Roche and Marsh (2002) showed that teachers' perceptions of their teaching grew in consistency with their students' evaluations,

suggesting that while SETs may not alter the teaching behaviours of teachers, the perceptions teachers have of their own practice may be altered by the process of evaluation by their students.

In the United States of America, Darling-Hammond and McLaughlin (1995) identified greater focus on the content, timeliness and modes of teacher continuing professional development as being prime considerations in ensuring the creation of a model best-suited to the changing needs of the United States' educational reform agenda. One development spurred on by this agenda, and which this PhD project uses as a guide, was the professional learning community (PLC) or community of practice. Louis and Marks (1998) propound the usefulness of PLCs as a means of maximising professional learning opportunities in schools and, as is the focus of much of the literature on the topic, highlight the shift from a concentration solely on teaching, to teaching *and* learning. Their work drew on the five key components of PLCs as defined by Newmann and Wehlage (1996):

- shared conceptualisations about children, school priorities, use of time and space in school, and teachers' roles and responsibilities;
- clarity of focus on student learning;
- a dialogue among teachers that enables reflection on teaching, student progress and curriculum;
- making the practice of teaching a more public / less private one;
- focusing on collaboration.

By integrating the core tenets of both solution-focused brief therapy and professional learning communities, this study seeks to create a safe, confidential environment for the teacher receiving the feedback, and one in which the focus is firmly on forward movement,

as opposed to a retrospective trawl through the 'what ifs' and 'if onlys' of their teaching practice.

2.16 Conclusion

This literature review serves as a survey of the landscape of the substantive and theoretical areas which were identified as relevant to this study: validity, feedback, student evaluation of teaching, goal-setting and teachers' continuing professional development. The evidence of the validity and impact of student evaluation of teaching instruments has been mentioned briefly here, but receives a more significant investigation and discussion later in this thesis.

Alongside the function of situating this thesis in the research which has come before, the review of the literature also serves to inform and shape the design of the intervention and the investigation which accompanies it. In stating this, however, it should be noted that the limitations placed on my work by the consequences of publication bias are acknowledged: I can only review and learn from the literature which has been published.

In the following chapters, I turn my attention to the investigation formed in light of my hypothesis and the literature reviewed.

Chapter 3: An overview of the investigation

This chapter marks a key point of departure from previous chapters. Having provided the relevant background and rationale for the study and reviewed the literature, the thesis now moves to describe in brief the investigation that was carried out. As part of this overview, the methodological underpinnings of the study design are set out. Following this chapter, full details of the separate components of the investigation (the feasibility study and its conclusions, as well as the main study and associated conclusions) are given.

As the thesis moves into describing the design of an evaluation and reporting the results of it, it is appropriate to include at this point the CONSORT (2010) checklist (Schulz, Altman, & Moher, 2010) as this offers a best-practice way of ensuring that appropriate pieces of information are included.

Table 3.1: CONSORT 2010 checklist of information to include when reporting a cluster randomised trial

Section/topic	Item no	Standard checklist item	Extension for cluster designs	Page no *
Title and abstract				
	1a	Identification as a randomised trial in the title	Identification as a cluster randomised trial in the title	1
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)[i,ii]		2
Introduction				
Background and objectives	2a	Scientific background and explanation of rationale	Rationale for using a cluster design	69
	2b	Specific objectives or hypotheses	Whether objectives pertain to the cluster level, the individual participant level or both	69
Methods				

Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	Definition of cluster and description of how the design features apply to the clusters	74
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons		
Participants	4a	Eligibility criteria for participants	Eligibility criteria for clusters	81
	4b	Settings and locations where the data were collected		
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	Whether interventions pertain to the cluster level, the individual participant level or both	
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	Whether outcome measures pertain to the cluster level, the individual participant level or both	
	6b	Any changes to trial outcomes after the trial commenced, with reasons		
Sample size	7a	How sample size was determined	Method of calculation, number of clusters(s) (and whether equal or unequal cluster sizes are assumed), cluster size, a coefficient of intracluster correlation (ICC or *k*), and an indication of its uncertainty	74
	7b	When applicable, explanation of any interim analyses and stopping guidelines		
Randomisation:				
Sequence generation	8a	Method used to generate the random allocation sequence		77
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	Details of stratification or matching if used	77
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	Specification that allocation was based on clusters rather than individuals and whether allocation concealment (if any) was at the cluster level, the individual participant level or both	84
Implementation	10	Who generated the random allocation	Replace by 10a, 10b and 10c	77

		sequence, who enrolled participants, and who assigned participants to interventions		
	10a		Who generated the random allocation sequence, who enrolled clusters, and who assigned clusters to interventions	84
	10b		Mechanism by which individual participants were included in clusters for the purposes of the trial (such as complete enumeration, random sampling)	84
	10c		From whom consent was sought (representatives of the cluster, or individual cluster members, or both), and whether consent was sought before or after randomisation	81
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how		
	11b	If relevant, description of the similarity of interventions		
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	How clustering was taken into account	186
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses		
Results				
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	For each group, the numbers of clusters that were randomly assigned, received intended treatment, and were analysed for the primary outcome	80
	13b	For each group, losses and exclusions after randomisation, together with reasons	For each group, losses and exclusions for both clusters and individual cluster members	80
Recruitment	14a	Dates defining the periods of recruitment and follow-up		81
	14b	Why the trial ended or was stopped		

Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	Baseline characteristics for the individual and cluster levels as applicable for each group	85
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	For each group, number of clusters included in each analysis	80
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	Results at the individual or cluster level as applicable and a coefficient of intracluster correlation (ICC or k) for each primary outcome	190
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended		
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory		
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms[iii])		
Discussion				196
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses		199
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	Generalisability to clusters and/or individual participants (as relevant)	200
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence		198
Other information				
Registration	23	Registration number and name of trial registry		
Protocol	24	Where the full trial protocol can be accessed, if available		203
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders		

** Note: page numbers optional depending on journal requirements*

3.1 Methodology

The investigation is composed of two main parts: a feasibility study and a main trial. The feasibility study was conducted in a single comprehensive school in north-east England, with ten teachers and their classes participating (leading to 227 student evaluation of teaching cases being generated), and set out to determine whether or not it was feasible (for the teachers, their students and this researcher) to collect student evaluation of teaching data in secondary schools.

The main study, a cluster randomised controlled trial, set out to evaluate the impact of a peer-coached feedback intervention for teachers conducted in school, with trusted colleagues assuming the role of peer-coach. This RCT recruited 36 schools, each of which was randomised to either the intervention arm (student evaluation of teaching feedback, a peer-coached meeting with a trusted colleague which was supported by resources designed to help teachers improve their practice in response to the data presented to them) or the active control arm (in which teachers received only the student evaluation of teaching feedback). The decision was taken not to incorporate a pure control group (doing 'business as usual') in the trial for reasons of pragmatics: in recruiting schools to the trial, it became clear from my discussions with Senior Leaders that the prospect of not receiving anything in return for participation (despite the waiting-list design offering the possibility of receipt of the intervention) would potentially have an adverse impact on retention plans. While the use of an active control condition - in which teachers received an SET feedback chart showing their class average ratings but no consultative peer-coached intervention - was pragmatic, it is acknowledged that the potential for reducing the detectable effect of the intervention is present.

This part of Chapter 3 describes both the process adopted to ascertain the validity of the instrument used (an adapted version of Marsh's (1982) Student Evaluation of Educational Quality instrument), and that used to determine the impact on student attainment of its use, in conjunction with a peer-coached feedback intervention for teachers. The instrument and the evaluation in the study were designed with the following research questions in mind:

1. Is the SecondarySEEQ instrument used in this PhD study a valid means with which to capture measurements of teaching effectiveness?
2. Does the presence of a termly, one-hour, peer-coached teacher consultation meeting based on SecondarySEEQ student evaluation of teaching data (treatment) create a larger effect on attainment than the absence of it (control)?

Beginning with a brief discussion of the rationale for the design, the chapter thereafter focuses on the validation process; the use of a randomised controlled trial is then discussed and described. Finally, the processes undertaken to collect data are described, as are the details of the intervention's content implementation.

3.2 Rationale for using an experimental design

The most robust means by which a causal inference may be made about the impact on a desired outcome of a specific intervention is a randomised controlled trial, or RCT, a point made clear in the literature (Eldridge & Kerry, 2012; Torgerson & Torgerson, 2008). To ensure the trial design was as sound as possible, cluster randomisation was used, and this will be expanded upon in subsequent sections.

The design of this study is most recognisable as adopting a pragmatic, mixed methods approach (Tashakkori & Teddlie, 2003), though philosophically it is situated in the discussion of paradigms as problematised entities (Coe, 2012). Mixed methods in this study is not viewed as a paradigm in itself (contrary to Onwuegbuzie, 2004), but as a tool to communicate that this is neither purely quantitative nor qualitative, as an undertaking. The notion of the separateness of paradigms seems somewhat unhelpful in the context of a study that adopts the pragmatist's approach of finding the appropriate methods to meet the desired aim. This study has an overarching quantitative bent, but the quantitative data are nothing without the qualitative data which accompany them; similarly, the narratives of teachers and students which form much of the latter data would be of interest but little use without their quantitative companions.

3.3 Randomised controlled trials

There is a strong argument that randomised controlled trials are the 'gold standard' of research designs (Torgerson & Torgerson, 2001), yet the notion is a flawed one. While there may be a hierarchy of investigations from a methodological standpoint, there is no such thing as a 'gold standard'; it should never be considered the route to an investigation free from certain kinds of bias, but rather, the best option available to answer a specific question. The job of the investigator is to select the most appropriate design to enable a question to be answered, rather than simply to select that which is labelled as the 'best'.

From the agricultural beginnings of the RCT with R.A. Fisher in the early part of the twentieth century - the language of which still resonates in the terminology of the RCT (Torgerson & Torgerson, 2008) - through Lindquist's (Lindquist, 1940) work in education on avoiding confounding through clustering, the applications of the approach have resonated

with social sciences research. It is in the clinical trials of medicine, however, that the RCT has been most used, and with which it has become most commonly associated. That said, the work of such organisations as the Centre for Evaluation and Monitoring[30] (CEM) at Durham University, the Coalition for Evidence-based Education[31] at York University and the What Works Clearinghouse[32] in the USA speak to a growing interest in evidence-based policy and practice in education. In early 2013, the MET Project's final report (Cantrell & Kane, 2013) was released to great interest; its suggestion that the effectiveness of teachers could be measured came as a result of a large-scale randomised controlled trial (though there is some dispute over the robustness of the claims made in the project (Rothstein & Mathis, 2013)).

The randomised controlled trial offers the best means by which selection bias, regression to the mean and changes over time can be controlled (Torgerson & Torgerson, 2008) by seeking equivalence in both known and unknown variables; the RCT offers the most robust way of controlling for both the observed and unobserved differences in participants.

The calls for greater use of evidence as a basis for decision-making in education have been heard clearly in the UK for well over a decade (Coe, Fitz-Gibbon, & Tymms, 2000). Moreover, advocates of the RCT design have become ever more vocal (Gorard & Cook, 2007; Torgerson & Torgerson, 2001), and to increasingly receptive ears.

The use of randomised controlled trials in the social sciences has, of late, begun to increase in popularity, and the relevance of their findings seen in a new light. Perhaps as a result of a declining global economy, a populace more interested in frugality and value for money,

[30] http://www.cemcentre.org/ accessed 20/1/13
[31] http://www.york.ac.uk/iee/coalition.htm accessed 20/1/13
[32] http://ies.ed.gov/ncee/wwc/ accessed 20/1/13

the randomised controlled trial in social sciences seems to be gaining in popularity. Notably, the release of *Test, Learn, Adapt* (Haynes, Service, Goldacre, & Torgerson, 2012) is suggestive of growing interest in RCTs at the policy level[33], both in the UK and other countries, and one might conjecture that this interest in ascertaining 'what works' prior to full implementation at scale has, in part at least, an economic motive.

Furthermore, the creation of the Education Endowment Foundation in 2011 signalled a shift in government thinking towards increased deployment of the RCT design in education contexts. Prompted by action from the Deputy Prime Minister's Office, the coalition government's endowment of £125 through the Department for Education[34] set out to address the destructive link between family socio-economic status and educational achievement. Since inception, the EEF has engaged over 1,800 schools in 56 investigations (Collins, 2013), many of which are randomised controlled trials; in June 2014, a survey by the National Foundation for Educational Research (NFER) found that 45% of school leaders were using the Sutton Trust / EEF Toolkit[35]. According to a presentation given by Sir Kevan Collins, Chief Executive of the EEF, in November 2015, this figure had risen to 65%[36].

Interestingly, the once-adversarial critic of government policy, Dr Ben Goldacre, has, of late, been visible in his support for the UK Department for Education's approach to policy-making informed by research evidence. Coming from a health science background, Goldacre's (2013) entry into the debate on the use of evidence in education heralded

[33] http://www.bbc.co.uk/news/uk-politics-19656595 accessed 16/1/13
[34] http://educationendowmentfoundation.org.uk/about accessed 20/9/13
[35] https://educationendowmentfoundation.org.uk/news/45-of-school-leaders-using-the-teaching-and-learning-toolkit/ accessed 12/11/14
[36] https://twitter.com/gl_assessment/status/662630671450570753 accessed 22/11/15

something of the shift in central government thinking. The sight of a popular[37] author and broadcaster sharing a Bethnal Green Academy stage with policy-makers and academics alike made for interesting viewing.

Clearly there is an environment emerging which suggests that there is support for the use of RCTs in education. Publishers of educational resources are beginning to recognise the commercial potential in interventions based on robust evidence, as indicated by Pearson's publication of a set of materials designed to support writing skills in secondary schools[38]. Interestingly, a prominent part of the marketing of this product focuses on the benefits of its basis in research using the RCT design, the findings of which originally appeared in the report on an ESRC-awarded grant (Myhill, Jones, & Bailey, 2011).

Perhaps this is how robust evidence in education will begin to be embedded in schools. Daniel Willingham's (2012) book (written out of "frustration with a particular problem: the word 'research' has become meaningless in education"[39]) is clearly written with practitioners in mind, and presents a model for evaluating a proposed intervention. Pearson's marketing of their writing skills publication follows this model almost entirely to the letter, stating clearly their claims for the intervention's impact, giving the credentials of the author and the research on which the product is based, highlighting the evidence offered and engaging teachers' professional desire to raise attainment in a crucial area. This method is nothing new, echoing the structure of rhetoric as described by Aristotle

[37] Goldacre had 292,725 followers on his official Twitter account https://twitter.com/bengoldacre accessed 20/9/13

[38] http://www.pearsonschoolsandfecolleges.co.uk/Secondary/GlobalPages/SkillsforWriting/SkillsforWriting.aspx accessed 20/9/13

[39] http://www.danielwillingham.com/1/post/2012/07/how-can-you-know-if-its-really-research-based.html accessed 03/04/13

(Burnyeat, 1994): ethos, logos, pathos. Perhaps the combination of research, economic imperatives, educational imperatives and rhetoric may offer a fruitful path for embedding evidence in education.

3.4 Design of the Secondary Student Evaluation of Educational Quality randomised controlled trial

In planning the RCT for this study, Torgerson and Torgerson (2008) highlighted the need for a high-quality and robust design. Moreover, the guidance offered in CONSORT (M. K. Campbell, Elbourne, & Altman, 2004) demonstrates the need for and process of high-quality reporting of trials. As such, these texts were instrumental in the design of this study. Indeed, the structure of much of this thesis follows the CONSORT statement and its extension to cluster randomised trials. In planning the design and implementation of an RCT, it seemed appropriate that one reverse-engineered to meet the needs of a CONSORT report would start life with the correct intent at least.

The eligibility criteria for participation as a cluster in this trial were as follows:

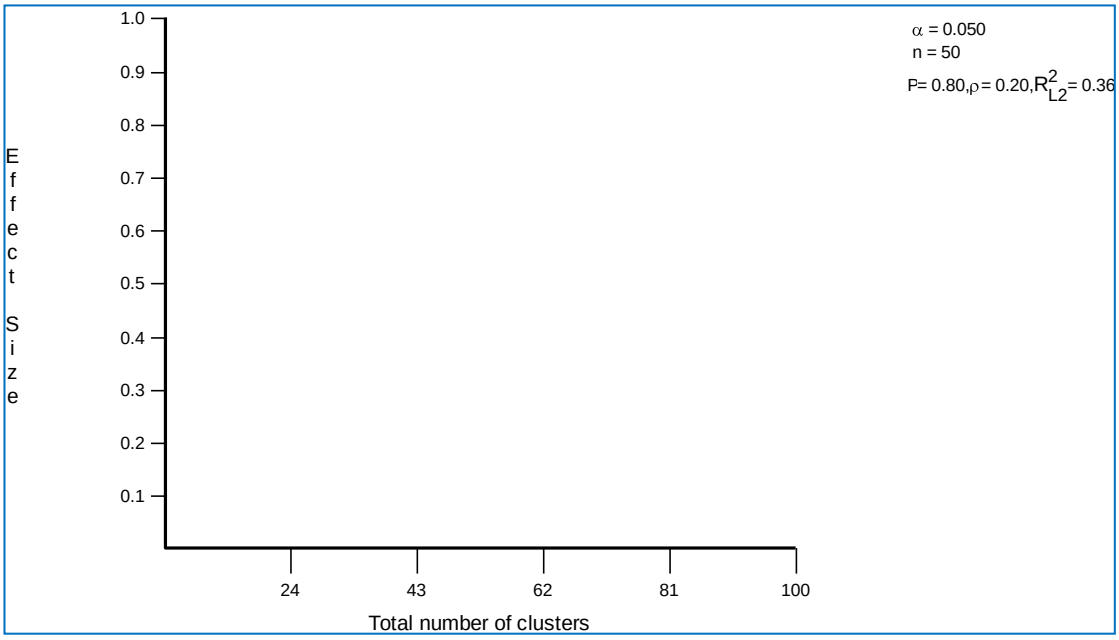
- Students must be studying subjects with the aim of achieving Advanced Level and Advanced Subsidiary Level qualifications;
- The school / college must be registered to use the Advanced Level Information System (ALIS) with the students studying for Advanced Level and Advanced Subsidiary Level qualifications.

3.5 Sample size calculation

It was estimated that an adequately powered evaluation would require a sample of at least 48 schools. Assuming random allocation of schools with 50 students participating in

each school and an ICC of 0.2; the availability of covariates with a multiple correlation of 0.6 with the primary outcome measure; a Type-1 error rate (alpha) of 0.05 and power of 0.8: with a total of 48 schools (24 experimental, 24 control), a 0.3 effect size could be detected. This minimum detectable effect size (MDES) is considered sufficient to be of interest if it is found in the trial results, but also low enough to find what is, in reality, a modest effect. Nonetheless, it should be noted that, were the intervention scalable (as it is designed to be), an effect size of 0.3 would be of considerable interest.

The graph (from Raudenbush's Optimal Design software[40]) shows the relationship between minimum detectable effect size and number of schools, with these assumptions:

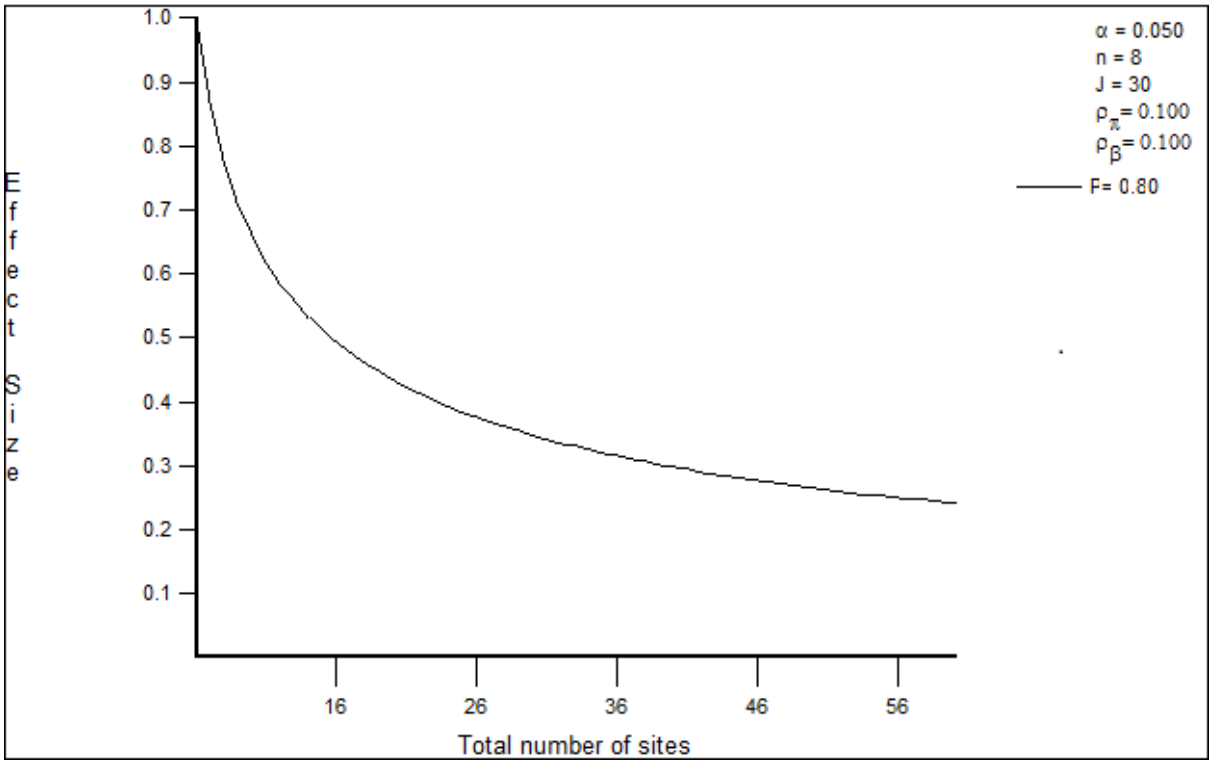


Ultimately, 36 schools were randomised and five (two in the intervention arm and three in the active control arm) did not receive the allocated intervention because they failed to

[40] http://hlmsoft.net/od/ accessed 01/11/11

respond to pre-intervention communications. These units remained in the study for analytical purposes, as the design called for Intention To Treat analysis.

It is worth stating, therefore, that this is an underpowered study. Nonetheless, with 31 (instead of the planned 48) clusters, and average size of eight teachers per cluster, the minimum detectable effect size rises only by 0.06 to 0.36.



Irrespective of this rise, an effect size of 0.3 would be considerable, when viewed in the context of small effects. Hedges (2013) notes that the following mathematics learning interventions' effects correspond with an effect size of *d*=0.26 (with student SD of *S*=38.4) and *d*=0.61 (with between-school SD of *S*=16.4[41]):

- Intervention A has an effect of 10 NAEP (National Assessment of Educational Progress) scale points (national SD of *S*=38.4);

[41] This is quite a high SD, but it serves a useful illustrative purpose.

- Intervention B can reduce the parental achievement gap between High School and college-educated parents by 34%;
- Intervention C has an effect of moving a school from the 10th percentile nationally to the 50th percentile.

As Hedges is at pains to point out, the first intervention appears to have a large effect, the second a modest impact and third is simply hard to judge, yet each is just one way to describe the same intervention effect. While the effect appears to differ due to the myriad ways in which it is observed, the fact remains that the actual effect of this intervention, if viewed through the lens of Jacob Cohen's guidelines would be considered small; equally, Hill et al. (2008) would also place this effect in the range between small and medium, though closer to small. But such distinctions are, in essence, unproductive, as they generalise in a manner that is devoid of meaning. One only need look at the follow-up work done on the Tennessee Class Size experiment to know that small effects scaled up can be significant, particularly in relation to both educational improvements and cost-benefit analysis (Krueger, 2000).

3.6 Covariates

The covariates used (GCSE subjects) had a multiple correlation of 0.56 with the primary outcome measure (Advanced Subsidiary and Advanced Level General Certificate of Education grades). In order to control for baseline differences, the Standardised Value Added (STDVA) variable will be used in the analysis presented in Chapter 7.

3.7 Randomisation protocol

A distance randomisation process was undertaken by Dr Adetayo Kasim, Research Statistician at the Wolfson Research Institute for Health and Wellbeing. The importance of concealing the randomisation process from the principal investigator (who recruited and registered participants for the trial) was not underestimated. Each school recruited was identified during the randomisation process only by a unique code of between two and four digits; it was not possible for Dr Kasim to identify any school by name, type or location. Moreover, the list of units to be randomised was sent securely from Durham, UK, to Dr Kasim in Stockton, UK, and the results returned by the same route.

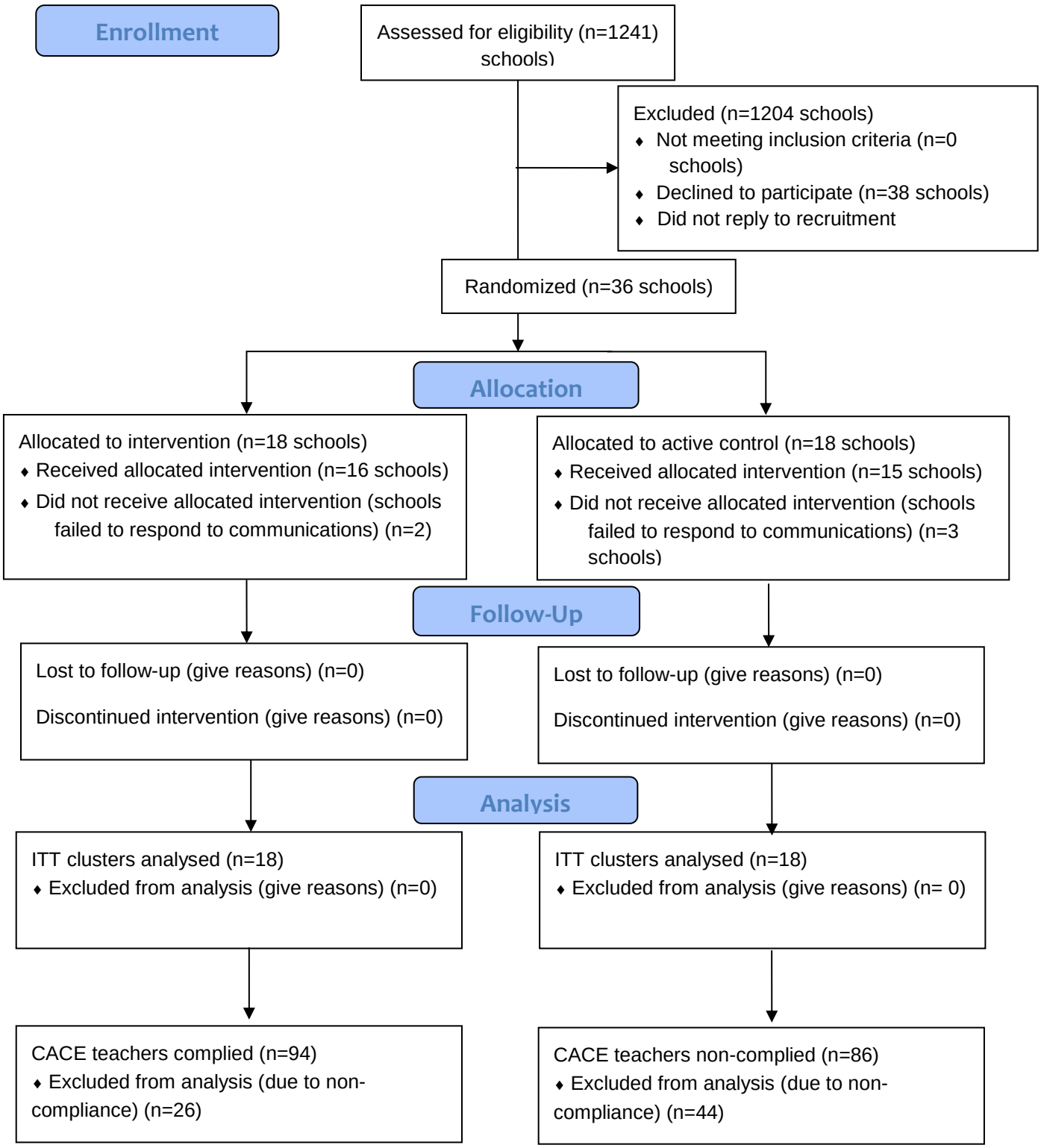
The 36 schools recruited for the study were stratified into five groups according to their median baseline scores (average GCSE points per school). There were eight schools in stratum 1 with median baseline score below 5.82 (a high C grade at GCSE), seven schools in stratum 2 with median baseline score between 5.82 and 6.16 (a low B grade at GCSE), seven schools in stratum 3 with median baseline scores between 6.16 and 6.67 (a mid B grade at GCSE), seven schools in stratum 4 with median baseline scores between 6.67 and 6.92 (a high B grade at GCSE) and seven schools in stratum 5 with median baseline score above 6.92. Schools in each stratum were randomised between intervention and control groups based on a permuted block randomisation scheme with fixed allocation ratios. Schools in stratum 1 were randomised with an allocation ratio of 1:1. Schools in stratum 2 were randomised with an allocation ratio of 1:2 in favour of the control group. Schools in stratum 3 were randomised with an allocation ratio of 2:1 in favour of the intervention group. Schools in stratum 4 were randomised with an allocation ratio of 1:2 in favour of the control group. Lastly, schools in stratum 5 were randomised with an allocation ratio of 2:1

in favour of the intervention group. The permuted block randomisation scheme was based on a combination of block sizes of three and four.

In summary, the randomisation results in 18 schools being randomised to the intervention group and 18 schools randomised to the control group. The intervention group consists of four schools from stratum 1, three schools from stratum 2, four schools from stratum 3, three schools from stratum 4 and four schools from stratum 5. The control group consists of four schools from stratum 1, four schools from stratum 2, three schools from stratum 3, four schools from stratum 4 and three schools from stratum 5. The randomisation scheme was implemented in R software.

By randomising all at once, the possibility of post-randomisation selection bias is controlled (Puffer, Torgerson, & Watson, 2003).

CONSORT 2010 Flow Diagram: SecondarySEEQ trial



3.8 Recruitment and enrolment

The eligibility for participation of schools (clusters) was assessed according to the following criteria:

- A cluster school must teach students with the aim of achieving Advanced Level and Advanced Subsidiary Level qualifications;
- A cluster school's students must have taken the CEM Advanced Level Information System (ALIS) assessment in either 2012 or 2011.

At the time the recruitment process began, 1241 institutions globally met the above criteria. Using the database of contact details held by CEM, the registered coordinator of CEM assessments in each school was emailed with details of the proposed trial; they were asked to respond via email or telephone asking for more details or offering participation in the trial.

Three emails were sent to school contacts registered on CEM's ALIS database: an initial email on 19/03/12, a follow-up email on 03/05/12 and a final email one week later. The data for emails opened are presented in Table 3.2 below.

Email name	Number sent	Number opened	% opened
SecondarySEEQ	1674	441	26.3
SecondarySEEQ follow-up	1674	312	18.6
SecondarySEEQ final	1674	333	19.8

Table 3.2: Recruitment email opening rates

The above does not represent a response rate for potential participants. It is an indication of the difficulty of communicating with potential research subjects. It is surprising to see such low rates of opening emails on three separate occasions, given that the email addresses used were those of the coordinators registered as points of contact for using the ALIS system, people who would be familiar with receiving communication from CEM and for whom there was already an investment in CEM systems. From the combined 21.5% of emails opened, 2.7% of schools eligible to participate (45) responded that they were keen to participate. Of these schools, 36 were randomised (2.1%) and 31 completed the trial (1.8%). It is proposed that changes in school coordinators responsible for the administration of CEM data, as well as schools discontinuing their contracts for assessments systems with CEM, are plausible reasons for emails not being opened in many of the cases identified above.

None of the above response rates is encouraging, but it should be noted that even to have 36 schools willing to participate came as a pleasant surprise. Given the nature of involvement in the trial - specifically with regard to the expectation placed on teachers to request feedback on their teaching from their students - it was heartening to receive a positive response of any degree.

The Education Endowment Foundation (EEF), whose name features very prominently in the field of education research in England, offers some interesting recommendations for recruitment to trials (Foundation, 2013). Analysis of their own projects found that

recruitment initiatives resulted in 33% of schools approached signing up to participate; the recruitment process for EEF projects took, on average, 15 weeks. The recommendation that is made in bold by the EEF is that three times as many schools as are required to generate the prescribed sample size should be approached. Interestingly, it was found that 23% of secondary schools approached agreed to participate, while the figure for primary schools stood markedly higher at 54%. Given that the trial designed for this PhD project was destined only to be populated by secondary schools with Year 12 and Year 13 students, it is useful to put the results of my recruitment efforts into this context. Nonetheless, it is perhaps an indication - and this is really only speculation - of the influence generated by a government-funded body using professional evaluators to work on their projects. The difference in perception between EEF evaluations and one conducted by a PhD student (albeit one with the very overt backing of the Centre for Evaluation and Monitoring at Durham University) is, potentially, a significant one.

The discussion of how schools were retained in the project is taken up in Chapter 6, as this is where the data collection in the main project is described in detail. Having established, however, how the sample size was calculated, and described the approach adopted to generate the sample, attention is now turned to measures adopted to control for bias.

3.9 Controlling for bias

In order to help provide unbiased estimates of the intervention's effects, CONSORT (2010) provides a set of standards for the reporting of trials; by adhering to these standards, possible threats to the internal validity of the trial are minimised: selection bias, in particular is minimised. Yet, even under the conditions of a correctly-randomised trial, the possibility still remains that the generated groups will differ by chance on a covariate of importance. In this trial, such 'chance bias' was minimised by stratifying clusters based on their average GCSE scores (see Table 3.3 below), an important predictor of the standardised value-added outcome.

Alongside the minimisation of chance bias, the following post-randomisation threats to the integrity of this trial are acknowledged and efforts made to mitigate them in the following ways:

Detection bias

The nature of the outcome measure (standardised value-added scores derived from A-Level outcome data) means that examiners responsible for marking examination papers were blind to the identity and cluster of each student, as well as whether or not their cluster school had been allocated to the treatment or control condition; this removed the possibility of detection bias.

Dilution bias

It was not possible to blind participants to the nature of the trial, but a waiting list design is a measure employed to reduce dilution bias; the hope was that, with this in mind, control group participant teachers would not seek alternative support for the interpretation of

their feedback charts, or for the improvement of their teaching in relation to the feedback given. Equally, because of the computer-delivered design of the information component of the intervention (the feedback), it is possible to ascertain which participants in the treatment group actually received the intervention, and to use an instrumental variable approach (CACE analysis) (Angrist, Imbens, & Rubin, 1996) to estimate the causal effect of treatment receipt.

Technical bias

The *R* software programme used for the randomisation procedure was operated by a professional statistician highly familiar with the software, and the resulting stratification was checked by me against the original baseline data to minimise the risk of a software problem creating a faulty allocation schedule.

By following the CONSORT statement and its extension to cluster trials, the reporting of this trial is designed to be as robust as possible; the potential for chance bias (at randomisation) and post-randomisation biases is acknowledged and is the reason for the inclusion of the CONSORT (2010) diagram, as well as the checklist provided earlier in this chapter. Furthermore, details of measures taken to address the potential for a range of biases are given later in this chapter.

Selection bias

As a key threat to the internal validity of the experiment, careful attention was paid to the need to control for selection bias. By using an appropriate randomisation protocol (Appendix 1a), selection bias was removed from this trial.

Subversion bias

Concealment of cluster allocation to either active control or intervention arms was an important feature of the process, in order to ensure that subversion of the allocation was eliminated as far as possible. Allocation in this trial was at the cluster (school) level, as detailed in the Randomisation Protocol (Appendix 1a). Viewed as the key foundation for avoiding bias in randomised trials (Eldridge & Kerry, 2012), the allocation procedure was completed by a professional statistician whose physical location was removed from mine, and the relationship with whom was sufficiently detached to accord an appropriate degree of separation. Dr Kasim (the statistician) was given a spreadsheet containing a numerical identifier for each cluster and the average GCSE points score for each cluster. The average GCSE points score for each cluster was calculated as the average of all students' average points score in the 2012 GCSE cohort.

Table 3.3 below shows the allocation of each cluster (school)

School	Stratum	Block size	Block id	Group
1	6.67<x<=6.92	4	5	Intervention
2	5.82<x<=6.16	4	2	Control
3	5.82<x<=6.16	4	2	Intervention
4	<=5.82	4	4	Intervention
5	6.67<x<=6.92	4	5	Intervention
6	6.67<x<=6.92	4	5	Control
7	> 6.92	4	6	Control
8	6.16<x<=6.67	4	2	Control
9	6.16<x<=6.67	4	2	Intervention
10	> 6.92	4	6	Control
11	6.67<x<=6.92	4	5	Control
12	6.16<x<=6.67	4	2	Intervention
13	5.82<x<=6.16	4	2	Control
14	<=5.82	4	4	Control
15	> 6.92	4	6	Intervention
16	<=5.82	4	4	Control
17	5.82<x<=6.16	4	2	Intervention
18	> 6.92	4	6	Intervention
19	<=5.82	4	4	Intervention
20	<=5.82	4	6	Intervention
21	6.16<x<=6.67	4	2	Control
22	<=5.82	4	6	Intervention
23	5.82<x<=6.16	3	3	Control
24	6.67<x<=6.92	3	3	Control
25	> 6.92	3	3	Intervention
26	6.16<x<=6.67	3	3	Intervention
27	> 6.92	3	3	Intervention
28	6.16<x<=6.67	3	3	Intervention
29	6.67<x<=6.92	3	3	Control
30	6.16<x<=6.67	3	3	Control
31	5.82<x<=6.16	3	3	Control
32	5.82<x<=6.16	3	3	Intervention
33	<=5.82	4	6	Control
34	6.67<x<=6.92	3	3	Intervention
35	> 6.92	3	3	Control
36	<=5.82	4	6	Control

Table 3.3: Allocation sequence

To protect the integrity of the allocation process, no discussion was held with Dr Kasim regarding the trial's aims, nor was any other information given regarding the identity of the

clusters. These decisions were made in light of descriptions of incidents in which clinical researchers have subverted random allocations by x-raying sealed envelopes and breaking open filing cabinets to access randomisation sequences (Schulz, 1995). While subversion was not considered a serious risk in this trial, attempts were made to adopt good practice nonetheless. Moreover, Boruch (1997) describes an American trial in which police officers subverted a trial of an intervention in which offenders were randomised to either a caution handed out by an officer, or taken to the police station. Under these circumstances, there is no question that such subversion will render the trial's findings meaningless; in the design of my study, it was imperative to ensure that correct procedures were in place to control for subversion bias.

Attrition bias

It is generally accepted that most trials will lose participants after randomisation has taken place. If the rate of attrition differs between the intervention and control arms, the bias introduced may be pronounced. One of the ways in which this trial set out to minimise attrition was through the use of a waiting-list condition for those schools allocated to the control arm. Additionally, Intention To Treat (ITT) analysis was programmed into the design to avoid some of the problems associated with attrition bias. Irrespective of their remaining in the trial or not, both pre-test (average GCSE scores) and post-test (average A-Level scores) could be collected for all clusters randomised, given that they would submit these routinely as a feature of their participation in the ALIS program.

Ascertainment, reporting or detection bias

The chief outcome for the trial is the average standardised value-added (STDVA) scores for each cluster. Trials often suffer from bias in the reporting of outcomes when there is uncertainty regarding their 'objectivity' (Torgerson & Torgerson, 2008), so the selection of a suitable outcome measure for this study was paramount at the design stage. One aspect of ascertainment bias which is negated by the use of STDVA scores is that none of the markers of A Level examinations knows whether the student whose work is being graded is nested within a control or an intervention school; also, the STDVA variable controls for baseline differences in student ability. Moreover, the computerised generation of the STDVA measure is viewed here as having reduced risk of bias due to the lack of human intervention in its creation.

Performance bias and resentful demoralisation

In order to recruit and retain teachers within schools in the trial, the control arm clusters received an active control intervention (detailed in Chapter 4). There was no feasible way that schools could have been recruited to a trial incorporating student evaluation of teaching ratings without offering some indication of the responses provided by students. The practice of engaging students in sharing their opinions on aspects of their education is one that has been poorly managed, and its functions and limitations poorly understood (BBC, 2010), so there was a perfectly reasonable reticence on the part of some schools and teachers to be involved in what was perceived by some as a 'student voice' activity that was devoid of the echoes of those voices in teachers' ears. Thereby, the active control condition of receiving feedback alone was implemented. It is important to recognise here that there are, potentially, serious implications associated with a small minimum

detectable effect size (MDES) and an active control condition. By presenting feedback to the control arm teachers, it is plausible to believe there may be an effect on their students' STDVA scores and, therefore, consideration must be given to the impact this has on the ability of the measure to detect a real effect of the intervention under investigation.

Teachers in control schools understood that they were in the control arm, though they were not told specifically what this meant in relation to teachers in the intervention arm. At no point was it revealed to one group the precise nature of the others' intervention. While this could not be defined as blinding, by any stretch of the imagination, it was the most pragmatic way in which to describe the trial to schools and to reassure them about the process of participation.

There has been debate over whether or not performance bias is, in fact, a bias at all (Torgerson & Torgerson, 2008). One could imagine that teachers in the intervention arm might 'perform' better than those in the control arm because they like the fact of the allocation more than those in the parallel arm. It could be argued that, if they like the intervention and their subsequent performance in the classroom - informed by the student evaluation of teaching data - has a positive impact on the outcome measure, this is part of the intervention.

In order to allay the concerns of resentful demoralisation amongst trial participants in the control arm, a waiting-list design was adopted. Teachers who registered for the trial were informed prior to randomisation that those allocated to the control arm would be offered the intervention package in the 2013-2014 academic year, subject to evidence of a positive impact. Due to the nature of data collection and analysis of the outcome measure, the impact of the intervention on student attainment was not fully known until late in 2013.

Consequently, the intervention package was offered to control arm schools with the caveat that, at the beginning of the 2013-2014 academic year, no conclusion could be drawn, and that schools adopted the intervention having been fully appraised of this fact.

One concern regarding resentful demoralisation that was considered in this trial was that teachers in control schools may not demonstrate the same attitude *to* their students towards being rated *by* their students as those in the intervention arm schools, thus leading to students reporting in a manner that may bias their ratings and, consequently, teachers responding to their feedback-only condition in a manner informed by their feelings towards being randomised to the control arm. Again, the use of the waiting-list design was intended to combat this, as was the information given to teachers informing them of the need for a control group to ensure that the research findings were as robust as possible. Nevertheless, such bias may occur regardless of the attempts to control it. In hindsight, and in an ideal world, a 'preference design' - in which schools whose teachers demonstrated indifference to their allocation condition would be allocated - could have been used. This would, however, have been something of a utopic approach and not one for which there was significant pragmatic support.

Hawthorne effects

Studies into industrial productivity at the Western Electrical Company's Hawthorne Works in Chicago gave a name to the effect of being made to feel singled out or important by being studied: the Hawthorne effect (Mayo, 2004). This effect, while identified as a possible source of bias, has not been studied extensively (McCarney et al., 2007), so drawing conclusions about its impact in the current study is a challenging activity.

Participation in a trial of a consultative feedback intervention's impact on student attainment at A Level may well be prone to Hawthorne effect bias. This is, perhaps, more of a concern in trials that have interventions which require a lot of time and attention to be given to those not allocated to control conditions - simply because it is the very time and attention given that may be having an effect, rather than the intervention itself.

To control for the potential of Hawthorne effect bias, the intervention and control conditions were designed to be implemented without significant input from me or other external actors. The intervention (described in Chapter 6 in detail) has as a key feature the time and attention of a colleague in school (as part of the peer-coached consultation meeting), but this is a defined part of the intervention, rather than an addition to it. Had I become more involved in supporting individual teachers in this process, I would have altered the composition of the intervention and biased the trial's results.

3.10 Analysis

The primary analysis of the cluster randomised controlled trial undertaken in this study is Intention To Treat (ITT) analysis (cases will be analysed as they were randomised); this type of analysis offers an unbiased indication of the effect of offering the SecondarySEEQ programme to teachers, but generally provides a conservative estimate of impact. To provide deeper insight into the effect of the intervention, Complier Average Causal Effect (CACE) analysis (Gerber & Green, 2012) is used; this form of analysis builds on statistical work by Angrist et al. (1996) to model the effect of the intervention for those participants who comply with their allocated treatment (conditions for compliance are specified in Chapter 7). The CACE estimate will be calculated using the following calculation:

ITT/proportion of compliers

This calculation assumes that there is evidence of compliance amongst participants, that the randomisation process was truly random, that outcomes cannot be influenced by the outcomes or assignment of others (the so-called stable unit of treatment value assumption, or SUTVA), that the effect of the treatment works through (is mediated by) compliance with assigned treatment (the so-called 'exclusion restriction'), and that there are no participants who do the opposite of their assigned treatment condition (the monotonicity assumtion).

Finally, analysis of the multilevel structure of the SET data will begin with exploratory structural equation modelling approaches but will, ultimately, adopt a confirmatory factor analytic approach, due to the a priori foundations of the study (Asparouhov & Muthén, 2009; Marsh et al., 2009).

Chapter 4: The feasibility study

4.1 Introduction

Having provided an overview of the investigation as a whole, this chapter serves to illustrate the design and implementation of the feasibility project. This project aimed to determine the extent to which collecting student evaluations of teaching in secondary schools was a feasible proposition; it also looked at the reactions of teachers and students to the process of rating teaching (in the case of the former) and having teaching rated (in the case of the latter). Finally, the project sought to develop and trial a framework for a peer-coached teacher consultation intervention, in which I played the part of the peer-coach.

It is worth noting that this project was conducted in a collaborative and exploratory manner, working with the teachers and students involved to develop both the instrument and the feedback intervention.

4.2 Pilot study or feasibility study?

> "A pilot study can be defined as a small scale study conducted in advance of a main research study with the aim of refining or improving the design of the main study or assessing its feasibility." (Eldridge & Kerry, 2012), p.93)

While the authors above speak of a pilot study as something which sets out to determine the feasibility of a design, this is problematic in itself and requires clarity for the purposes of this study. In making a distinction between 'pilot' studies and 'feasibility' studies, Arain,

Campbell, Cooper, and Lancaster (2010) point out that their findings suggest those with the latter label operate with greater methodological flexibility than those termed 'pilots'. They note that there appears to be greater methodological rigour in pilot studies, citing sample size estimation, randomisation and control group allocation as features. In seeking to address the apparent 'grey area' between pilot and feasibility studies, the authors suggest the definition provided by the NHS National Institute for Health Research[42] (NIHR) is the most helpful: "pieces of research done before a main study in order to answer the question 'Can this study be done?'". The definition continues with a list of key features of a feasibility study: the standard deviation of the outcome measure, willingness of participants to be randomised, availability of data, time needed for data collection and analysis.

Because this PhD study used its feasibility study to determine broadly the willingness of participants to be randomised in a trial, their thoughts about being involved in a student evaluation of teaching trial, and also sought the guidance of its participants in shaping the instrument to be used in the main study, the nomenclature ('feasibility study' rather than 'pilot study') is appropriate. Furthermore, it truly set out to assess whether or not engaging secondary school teachers in a UK school in the process of having their teaching rated by their students was a feasible, practicable proposition, both from their point of view as busy professional teachers, and from my point of view as a researcher attempting to devise a scalable design.

4.3 Feasibility study design considerations

[42] http://www.nets.nihr.ac.uk/glossary?result_1655_result_page=F accessed 20/6/14

My feasibility study did not use randomisation as there was no control or comparison condition (again, supporting the semantic decision discussed above), and did not fall into the trap of formal hypothesis testing that many pilot or feasibility studies seem to (Leon, Davis, & Kraemer, 2011). Furthermore, the sample size was based on what was pragmatic for both the school and for me, chiefly in terms of the time available for the former to devote to the study; Leon et al. (2011) suggest that this is an entirely defensible approach. Nonetheless, it must be noted that there were significant limitations to the feasibility study: it failed to address the need to understand the willingness of participants to be randomised; also, it did not go beyond the bounds of a single school and, as such, did not address the need to understand the willingness of school leaders more broadly to allow their schools to participate in the main trial.

Therefore, a small-scale feasibility study of a student evaluation of teaching instrument and associated consultative feedback intervention was conducted, involving ten A Level teachers of eight subjects, and 157 Year 12 and Year 13 students at a large urban comprehensive school in County Durham.

4.4 A note on sample size

While some argue that using pilot studies to estimate effect sizes for the purpose of informing subsequent larger-scale hypothesis-testing studies is an appropriate course of action, Kraemer, Mintz, Noda, Tinklenberg, and Yesavge (2006) argue that this approach is inappropriate; indeed Leon et al. (2011) state plainly that the results of pilot or feasibility studies should not be used because of the problems that stem from drawing such conclusions from small samples. It is clear that, with small samples being used as the

foundation for effect size calculations in subsequent main studies, the risks of false positive (Type I error) and false negative (Type II error) results emerging is high.

4.5 Purpose

Defining the purpose of a pilot or feasibility study is recommended in the literature (Arain et al., 2010; Arnold et al., 2009; Eldridge & Kerry, 2012; Leon et al., 2011). Consequently, the feasibility study designed for this PhD set out to achieve four aims:

1. An exploration of the feasibility of A Level students evaluating their teachers, and an investigation of the students' attitudes towards the evaluation process;
2. Gathering teachers' opinions on the suitability of the instrument and its items for evaluating teaching at A Level, as well as their perceptions about being evaluated by their students;
3. Trialling the online survey software (Cvent);
4. Trialling a consultative feedback intervention session.

In setting these goals, the study retained the kind of flexibility required for it to be a learning experience that would facilitate an iterative, developmental process, with the participating teachers informing that development. No pretence is made that the more rigorous methodological considerations of randomisation and sample size estimation were factors in this study.

4.6 Overview of the feasibility study

The sample of ten teachers in the feasibility study was recruited as a result of a pre-existing connection between the Director of CEM and one of the Assistant Heads at the feasibility study school. The school has worked with CEM in the past on research projects, and the leadership team was amenable to collaborating on this project. The convenience sample that emerged was suited to the needs of a feasibility study, insofar as the teachers therein taught a range of subjects (the core subjects of English, Maths and the Sciences were represented), and had a broad range of teaching experiences and number of years in the classroom. While there is no way that they could be considered a representative sample, it was encouraging at least to have participating teachers from a variety of subject areas.

There were five stages to the feasibility study fieldwork:

1. Focus group with teachers (8/2/12)

This session introduced the feasibility project to the volunteer teachers. During the meeting, the teachers read through the instrument and gave first impressions orally. They also took copies of the instrument away with them and were asked to email further thoughts on the suitability of specific items. Finally, teachers were asked to give their views on the prospect of having their teaching evaluated by their students.

2. Adaptation of items (9/2/12 - 16/3/12)

In light of discussion with teachers and students, the items from Marsh's original Student Evaluation of Educational Quality instrument (Marsh, 1982) were adapted to be more appropriate to A Level students.

3. Exploratory interviews with students (13/3/12)

A group semi-structured interview with three Year 12 and three Year 13 students was conducted in the feasibility study school. This exploration focused on interpretation and

comprehension of the survey items, and on attitudes of students towards evaluating their teachers.

4. Online student evaluation of teaching survey (22/3/12 - 28/3/12)

Using Cvent (a commercial online survey package), 157 students evaluated the teaching they had received during the Easter Term 2012.

5. Feedback

Once the student surveys had been downloaded, the individual student data were aggregated to the class level and displayed in the form of a graph. These formed the central artefact in the consultative feedback discussion held with each teacher.

The following sections detail the five stages of the feasibility study project set out above: the initial focus group with teachers; the adaptation of items from Herb Marsh's original Student Evaluation of Educational Quality (Marsh, 1982) instrument; exploratory interviews with students; the online survey taken by students in the feasibility study school; feedback. Analysis of the data is presented in Chapter 6.

Ethical approval for the feasibility study was received (by email) from Durham University's School of Education Ethics Committee on 13/12/11 (Appendix 3a).

4.7 Focus group with teachers

The chief functions of the teacher focus group discussion were to introduce the feasibility study project to the interested staff and to gather feedback on the instrument. It was of interest to me to accrue teachers' feedback on the following areas:

- perceptions of being evaluated (or 'rated') by their students;

- perceptions of the face validity of the proposed SET (SEEQ) items;
- perceptions of the usefulness of the proposed feedback materials.

As one aim of this study is to validate the adapted student rating instrument, the initial responses of teachers heralded the start of the validation process. Equally, a collaborative, iterative approach to the development of the instrument and feedback materials and processes was important in securing the trust of the teachers involved; my own background in teaching and school management helped with this.

To be skilled in the facilitation of group discussions, the management of conflict and in enabling non-judgemental interaction between focus group participants is an important trait for the moderator or researcher responsible for a focus group (Halcomb, Gholizadeh, DiGiacomo, Phillips, & Davidson, 2007; Toner, 2009). While I received no formal training in the conduct of focus groups, my reading prior to the feasibility study stage focus group meetings indicated that a structured and defined, yet sensitive and flexible approach was needed from me in my role as discussion facilitator.

4.8 Focus group composition

The size of the focus group was determined by the number of teachers interested in participation in the project. Though aware of the suggestions of ideal group size (between four and twelve participants) present in the literature (Cousin, 2009; Hopkins, 2007; Vaughn, Schumm, & Sinagub, 1996), I had no way to control the composition of this group, nor did I wish to. Indeed, allowing those who wished to attend was crucial in presenting a

very approachable demeanour for the study, given the nature of the task I wished teachers ultimately to undertake.

4.9 Institutional support

Prior to the teacher focus group meeting, the Assistant Head at the school had emailed colleagues to alert them to the event, and had indicated that there was institutional support for it. The fact that there was a clear and positive statement from the management of the school regarding the project later became a key element in the recruitment phase of the main trial. In later stages of the project, such support from the leaders of schools became crucial both in ensuring that time and resources were secured, as well as in impressing on teachers the perception of value in the activities involved in participation.

The term 'focus group' is used here as it has widely become adopted in education to refer to research in a group setting (Gibbs, 2012). The process of discussing in the teacher focus group was very much an iterative one, in which opinion was gauged and the interpretation of the language of survey items engaged in. It was during this phase of the project that some teachers expressed concern regarding the comparative items (37 and 38 in the list below). There was a feeling that comparing one's own performance with colleagues was not of interest, given the perceived differences in teaching different subjects; it was felt that being rated in this way was not productive, and had echoes of a divisive ranking system. The Breadth of Coverage items (28 - 31) were perceived by one teacher as being redundant in some subject areas; the teacher thought that these items may also be difficult for students to comprehend as their meaning was intended, something that I followed up in the exploratory interviews with students.

4.10 Adaptation of items

The original instrument used as the basis for the feasibility study project and subsequent main trial was Marsh's Student Evaluation of Educational Quality survey (Marsh, 1982) (Appendix 2a), an instrument which has had significant usage in higher education contexts for more than 30 years (Marsh, 2007c). Marsh concluded that, when measured appropriately, student evaluations of teaching are:

- Useful in improving the effectiveness of teaching when they are associated with suitable consultative strategies;
- Reliable and stable;
- Multidimensional
- A function more of the teacher than of the subject or course that they are teaching;
- Relatively valid when compared with other indicators of effective teaching;
- Relatively unaffected by variables suggested as potential biases (class size, workload, prior interest in the subject, leniency of grading).

The original Student Evaluation of Education Quality (SEEQ) instrument which was adapted for use with Key Stage 5 students in this project is found in Appendix 2a. The 39 items used in the feasibility study are listed below:

**1. What is your first name?(*Required)*

**2. What is your last name?(*Required)*

**3. Are you male or female?(*Required)*

**4. Which year group are you in?(*Required)*

**5. Select the surname of the teacher whose lessons you are going to comment on.*

*(*Required)*

**6. Which subject do they teach you?(*Required)*

**7. What is the name of class / set?(*Required)*

8. During the Spring Term (January - end of March this year), I found this teacher's lessons intellectually challenging and stimulating.

9. During the Spring Term (January - end of March this year), I learned some valuable things in this teacher's class.

10. My interest in this subject increased during the Spring Term (January - end of March this year) as a result of this teacher's lessons.

11. I learned and understood what my teacher has taught me during the Spring Term (January - end of March this year).

12. During the Spring term (January - end of March this year), my teacher was enthusiastic about teaching my class.

13. During the Spring Term (January - end of March this year), my teacher was dynamic and energetic when teaching my class.

14. My teacher enhanced their teaching with the use of humour during the Spring Term (January - end of March this year).

15. My teacher's style of teaching held my interest during lessons during the Spring Term (January - end of March this year).

16. My teacher's explanations were clear in my lessons during the Spring Term (January - end of March this year).

17. During the Spring Term (January - end of March this year), the resources and other materials my teacher gave the class were well-prepared and carefully explained.

18. During the Spring Term (January - end of March this year), I knew where lessons were going because my teacher explained the lesson objectives clearly, then taught in a way that helped me to achieve them.

19. My teacher's lessons enabled me to take notes during the Spring Term (January - end of March this year).

20. During the Spring Term (January - end of March this year), my teacher encouraged students to take part

in class discussions.

21. My teacher invited students to share their own ideas and knowledge in my class during the Spring Term (January - end of March this year).

22. During the Spring Term (January - end of March this year), I was encouraged to ask questions and my teacher gave me meaningful answers.

23. I was encouraged to express my own ideas and / or question my teacher during the Spring Term (January - end of March this year).

24. During the Spring Term (January - end of March this year), my teacher was friendly towards individual students in my class.

25. During the Spring Term (January - end of March this year), my teacher had a genuine interest in the individual students in my class.

26. My teacher made students feel welcome in seeking help / advice in or outside of lessons during the Spring Term (January - end of March this year).

27. My teacher was adequately accessible to students outside of lessons during the Spring Term (January - end of March this year).

28. My teacher contrasted the implications of various theories during the Spring Term (January - end of March this year).

29. During the Spring Term (January - end of March this year), my teacher presented the background or origin of ideas / concepts developed in class.

30. During the Spring Term (January - end of March this year), my teacher presented points of view other than his / her own when appropriate.

31. During the Spring Term (January - end of March this year), my teacher adequately discussed current developments in the subject.

32. The feedback my teacher gave on my assessed work during the Spring Term (January - end of March this year) was valuable.

33. The ways in which my teacher assessed my work during the Spring Term (January - end of March this year) were fair and appropriate.

34. The assessed work and examinations I did during the Spring Term (January - end of March this year) tested what I was taught in lessons.

35. Homework set by this teacher during the Spring Term (January - end of March this year) was valuable.

36. During the Spring Term (January - end of March this year), the homework set by this teacher contributed to my appreciation and understanding of the topic(s) taught.

37. Overall, how does this teacher's lessons compare with other teachers' lessons in this school / college?

38. Overall, how does this teacher compare with other teachers in this school / college?

39. If there is any other CONSTRUCTIVE FEEDBACK you would like to give your teacher about their teaching last term, please do so here.

Items 1 - 6 were designed to capture information that would enable individual students' evaluation data to be aggregated to the class level, while still allowing me to identify the individual - an important feature to be trialled, as matching against ALIS data and A Level outcome data would be crucial to the main trial.

The remaining items underwent only minor changes from the original SEEQ statements. Wording to indicate a specific time period ("January - end of March this year") was included to focus students' attention, and possessive pronouns were included to reinforce the notion that the rating was being given to and for a specific teacher's teaching. Essentially, though, there were few changes made to the original SEEQ instrument's items, and this remained so in the main trial, with the notable exception of the comparator items towards the end (items 37 and 38 in the list above).

Work on Marsh's SEEQ demonstrated the multidimensionality of teaching through the presence of nine teaching factors:

1. Learning and academic value
2. Teacher enthusiasm
3. Organisation and clarity
4. Group interaction
5. Individual rapport
6. Breadth of coverage
7. Exams and grading
8. Assignments
9. Workload and difficulty

Factor analyses (Marsh & Roche, 1993) have demonstrated that the grouping of items into these categories best accounts for the variation in student responses and teacher self-evaluation responses.

Analysis of the data collected from the online survey could only ever be interpreted within the limitations conferred by the design of the feasibility study (non-randomised, small convenience sample). Nonetheless, the structure of the data should be explored and its reliability examined; description and discussion of the factor analytic results and reliability statistics is found in Chapter 7.

4.11 Exploratory interviews with students

The transcript of the exploratory interview held with the three Year 12 and three Year 13 students in the feasibility study school is found in Appendix 3b. The purpose of this group interview was to gather data on the perceived meanings of the items proposed for inclusion

in the feasibility study instrument. Five students were asked by the Assistant Head of the feasibility study school to take part in the activity; I was the only other person in the room for the duration of the interview, which was recorded and later transcribed (Appendix 3b).

The literature suggests that a group of between four and twelve people is an appropriate composition of a focus group (Hopkins, 2007) so my group composition of six is defensible; in reality, gathering more students together from the feasibility study school would have been difficult, due to the constraints of teaching time. In no way could this sample be considered representative of the target student population; rather, it was a sample of convenience formed more by the exigencies of the school day rather than the pure needs of a research project. Given the purpose of the feasibility study, however, the derivation of the sample is, again, unproblematic. Nonetheless, the dialogue that came out of the meeting offered rich insights from a small number of students, all of whom were later involved in the evaluation of their teachers' teaching.

4.12 Online student evaluation of teaching survey

The purpose of the instrument was to capture students' subjective feelings about the quality of teaching they had received from a specific teacher. The items make up 9 factors (above) identified by Marsh and Roche (Marsh & Roche, 1993), with each factor comprising between two and four items. Once collected, student evaluation of teaching data from each student in a single class were aggregated to the class level using the following process:

- Each student's rating for each item on each factor was identified and an average factor rating for each student calculated simply by taking the mean of these numbers;
- Each student's mean rating for each factor was added to those of other students in the same class and a mean for each factor at this class level calculated.

The nine factors on the SecondarySEEQ scale were kept from Marsh's original instrument both at his request and to align with *a priori* structures, as well as with a keen eye on pragmatism. While Marsh suggested that there may be benefit gained from trialling additional subscales and items, the restrictions of a single novice researcher only working on the project meant that only Marsh's original items were used. To write new items, and then trial these, would not have been feasible. Instead, focus was placed on the adaptation of items and their ecological validity in the novel context they found themselves.

4.13 Instrument trialling

The multidimensional nature of teaching is discussed in the sizeable body of literature on SETs (Marsh, 1987). One thing that becomes clear in the literature is that inappropriate or poorly-constructed items will generate poor-quality data: asking appropriate questions in suitable ways is crucial to the collection of useful data. The widespread use of Marsh's SEEQ in higher education contexts brought a degree of confidence regarding the suitability and quality of the items used to this researcher. It should be noted, however, that, in an email to me dated 24/11/11, Marsh stated that he would have included more items in his instrument, had this been possible at the time of his initial work. With this in mind, the decision not to trial new items was a difficult one to take but one that was justified by the

need to develop a scale for use with younger students (16 - 18 years old), for whom a longer instrument may have led to fatigue or resistance.

Table 4.1 below shows summary statistics for those students and teachers involved in the feasibility study.

Teachers	10
Students	161
Average survey response time	6 minutes, 53 seconds
Year 12 students (%)	57%
Year 13 students (%)	43%

Table 4.1: Summary statistics for students involved in the feasibility study

Table 4.2 below shows the number of survey responses by subject:

Biology	30
Chemistry	70
English	60
ICT	30
Languages	14
Maths	22
Physics	1

Table 4.2: Numbers of feasibility study survey responses by subject

The student survey was administered online using a commercially-available software package called Cvent[43] and the URL sent to the Assistant Head at the feasibility study school, who acted as the coordinator.

4.14 Feedback

The value of feedback has been well-documented (Askew, 2000; P. A. Cohen, 1980; Gregory et al., 2011; Hattie & Timperley, 2007; Kluger & DeNisi, 1996); work has also been done in higher education contexts to support the interpretation and understanding of student evaluation of teaching ratings as a means of generating improvement in the effectiveness of teaching (Dominick et al., 1997; Marsh & Roche, 1993; Penny & Coe, 2004; Rindermann, Kohler, & Meisenberg, 2007). The mechanism for feeding back evaluation information to the ten teachers involved in this feasibility study project was a one-to-one meeting with each teacher, in which I presented a class-level graph of the student evaluation of teaching data, an example of which is seen below in Figure 4.1 (the complete feedback sheet is found in Appendix 3c). In trialling this approach, I wanted to assume the role of 'trusted colleague' which would be so integral in the main study. While, obviously, I was not actually employed by the school, and thereby not a 'colleague' in the pure sense of the word, my background in teaching enabled me to approach this task appropriately.

[43] Cvent: www.cvent.com

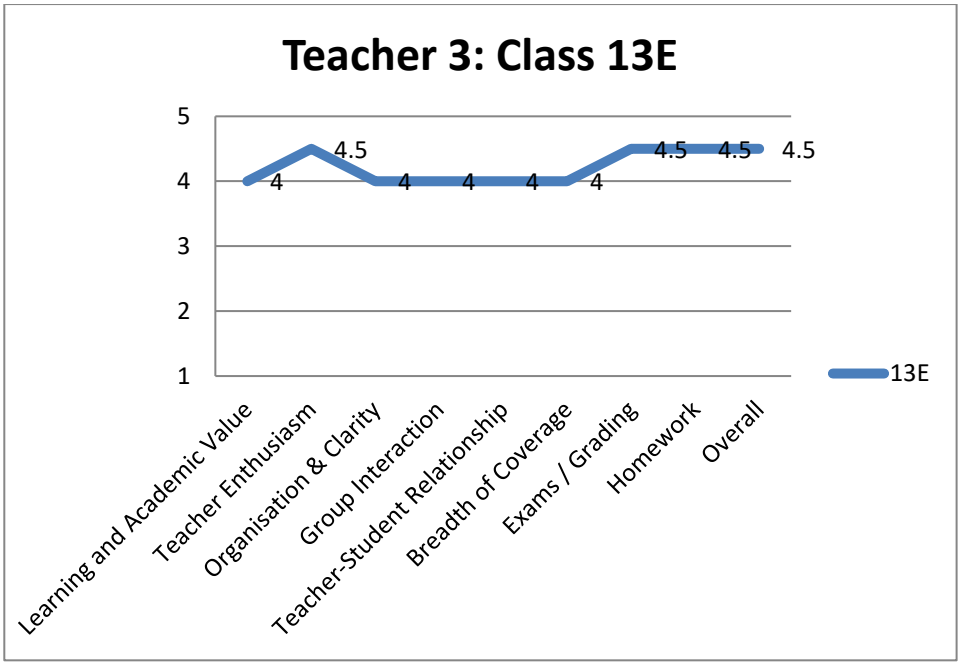


Figure 4.1: Example of a feasibility study feedback graph

Where teachers had multiple classes involved in the feasibility study, a single graph showing data for all classes was presented, as illustrated in Figure 4.2 below.

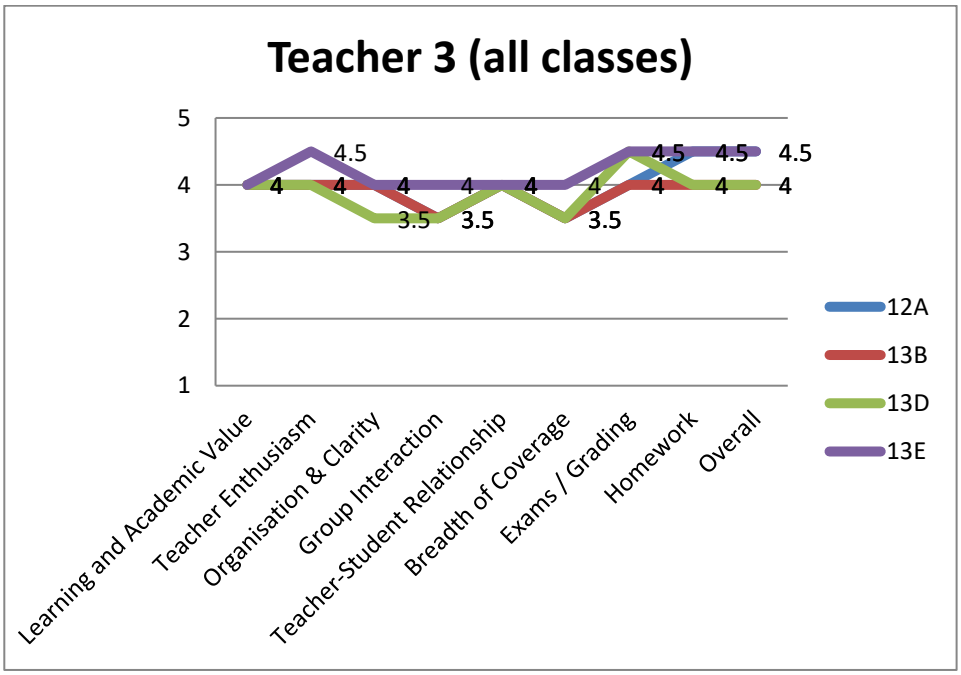


Figure 4.2: Example of a feasibility study feedback graph showing all classes for a single teacher

Each meeting was scheduled to last one hour and followed the following format:

1. How did you feel about having your students evaluate your teaching?
2. What do you think the data presented on your graph mean?
3. How do you feel about the evaluation data your class has given you?
4. What do you think these data mean?
5. What might you do with these data?

Comments that had been made by students during the survey were also presented to the teachers without editing. Each meeting was recorded and analysis of the transcript data is presented in Chapter 5.

The consultation meeting was designed around goal theory (Locke & Latham, 2002) and solution-focused brief therapy (SfBT) (Gingerich & Eisengart, 2000), greater discussion of which is presented in Chapter 6.

Chapter 5: Analysis of the feasibility study findings

The feasibility project set out to gather data and present both analysis and interpretation of them in accord with the following objectives:

1. An exploration of the feasibility of A Level students evaluating their teachers, and an investigation of the students' attitudes towards the evaluation process;
2. Gathering teachers' opinions on the suitability of the instrument and its items for evaluating teaching at A Level, as well as their perceptions about being evaluated by their students;
3. Trialling the online survey package (Cvent);
4. Trialling a consultative feedback intervention session.

This chapter provides analysis and interpretation of the data collected, and provides discussion of them in light of the above aims. Furthermore, as this study adopted ostensibly a confirmatory approach, reliability data in the form of Cronbach's alpha is presented below both for the instrument as a whole and for each subscale, concluding as it did that no item or scale should be dropped from the trialled version of the instrument when it came to the larger-scale main project trial.

5.1 Descriptive statistics

161 students and ten teachers were involved in the feasibility study. Of these, 161 completed the survey instrument online, which created 214 usable cases (some students

failed to respond to all of the items). 36% of students were male and 64% were female. 57% were studying in Year 12 and 43% were studying in Year 13. The ten subjects they studied and teachers of which received ratings from them are listed below:

Art	Biology
Chemistry	Drama
English	Ethics & Philosophy
History	ICT
Languages	Physics

Table 5. 1: List of subjects involved in the feasibility study

The average time taken to complete the survey fully was 6 minutes and 53 seconds; the shortest completion was in 2 minutes and 53 seconds, while the longest respondent took 13 minutes and 15 seconds.

214 usable cases were derived from those collected during the feasibility study. One of the chief interests of the analysis of these data at this early stage was their internal consistency, as measured by Cronbach's alpha (although it is recognised here – and discussed later in this thesis – that there are problems with the use of Cronbach's alpha for such purposes as drawing conclusions about internal consistency). As such, below are presented reliability statistics for the instrument as a whole[44].

[44] Reliability statistics are presented for the scale as a whole, though it should be noted that this does not imply that a unidimensional construct is under investigation. The construct 'effective teaching' is measured

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.953	.953	31

Table 5.2: Reliability statistics

The data in Table 5.2 above suggest that, as a whole scale, there is good internal consistency to the data.

Alongside looking at the reliability of the data using Cronbach's alpha, a brief exploration of the same feasibility study data set suggested that the sample size was adequate for the purpose of factor analysis, though this was not a planned analysis for this part of the study. Hutcheson and Sofroniou (1999) suggest that sampling adequacy measured on the Kaiser-Meyer-Olkin (KMO) test may be considered 'superb' if they are greater than .9 and in the case of the feasibility study data, the value is .921, as shown below:

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.921
Bartlett's Test of Sphericity	Approx. Chi-Square	4332.479
	df	465
	Sig.	.000

by this instrument multidimensionally, but for the purpose of initial internal consistency testing, this statistic suffices.

The KMO measure indicates that the sample is, in fact, large enough for Principal Component Analysis (PCA). When looking along the diagonal of the anti-image correlation matrix, we can see that no single variable has a Measure of Sampling Adequacy (MSA) lower than .768, again suggesting an adequate sample size.

Feasibility study objective 1: An exploration of the feasibility of A Level students evaluating their teachers, and an investigation of the students' attitudes towards the evaluation process.

Very simply, one of the main foci for this part of the study was to ascertain whether or not it was feasible for Year 12 and Year 13 students to provide student evaluation of teaching (SET) data. In essence, this was a matter of pragmatics and acceptability: could the data be collected without too great a disruption to the school day and to the classes of those whose students were providing evaluation data? And did students and teachers deem the activity sufficiently appropriate and acceptable for them to participate in the spirit of the project?

5.2 Pragmatics

The practical considerations of collecting SET data for the feasibility study were twofold: the willingness and consent of a school and volunteer teachers to have their teaching evaluated by their students; and the creation of a system to collect the SET easily and confidentially.

The school with which I worked on the feasibility study is a high-performing comprehensive school, and there were well-established relationships between teaching staff there and

University staff with whom I worked. The leadership of the school was keen to participate in the trial.

Feasibility study objective 2: to gather teachers' opinions on the suitability of the instrument and its items for evaluating teaching at A Level, as well as their perceptions about being evaluated by their students.

5.3 Acceptability

To ascertain the perceptions teachers had of their teaching being evaluated, and to understand more about the experience from a teacher's point of view, seven teachers from the feasibility project school volunteered to be interviewed. Individual, semi-structured interviews were conducted by me during the month of June 2012; all interviews took place in the feasibility project school and were recorded in mp3 format. Key phrases from the recordings are found in Appendix 3d. The following are the topic headings and questions used to guide the conversations:

- Initial reactions
- How would you describe the experience of having your students evaluate your teaching?
- Do you think you can rely on these data as being an accurate evaluation of your teaching during the Easter Term?
- Do you perceive that it is your teaching that has been evaluated or you as a person?
- What uses could you make of your data?
- What are your interpretations of these data? What do you think they mean?

- Do you think this [format] is the most effective way of presenting these data? If not, how could it be improved?
- What will you do with these data?

Initial reactions

None of the teachers interviewed said that there was anything particularly surprising in their data when asked for their initial reactions upon seeing their graph for the first time. Comments such as "That's pretty much as I expected" and "If I'd have drawn the graph myself it would have looked like that" are indicative of the narrative that emerged from all teachers, even from those for whom there were noticeable 'dips' in the data. In these cases, mild surprise was noted, but this was soon replaced with a genuine curiosity to know how the lower ratings might be addressed, and so, raised.

How would you describe the experience of having your students evaluate your teaching?

The themes that emerged very strongly from the majority of responses to this question were of nervousness and apprehension, coupled with a resolve to know more. Several teachers used the term "nerve-wracking" but there was an associated excitement and interest in what their students had said about their teaching. Such nervousness and excitement seemed to be outward expressions of the seriousness with which these teachers took their jobs, and the degree to which they truly wanted to know of areas for improvement. Teachers spoke of students' "right to feed back" and a desire to receive information pertinent to their work with a particular group.

Despite the question asked in this section being about the experience of having their teaching evaluated, several teachers began to engage in ruminations on what they could

do to improve their work in light of the feedback presented to them, such appeared to be the power of presenting evaluation feedback of this kind.

Do you think that you can rely on these data as being an accurate evaluation of your teaching during the Easter Term?

There was some equivocation in several responses to this question; this lack of certainty seemed to be drawn from the reliability of data being drawn from small groups of students, as well as from the fact that students were perceived by some teachers as being unable to differentiate between the teaching that had occurred during the course of one term, and the teaching that they had received during the whole academic year. Students found it difficult to remember what had happened: one teacher commented that "It was a bit odd looking back retrospectively", while another suggested that "[t]hey always look back with rosy-eyed vision" when time has elapsed (between the end of one term and the beginning of the next).

There was a feeling amongst some of the teachers interviewed that these two factors (small class sizes and the ability of students accurately to reflect on teaching from a specified period, rather than just a general perception of teaching) would have reduced the accuracy of the evaluation data. Nonetheless, there was still a clear interest in investigating them further and using them as the basis for discussions around possible changes which could be made to their teaching.

Do you perceive that it is your teaching or you as a person that has been evaluated?

The reason for asking this question was to find out if teachers perceived a difference between the role for which they were employed (the teacher) and the activity in which they were engaged (teaching).

The consensus among the seven teachers interviewed was very much that separation of teaching and the teacher for the purposes of this feasibility study was very difficult, if not impossible. Most teachers thought that both the teacher and their teaching were being evaluated, with one commenting "I don't think they're mutually exclusive", and another stating that "You can't separate them". Such responses were very interesting, particularly in light of the early thinking I did around the semantics of student evaluation of teaching. I was very keen from the outset (and remain so) to express to those teachers involved in the feasibility and main projects that it was their teaching that was to be evaluated. The rationale for this was partly drawn from a desire to protect individuals (it is, perhaps, easier to cope with negative comments about one's teaching than it is to cope with comments about the individual as teacher) and a keenness to focus attention on malleable traits and actions. Without such focus, engaging teachers in the act of improving their practice would be more challenging.

What uses could you make of your data?

This question was included to try and activate teachers' thinking about potential uses of the data without the bias of my own ideas.

Where teachers had (relatively) lower ratings on their graphs, this question seemed to draw teachers' thinking. "… Maybe I could do something about that" said one teacher, while another spoke of "Looking at the weaker areas and trying to improve them". None of this

is ground-breaking information, but it was critical to see teachers suggest that they would seek to address areas for (perceived) improvement in some way or other. It became clear during these conversations that there was a need for a simple way for teachers to access the help and support they needed when it came to addressing relatively low ratings from students. It was generally at this point that I made mention of the plan to provide a set of materials for teachers which would offer specific strategies designed to improve teaching; this was met with approval and seemed to offer a sensible next step for using the data.

What are your interpretations of these data? What do you think they mean?

This question was planned to allow teachers to express their own meanings of the data, without the bias of my prior knowledge. In reality, the question didn't perform well as it had become almost redundant due to the discussions which had already taken place during the interview. Nonetheless, comments such as "The 3.5s give me something to focus on and I think I would be arrogant to think that there is nothing that I could improve" and "It's making us (sic) think of all sorts of things… it throws up lots of questions" spoke of these teachers' sense; they are aware that they can adapt their practice to effect positive change.

Do you think this [format] is the most effective way of presenting these data? If not, how could it be improved?

One goal of the feasibility project was to gauge reactions to the format in which the feedback was presented (as a line graph). The decision not to use a bar graph (which would have been a more logical format for data such as these) was one take chiefly through pragmatism: I wanted to be able to superimpose multiple classes' data for a single teacher

on one sheet. While it is entirely feasible to do this with a bar chart (and even to display a trend line with it), I wanted to keep the chart as simple and uncluttered as possible.

It came as no surprise that one teacher suggested using a bar chart, though she herself acknowledged some of the difficulties associated with putting multiple classes' data onto a single sheet. What was of greater interest to me, however, was the recurrent theme of a desire to see the range of scores from a particular group of students. Almost unanimously did the teachers in this group ask for a way to represent the spread of scores from their class; I considered confidence intervals, but did not feel that they would be interpreted correctly and so chose not to include them when it came to the design of the feedback for the main trial.

What will you / would you do with these data?

In this part of the interview, I was keen to have teachers consider next steps that might be taken in light of the feedback they had received. Inevitably, where there were relatively lower ratings on a chart, individuals spoke of trying to find ways to "do more" of a particular thing (such as group interaction), while one teacher spoke of returning to the classroom with "a spring in my (sic) step". While laudable, such comments point towards a fairly superficial response to the data, though this is entirely understandable and indicates, I think, the need for support materials and a consultation meeting to be provided. Without some medium by which a teacher can begin to get at the pedagogical heart of the matter when interpreting their data and considering what to do next - and in a well-supported and constructive environment - there would be little credence given to the assumption that they would have the time or capacity to be able to make significant use of the data.

Consequently, the limited responses to this question were instrumental in causing me to think very carefully about ensuring that teachers were not left isolated when they saw their data, and that there was a clear path that could be followed to access support and guidance as they sought to interpret and use the information gleaned and fed back to them.

The chief purpose of the feasibility study was to generate an understanding of the feasibility of collecting SET data in a UK secondary school; in order to do this, the following data were collected and are discussed below:

- Student comments on each item proposed for inclusion in the SecondarySEEQ instrument (transcribed in Appendix 3b);
- Student comments made in addition to the SET ratings given on the SecondarySEEQ instrument;
- Teacher comments about the feedback they received from their students.

The comments made by students in the talk aloud exercise (Appendix 3b) were highly illuminating in gaining an understanding of the meanings they attached to each of the items, but also in understanding more broadly their perception of teachers asking them to rate teaching. As can be seen from the transcript, students often made evaluative comments ("…she is a very good teacher…") without necessarily elucidating the specific aspects or features of teaching which they believed were indicative of being "very good". Equally, some comments demonstrate a keener insight into the actions and behaviours of teachers which led to certain valued outcomes for students; the request of one student that their teacher insist on the completion of homework[45] (as a means by which the value

[45] Interestingly, when this comment was discussed with the teacher concerned, she immediately stated that she knew the reason for such a comment. She had set homework for Year 13 students which some in the class had handed in, while others had not; her response to those not handing in work was that she would

of the task might be increased) is just one instance which suggests that students are capable of insight into the logic of teaching and learning. The comments made are shown below (not all teachers' teaching was commented on).

Student comments about Teacher 1	"More time spent looking at exam questions and how to answer them for the topics we are studying, mock questions etc."
	"Teach what we need to learn specifically relevant to the coursework. Perhaps be more friendly. Do not criticise students because it depresses and discourages students. Perhaps better feedback on essays."
Student comments about Teacher 2	"Teacher 2 is one of the best teachers for providing us with clear information. Her lessons allow us to listen to her explaining and take good notes."
	"Sometimes she can get impatient with certain students. I feel like I don't know where I stand with her and that I don't know what to do."
	"Overall I think she is a very good teacher. Very good at describing things in a simple way."
	"I personally would enjoy greater discussion of current developments in Subject 3, although I appreciate this is not always possible due to the limitations of the curriculum and the task of maintaining the attention and interest of everyone within the class."

not 'chase' them as they were in their final year of school. She told the class that if they handed work in, she would mark it and give feedback, but would not spend time trying to get work out of those who had not submitted it in the first place.

Student comments about Teacher 3	"I find working in groups a great way to gather different interpretations for the poems. It works effectively."
	"To encourage wider reading of essays on the texts, perhaps specific examples could be given as homework."
	"While discussions in class were encouraged, the material analysed was tackled by manner of the teacher reading it to the class, which failed to engage the class and grasp attention. This then had an effect when it came to discussion as people failed to pay attention to the material and had little with which to contribute. I would encourage alternative methods of keeping the class engaged. There was not much versatility in lesson plans, and while most enjoyed the subject and the discussions, lessons became more and more predictable and less enjoyable. The accessibility of the teacher is great and the class is usually at ease to ask questions or seek extra help. The board regulations do sometimes tend to make the teacher overly worried and this affects the amount of help she gives students in a negative way."
Student comments about Teacher 4	"He goes through things in a good level of detail so that we are given the ability to apply what we are taught to most situations, rather than some teachers who teach the basic concepts then leave the rest to us."
	"My teacher has been a very good teacher this year. In previous years, I have had teachers who move too quickly with work and I felt like I was being left behind. Teacher 4 explains the work very clearly. He makes sure that everyone in the class understands what they are doing. He has a good balance between being nice and being strict. He always makes sure he is available out of lesson time to go over work."

	“He is very approachable.”
	“More class discussion would be appreciated and so not skim over some chapters just as quickly as we have done would be appreciated also.”
	“His lessons are fun and informative, and concise, with lots of explanation.”
Student comments about Teacher 5	“I particularly like that Teacher 5 marks our work fairly and gives feedback on how we can improve. She is very organised and helps students if they are having trouble with work. Her teaching makes me enthusiastic about the topics we learn.”
	“Sometimes the lessons have too many practicals which we've done before in other years and know the outcome so they are unhelpful and limit the time spent on theory which could be expanded past the syllabus into more interesting / current areas of Subject 2 which are still related to what we need to learn anyway.”
	“Continue to be a great teacher.”
	“Insist on homework being completed, as this will make it more valuable to learning. Focus on note-taking, whilst expanding on what we are writing, so we have solid revision notes for exams.”
	“I would like more time in lessons to take notes.”
	“Less self-directed learning and more note-taking in lessons.”

Student comments about Teacher 6	"The detail is good, but I would feel more comfortable for the exam if we did overview work first for the text and ensure we look at all the chapters in equal detail, even if that means less detail initially."
	"The lessons are engaging and really inspiring; Teacher 6 makes you genuinely interested in reading and her passion really comes across. But it felt like we didn't have enough focus on exam technique and could spend too much time doing contextual exploration which, though interesting, detracted from time which we really needed for more exam-focused activities."
	"Made me appreciate Subject 4 much more than I initially did at the beginning of the year."
	"Perhaps more lecture-style lessons where we are told the foundations and the basics, then, we can expand on that and go into more depth. The teacher has brilliant knowledge but perhaps is not always told."
	"Very fun, interesting lessons but perhaps need a bit more structure and focus."

Table 5.3: Feasibility study student evaluation of teaching comments

Chapter 6: Main project data collection

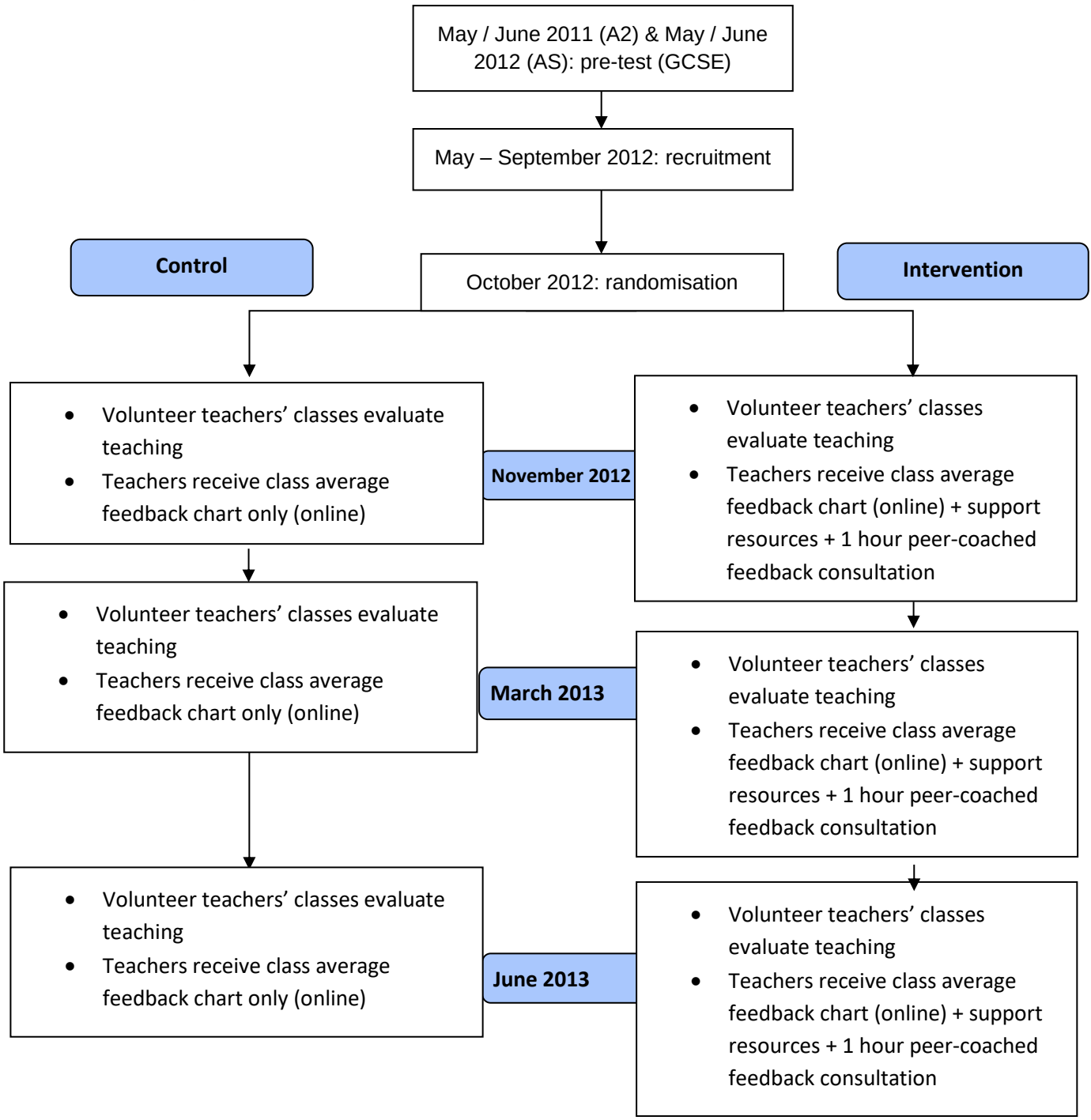
This chapter describes in detail the design of the student evaluation of teaching (SET) instrument used in the main project (based on the one trialled in the feasibility study); it will be referred to as the SecondarySEEQ instrument, reflecting the use of the Student Evaluation of Educational Quality (SEEQ) instrument (Marsh, 1982) in the secondary school context. As part of the evaluation design section, the chapter describes the recruitment and retention practices which were seen as crucial to the success of the study.

Thereafter, the discussion turns to the creation of an environment of trust in which those teachers participating in the study could easily see high potential benefit and low threat as a consequence of involvement. In this environment, specific rules for the uses of SecondarySEEQ data were created, and detailed guidance given for those in the intervention group regarding the role of the peer-coach colleague. These rules and guidance instructions are detailed in this chapter.

Having established a description of the overarching environment in which the SecondarySEEQ data were collected, attention turns to detailed descriptions of the intervention and control conditions; specifically, focus is placed on the application of the principles of solution-focused brief therapy (a form of cognitive behavioural therapy) and teacher learning communities in the intervention arm clusters.

In order to clarify the chronology of the intervention under evaluation, the following visual representation (based on the CONSORT (2010) flow chart) is provided:

SecondarySEEQ timeline



6.1 Design of the SecondarySEEQ instrument

The instrument that was eventually used in the main project was the product of theory (based on the significant development and applications of the original SEEQ) and collaborative enquiry (on the part of teachers and students at the feasibility project school and my own). This iterative process created an instrument whose content was almost identical to its predecessor, though the inclusion of language which is more accessible to secondary school students, as well as language which points more directly toward an individual teacher's teaching during a specified time period (the preceding term) mark the points of departure. As has already been said, the instrument remained very similar to the original partly at the request of Herb Marsh, and partly because of the lack of time to trial new items. The final 40-item version of the instrument that was trialled in the main project is found in Appendix 6a.

6.2 Recruitment

From the beginning of this PhD study, attention was given to the perceived potential for school and teachers to be sceptical of the value of participation, especially given the precise requirements expected of participants. Sample size calculations referred to earlier in this thesis indicated that some 48 schools would be required to detect an effect size of 0.3; this seemed like an ambitious target, but the degree of interest from schools was heartening to the researcher, and the ultimate sample of 36 schools recruited was a very encouraging sign indeed in the appetite for teaching effectiveness measures to be gathered and feedback presented to teachers. As the details of the recruitment process have already been given, the discussion will now move on to the measures taken to retain clusters in the trial.

6.3 Retention

The approach to retention adopted in this project can be outlined as follows:

- Build trust with school leaders;
- Make explicit all of the steps required to be taken by those participating (both schools and teachers within schools);
- Communicate clearly and in a timely manner the actions to be taken by participants;
- Respond quickly and accurately to questions;
- Be responsive to technical and practical issues that arise during the trial and document changes that had to be made accordingly.

The points above are discussed in detail below, and clear descriptions of the retention practices given.

6.4 Building trust with school leaders

Having established a relationship with school leaders in each cluster by communicating with them only during the recruitment phase, I felt strongly that two features of the project would be central to the successful retention of schools: my own experience as a teacher and school leader, and the presence of the Centre for Evaluation and Monitoring (a page giving basic information about the trial was created on their website and a small team assembled to assist with the creation of a database and a website designed to give feedback to teachers). The trust engendered between school leaders (who would act in a coordination capacity throughout the project, as well as offering leadership-level encouragement to those participating) and me was one of the most significant factors in the retention of 31 schools of the 36 randomised.

6.5 Creating a reflective working environment: building teacher trust

Building trust with individual teachers involved in the project was also seen as critical to the successful retention of the sample. The premise supporting this assumption was that, faced with students providing ratings of a teacher's teaching, negative feedback could easily lead a teacher to believe that continued participation was an undesirable and extraneous activity, potentially prompting a teacher to quit.

As with school leaders, my own background in teaching and the support of CEM were two key features of the foundations on which a relationship of trust was built. Allied to this was a communications plan designed to be as flexible (within the requirements of a faithfully-implemented intervention), as responsive (a direct phone line was set up, as was a dedicated email address and database) and as clear as possible (a set of protocols was drawn up for recurring issues).

6.6 Making explicit the steps to be taken

To provide clarity and to remove as many opportunities for misunderstanding were seen as important steps in the retention of teachers and clusters. At the point of registering interest in the project, teachers were sent an overview of key dates (for having their students provide feedback online using the SecondarySEEQ instrument, and for when class-average feedback would be ready to view). Thereafter, reminders of key dates within each term were sent during the relevant term, so that more timely reminders were created.

Emails were also sent both to the school leader coordinators and to individual teachers participating in the project, asking the latter in advance to book computer rooms for students to provide feedback on their teaching using the online SecondarySEEQ instrument, and asking the former to assist wherever possible (for instance, with

scheduling problems in schools where multiple teachers were participating). These emails were sent three weeks prior to the data collection point each term (when SecondarySEEQ ratings were collected), and then again two days in advance of the date set.

6.7 Communication

In a trial such as this, where clusters are located across the UK (from the south coast of England all the way up to Scotland, and as far west as Northern Ireland), the physical presence of this researcher (the only one involved in the conduct of the evaluation) in schools was not feasible. To counter the potential effects of demoralisation that were viewed as possibly affecting teachers during the course of the project, highly-effective communications (both proactive and reactive) were seen as vital. To this end, the dedicated email address and telephone line enabled me to provide support to teachers during term time for the duration of the trial. While this was imperfect (in the early days of the trial, the volume of emails and telephone calls was high and difficult for me to respond to in as timely a manner as intended), it was a useful tool in the retention of teachers.

6.8 Responsiveness to technical problems

The intervention relied to a large extent on technology: the SecondarySEEQ instrument was delivered online; communication between this researcher and teachers participating was via email and telephone; a database and website for processing student evaluation of teaching data and presenting resulting class-average feedback to teachers were created. Such heavy reliance on technology was an important part of the design of the intervention: the creation of an intervention that was scalable was paramount. Yet, reliance on technology brings with the potential for damaging problems. During the trial, I was

supported by several members of CEM staff who created and maintained the database and website used; without their vigilance and hard work maintaining the tools I used, there would have been a very different outcome to this project.

Most of the issues that arose were drawn to my attention in the first instance by teachers who made contact either by email or telephone. Thereafter, I would request technical support; necessary investigations would be undertaken and solutions implemented. Once fixed, communication would be made by me to the relevant teacher(s) to let them know the outcome and to advise regarding future actions. If the matter was applicable to all participating teachers, a mass email would be sent.

6.9 The intervention

The intervention under evaluation in this PhD project is most simply described as a peer-coached consultative feedback intervention for teachers, based on student evaluations of teaching data. The intervention was delivered in 15 schools over seven months between December 2012 and June 2013.

The overarching idea of the peer-coached consultative feedback intervention was to create a cyclical process which provided a rationale for an improvement goal to be set (generated by the student evaluation of teaching feedback received), a framework to help teachers interpret and understand their SET data (the peer-coached consultation meeting) and to commit to some implementation intentions (Gollwitzer, 1993) (using the resources provided on the SecondarySEEQ website). All of this was designed in a such a way as to provide feedback at three time points during the academic year (once per term, at the end of term), a relatively unusual arrangement for feedback on teaching (which often happens

only once or twice a year for UK teachers and is generally retrospective, negating the possibility of improvement during the academic year).

Peer coaching has its origins in the early 1980s, beginning as a tool used in the United States to support the implementation of new curricula and teaching techniques (Showers & Joyce, 1996). It is not a technique that could be considered 'revolutionary' in schools; teachers around the world are familiar with collaborative working, sharing experiences, planning units of work and lessons together, and engaging in collegial working generally. As such, the collaborative processes of planning, observation and feedback using peer coaching have most often been applied in informal settings, serving formal evaluative functions less often (Ackland, 1991). It is in an informal setting (a physical environment chosen by a teacher, accompanied by a trusted colleague, engaged in conversations about professional practice) that this study's intervention is situated.

6.10 The purposes of peer coaching

Wong and Nicotera (2003) usefully classify the most commonly-cited peer coaching purposes found in the literature: establishing a culture of standards and expectations; improving instructional capacity; supporting a process of ongoing evaluation; connecting classroom practice to the policy context. What is clear from the literature is that there is almost no support for the use of peer coaching as part of a formal teacher evaluation process, though Wanzare and da Costa (2000) do lend support to this notion. Again, this study aligns with the informal, diagnostic and formative functions of low-threat (to the individual teacher) improvements to teaching. Interestingly, putting SET data to use for formative purposes is not as common as perhaps it might be (Pallett, 2006), perhaps as a result of a greater interest in summative judgements, the lack of properly-developed and

validated instruments, as well as a dearth of mentors to support formative purposes (Benton & Cashin, 2012).

In this PhD study, peer coaching is adopted as a tool to enable teachers to reflect upon their class-average student evaluation of teaching (SET) data, and to set targets for their own teaching of the class from which the ratings came. As is seen in Wei et al. (2009) and Darling-Hammond and McLaughlin (1995), amongst others, this model of sustained (once per term for each term of the academic year) 'mid-flow' reflection on the quality of one's teaching (with a view to its subsequent improvement) is not a systematically-applied technique, despite there being plenty of support for the construction of CPD for teachers in this manner.

The rationale for peer coaching is explained in greater detail in this chapter, but briefly it is summarised as providing a safe, confidential context in which challenging ideas may be discussed and the professional advice of a trusted colleague sought, if desired.

6.11 Semantic variation

Peer coaching as a term is defined in multiple ways in the literature, and perhaps this variety is indicative of an approach whose name covers such disparate activities that its meaning has no singular understanding. There is a good deal of semantic variation in the literature on peer coaching, with the technique being referred to as:

- 'technical coaching' and 'team coaching' (focusing on incorporating new techniques into teaching) (Showers & Joyce, 1996);
- 'collegial coaching' and 'cognitive coaching' (focusing on the refinement of existing practice through professional dialogue and reflection) (Ackland, 1991);

- ‘challenge coaching’ (focuses on the ‘treatment’ of a specific problem) (Wong & Nicotera, 2003).

With so many different versions of peer coaching mentioned in key texts in the literature, we get a sense that a poorly-defined construct is at play, and this causes concern. This study aligns with the notion of ‘challenge coaching’, as the feedback provided to teachers as part of the intervention raises the idea of an area requiring improvement or, in other words, a perceived problem (relative to the teachers’ other feedback rating scores) to be solved. As such, where peer coaching is referred to in this thesis, the strand of this that is implied is ‘challenge coaching’ (Wong & Nicotera, 2003). SET data with a diagnostic purpose were issued to teachers in order to inform the conversation in their peer-coached meeting, as well as the subsequent decisions and actions they took.

Accepting - though not, by default, condoning - that there is a variety of terms indicating the same (or very similar) construct, it is useful to draw together the three key aspects of peer coaching found in the literature and usefully tabulated in Wong and Rutledge (2006) and shown, below, in Table 6.1:

Table 6.1: Peer coaching strategies identified in the research literature

Building institutional capacity

Provide training for coaches
Determine logistics, such as incentives and class coverage
Establish an ongoing process supported by modelling, coaching, collaboration, and problem solving
Focus on linking new information to existing knowledge, experience, and values
Coaches give advice about instructional content and strategies
Coaches share new ideas on curriculum and instruction
Coaches demonstrate classroom instruction

Development of a professional culture

Establish rationale for peer coaching program at school
Set criteria for the selection of coaches
Redesign a coaching culture that values collegial interaction and professionalism among participants
Define the roles of mentor and mentee: reciprocal or expert coaching
Have all teachers agree to participate in peer coaching

Ongoing support for evaluation

Make peer coaching distinct from teacher evaluation
Coaches observe classroom instruction regularly
Coaches provide support and companionship
Coaches consult about lesson plans and objectives
Coaches provide feedback and assistance through analysis of teacher application of strategies
Allow for experimentation of teaching strategies
Encourage reflection on the teaching and learning process
Evaluate the peer coaching program to determine the impact of the initiative on students and teachers

Source: Wong and Rutledge (2006)

Using the headings from Table 6.1, it can be concluded that the purpose of the peer-coached consultative feedback meeting in this PhD study was to build the capacity of teachers involved within their institution and to support ongoing evaluative processes in school. Such an approach has its pedagogical and philosophical underpinnings in the suggestion to 'love the one you're with', made by Wiliam (2007) in support of teacher learning communities (TLCs). The peer-coached model adopted in this intervention study is not a pure TLC, as defined by Wiliam, by virtue of its composition and confidential nature; it is, rather, a "building-based" and flexible (Wiliam (2007), p.197) approach that seeks gradual change in teacher behaviour and action.

For the purposes of this intervention's goals (to improve teaching and thereby student achievement), the following strategies found in the literature (and taken from Table 6.1) were viewed as offering support to the choice of peer coaching as an appropriate modus operandi:

- Focus on linking new information to existing knowledge, experience, and values;
- Coaches give advice about instructional content and strategies;
- Coaches share new ideas on curriculum and instruction;
- Make peer coaching distinct from teacher evaluation;
- Coaches provide support and companionship;
- Coaches consult about lesson plans and objectives;
- Encourage reflection on the teaching and learning process.

The final strategy listed in Table 6.1 ('Evaluate the peer coaching program to determine the impact of the initiative on students and teachers') is the focus of this PhD investigation's randomised controlled trial, as it seeks to ascertain the impact of this peer coaching programme on both students' evaluations of teaching and students' A Level grades.

Volunteer teachers' classes evaluate teaching

The following section serves to detail the specific instructions given to, and action taken by, the teachers involved in the evaluation, whose schools had been randomised to receive the intervention.

During the week commencing 19th November 2012, teachers who had volunteered to participate in the trial and whose schools had been randomised to receive the intervention were requested to take their classes into a computer room (or similar facility in school) to complete the SecondarySEEQ instrument online, using the web-based survey software provided by Cvent[46]. The licence for this software had been purchased by CEM, Durham University, and the security of the website reviewed by me in a meeting with a representative from Cvent.

An email was sent to all teachers in the intervention arm detailing the actions they should take in preparation for their students' completion of the survey. This email is found in Appendix 6b, but the paragraph containing the instructions to be given to students is printed below:

> 'You're going to take a survey that has been sent to us by a researcher at Durham University. The survey will ask you about my teaching and it shouldn't take very long to complete. I've volunteered to be part of the researcher's project and am interested to find out what you think of my teaching. As such, I want you to be honest when you give your responses. Even though you will be asked to put your name on the survey, I will not see this when I get the results back; I will only see class averages for our class and not any individual responses you give. Please don't talk during the survey - allow each other quiet time to think carefully. When you finish, please follow the instructions I've put on the board. Once everyone has finished, we will continue with our lesson as normal. Please now type this address into the browser and then follow the instructions on the survey website.'

[46] www.cvent.com accessed 02/03/14

I asked teachers to ensure that, as far as possible, students completed the survey during their normal class time, and with the teacher whose teaching they were evaluating present in the room, but working on a task that would necessitate their sitting quietly at the desk until all students had completed the survey. In doing so, I attempted to provide parity for all classes, at least as far as it was possible to do so.

Anecdotal feedback from teachers (telephone conversations and emails) suggested that classes mostly completed the survey supervised by their teacher during class time, though some reported that students completed it at home, or during break times at school. While this was not ideal, it is very much the reality of school life: students are absent from a lesson, or priorities change, necessitating rescheduling. In order for teachers not to feel too great a pressure, I was flexible when asked (on only a handful of occasions) if it was permissible for students to complete the survey out of normal class time.

While the one-week window for completion of the survey may seem variously naïve, pressurised and unfeasible, it was chosen purposefully. My experience in schools suggested that, as Parkinson's Law[47] states, work expands to fill the time allotted to it; therefore, a restricted time period (at least a perceived one) for survey completion seemed sensible, but only with the *de facto* addition of a further two weeks. In reality, therefore, I set aside three weeks for the survey to be completed, but told teachers there was only one. As expected, the entire three-week period was required for completion of the survey data.

[47] http://www.economist.com/node/14116121 accessed 25/6/14

Teachers receive class average feedback chart (online), support resources, and a one-hour peer-coached feedback consultation

In order for teachers to receive their class average SET feedback, support resources and materials to help them prepare for their one-hour peer-coached feedback consultation with a trusted colleague, they had to access a secure website. The following paragraphs serve to describe the design and functioning of this website and its associated database.

6.12 Website and database

Created by developers at CEM, the website and database used by the SecondarySEEQ project gave the highest priority to the security and accuracy of the data being collected and processed. Hosted on CEM's secure servers, the actual (and perceived) high level of security was an integral component of the study gaining and retaining the trust of teachers who were consciously putting themselves in what - for some, at least - was a position of vulnerability.

A critical component of the security was the manner in which teachers logged in to the website to access their data. To effect this, complex passwords were randomly generated in CEM, then printed and the individual sheets sealed in separate opaque envelopes; these were then sent by Royal Mail (TM) Recorded Delivery to the teachers in schools. Sending the passwords by mail was deemed to be the most secure method, as the Postal Services Act (2000) states that:

> "A person commits an offence if intending to act to a person's detriment and without reasonable excuse, he opens a postal packet which he knows or reasonably suspects has been incorrectly delivered to him."

The distribution of complex passwords led to two consequences that had, I believe, a negative effect on the retention of teachers in the study. Firstly, the passwords arrived by letter, many of which were mislaid or lost by teachers (the first request for a replacement arrived two hours after the Royal Mail website reported its delivery at the school concerned), necessitating replacement passwords to be generated and sent out. Secondly, the passwords were so complex as to confuse many teachers[48]. In the weeks following the distribution of login details, I received a high volume of phone calls and emails. While I had expected some, the quantity was indicative of the need for a revised approach, should this intervention be replicated. Queries were dealt with and, eventually, teachers were able to log on to the site and began to navigate it, in advance of their data being uploaded. Nonetheless, the potential for resentful demoralisation caused by the frustration of access difficulties should not be underestimated and will be discussed fully in Chapter 7.

As already mentioned, one feature of this project that cannot be stressed enough is the role the security of the data played in the creation of the appropriate context for the peer-coached consultation meeting. Only the teacher concerned, the database developers and I had access to a teacher's login details. Thereby, the onus was on the teacher to decide with whom she would share her data, and in what manner. The 'social evaluative threat' described by Wilkinson and Pickett (2009) was a critical guiding factor in the necessity of such confidentiality: even the notion that a teacher was being judged by anyone other than the trusted peer they invited to be their peer coach (and peer coaches were specifically instructed not to judge anything about the teacher's actions or behaviour) would have

[48] Had teachers received their passwords electronically, copying and pasting the password text into the SecondarySEEQ website would have made this process easier and more efficient.

created such a threat. But it is worth considering the potential positive impact of such a threat, an idea that will be developed later in this chapter.

6.13 Feedback

There is strong support in the literature that the effects of feedback combined with consultation are greater than when feedback alone is provided (Brinko, 1990; Hampton & Reiser, 2004; Marincovich, 1999; Penny & Coe, 2004). From their meta-analysis, Penny and Coe (2004) concluded that the strategies listed below offer the most effective conditions for a consultation:

1. Active involvement of teachers in the learning process;
2. Use of multiple sources of information (e.g., videotapes);
3. Interaction with peers;
4. Sufficient time for dialogue and interaction;
5. Use of teacher self-ratings;
6. Use of high-quality feedback information (e.g., student ratings);
7. Examination of conception of teaching; and
8. Setting of improvement goals.

This study's intervention adheres to all of the above, except for number 5 (teacher self-ratings), as this was not included in the project's plan.

Once students had completed the online survey, data were uploaded to the SecondarySEEQ website by me. Teachers were then able to log on to the website and view their class average SET data (on average, teachers did this three times in the first term, and twice per term in the other two terms). An example of this (showing each of the lines

representing class-average student rating data for each of the three terms of the school year) is given below in Figure 6.1.

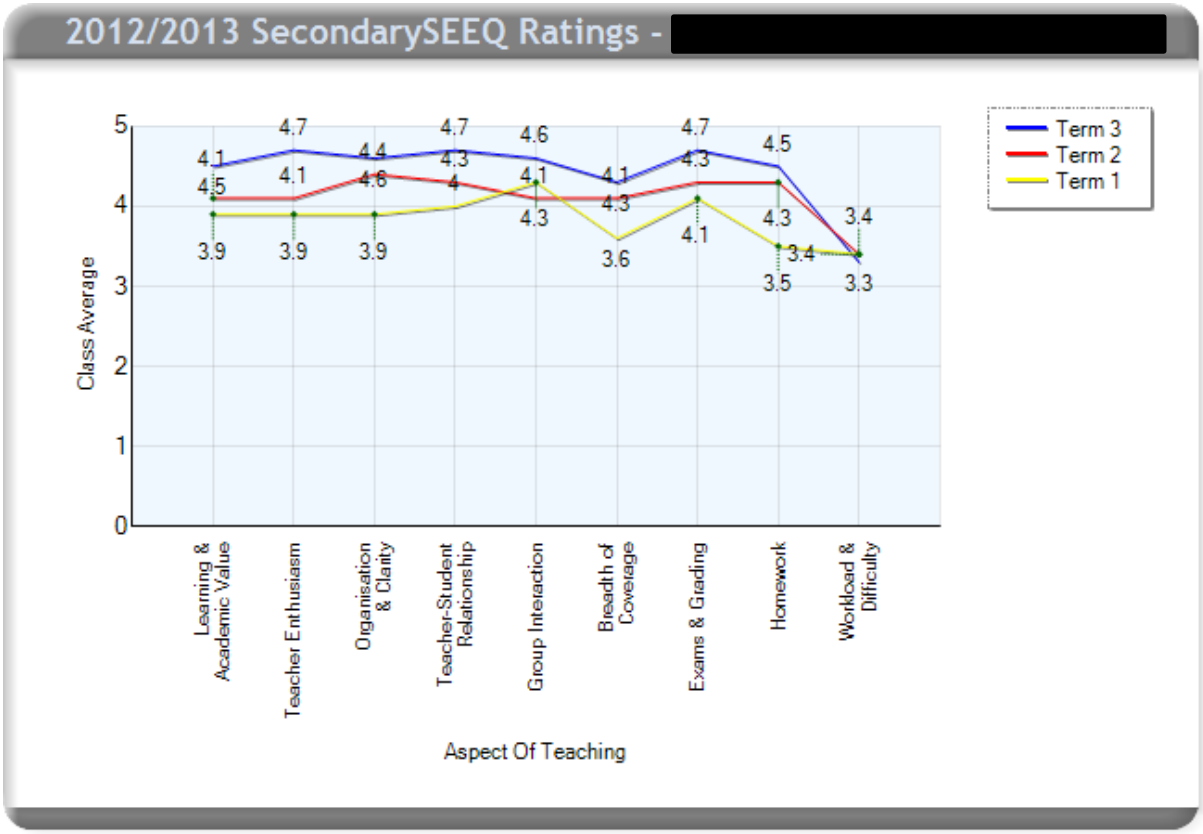


Figure 6.1: Example of multidimensional SecondarySEEQ SET feedback

Prior to gaining access to the data (as illustrated above in Figure 6.1), teachers had to accept a set of terms and conditions which had been developed by Durham University's Legal Department (see Appendix 6c). Thereafter, they were taken to a video of me introducing the process of reviewing feedback, setting up a consultation meeting with a peer coach, and thanking teachers for their participation. On this latter point (of gratitude), I made a particular effort in every communication I had with teachers (bulk emails, personal emails, telephone conversations, letters) to thank them for their participation and to highlight how their involvement in research was, more broadly, a positive and laudable thing.

Having watched the video, teachers accessed their data. They were encouraged to reflect on what the data might mean (how they might be interpreted within the context of the class from which they had been collected) and what approaches might be taken to improve the ratings in subsequent SET cycles during the academic year. Teachers were also encouraged to ensure that they didn't dwell too long in isolation on their data - the justification for this being that, without the presence of a trusted colleague, low ratings might have a demoralising effect which would adversely affect the intended outcome of the process (to provide the foundations for the improvement of teaching). Benton and Cashin (2012) suggest that it is in these moments of self-critical reflection that the cognitive dissonance created when a teacher's student ratings do not represent their own view of their teaching arises; they contend that it can be a transformational aspect of the process. It is at this point that teachers may be ready to recognise that there is something about their teaching that requires adaptation (Weimer, 2009).

The peer-coached feedback consultation meeting was designed with the guidance of goal theory (Locke, Chah, Harrison, & Lustgarten, 1989; Locke & Latham, 1990, 2002), solution-focused brief therapy (Corcoran & Pillai, 2007; Gingerich & Eisengart, 2000; Kim, 2007), and intention implementation (Gollwitzer, 1993; Hoell, Witt, Gilbertson, Ranier, & Freeland, 1997) kept in mind. Critical to the hypothesised effectiveness of the intervention was that the peer coach should be a colleague whom a teacher trusted and whose teaching they respected. In order to provide as pragmatic a framework as possible, these were the only two conditions to which the choice of peer coach was held. Thereby, a teacher could work with someone who knew their subject or was external to it, someone junior to them or someone senior. It may be considered retrograde to the goal of the intervention to seek a

peer coach who does not teach the subject of the teacher in question, but the notion that 'effective teaching' (when measured by the SEEQ) is affected by the subject taught, is not one supported by Marsh (2007c); this is echoed here. While the construct and the criteria are the same, the underlying practices of teaching in different subjects vary, hence the different responses and improvement strategies adopted by participants in the trial.

Teachers involved in the intervention arm were asked to input their peer coach colleague's name on their SecondarySEEQ web page. The rationale for this was to continue a process of commitment on the part of the teacher; Seijts and Latham (2000) as well as Locke and Latham (2002) suggest that some form of public declaration of goals increases commitment to them. While the definition of 'public' is entirely based on context, it is conceived in this case to relate to the peer-coached feedback consultation meeting. Perhaps this is not public enough. The low social evaluative threat (Dickerson, Gruenewald, & Kemeny, 2004; Gruenewald, Kemeny, Aziz, & Fahey, 2004; Wilkinson & Pickett, 2009) posed in an environment chosen by an individual who is accompanied by a trusted colleague is, perhaps, too small to create the potential for public shame associated with the failure to achieve goals. Nonetheless, for the purpose of this study, the notion of making goals 'public' is defined as negotiating and sharing them with one other colleague.

While many would argue that public shame is undesirable and unproductive (I am one of them), it could be contended that the threat of public shame associated with small social evaluative threats has a positive effect on goal commitment. A more positive and, perhaps, more productive way to look at this notion is with reference to the idea of social investment (Lodi-Smith & Roberts, 2007). Where an individual is 'invested socially' in their work, there is a greater probability of them committing to goals than if they are not; teachers who

volunteer to participate in a programme designed to improve their teaching may plausibly be defined as demonstrating this trait. Moreover, it is suggested by Gollwitzer (1993) that the generation of "implementation intentions" (defining the 'when, where and how' of the implementation of strategies designed to achieve a goal, as well as the specific process that will be undertaken to pursue the goals) has a positive impact on the achievement of defined goals. This is a concept that is realised in Marsh and Roche (1993), in which structured interventions for teachers were implemented, using the work of R. C. Wilson (1984, 1986) as a basis (but acknowledging methodological weaknesses in the latter's study designs which may have adversely affected the reliability of the findings and the validity of the claims made). The former authors consciously increased the methodological rigour of their study and, building on this work, my PhD evaluation adopted a randomised design (thus diminishing - though, clearly, not negating - the possibilities of dubious results) to evaluate an intervention that used online feedback and a shorter feedback consultation than in Marsh and Roche (1993).

Ratings procedures administered online (as opposed to on paper) have increased in number in higher education contexts as computers have become ever more numerous and capable of delivering the instruments with greater ease for users. Moreover, the speed with which data can be processed and analysed increases the timeliness of the feedback generated. One significant advantage for this study was the fact that a set of uniform procedures for all classes could be put in place, a feature noted in the literature as useful (Layne, DeCristoforo, & McGinty, 1999; Sorenson & Reiner, 2003).

It was pointed out by Sorenson and Reiner (2003) that lower response rates to fixed items was a feature of online student ratings data, and that this presented a threat to the ability

of the data to represent the class. In this study, all items were made mandatory (students had to provide an answer before they could move on and complete the survey), eradicating this problem. Feedback from students who took the survey, however, suggests that a sense of frustration at being asked the same questions repeatedly over three terms set in; this may have had an effect on the way they responded to items, thus having an impact on the validity of the claims that could be made. Nonetheless, the procedure described by Marsh (1982) to use class-average ratings accounts for this problem to an extent, but - clearly - classes in which there are fewer students are more prone to the effects of outliers.

Research shows no significant differences between the reliability of online and paper surveys (Benton & Cashin, 2012), supporting their use. Above all, the ease with which a survey can be administered to several thousand students is far greater than with paper, and the costs associated far smaller. There is no sound reason not to use online administration of SET survey instruments.

The guidance given to teachers and their peer coaches for the preparation and conduct of their meeting was as follows (a full set of slides is provided in Appendix 6d):

- Set aside one hour for this meeting;
- Keep the relevant class' data chart in view during the meeting (either on paper or on screen);
- Open the relevant teaching strategies PDF (linked on the teacher's feedback page);
- Recognise and agree on the purpose of the meeting (to interpret the data and set some goals for teaching in the subsequent term);

- Peer coaches are to be non-judgmental, disciplined (in keeping the meeting on-track), empathetic, attentive, accepting, warm, genuine, curious and in possession of a sense of humour.

Subsequently, the meeting guide video teachers and the peer coaches watched was used to provide cues for their discussion. By creating a set of six steps, each of which had specific questions to be answered or tasks to complete, I created a 'third person' in the meeting (me), as a guide to ensure that certain key facets of the goal-setting process were present. These steps were as follows:

- Step 1: What do you think the data presented on the graph mean?
- Step 1: Why do you think your students have given you these evaluations of your teaching?
- Step 2: Consider the two aspects of your teaching for which you received the lowest evaluations. If, when we meet towards the end of next term, you see improvements in these aspects of teaching, what helpful things might you or others have done?
- Step 2: What would have been the first thing that happened?
- Step 3[49]: Something amazing happens. Positive changes happen without your knowledge. When you realise the changes have happened, what would be the first, specific indication that your positive changes had been made and had worked?
- Step 4: If these amazing positive things had happened, how would the students in your class know?
- Step 4: How do you think your students would react?

[49] This part of the intervention seems, at first sight, a little bizarre. Nonetheless, it is taken directly from the practice of SFbT where it is known as the 'miracle question'.

- Step 4: What else would be happening that would indicate the amazing positive changes had taken place?
- Step 5: Decide on and record the specific goals [Mission Texts] you want to achieve in your teaching with a specific class.
- Step 5: Select and record the specific strategies you will use to try and achieve improvements in next term's evaluations.
- Step 6: Use the Reflection page to record your thoughts and observations on your experience in the SecondarySEEQ Project.

In light of the single hour suggested for the duration of the peer-coached meeting, the website was programmed to link to resources dedicated to providing strategies for the improvement of teaching in only the two lowest dimensions of teaching. The expectation that teachers would be able to cope with the cognitive demands of the task and successfully set goals (and strategies for their achievement) within an hour is a high one; according to those higher education instructors whose SET ratings have improved their end-of-course evaluations, however, significant effort is not required to make meaningful changes (McGowan & Graham, 2009). Small adjustments made to specific areas of teaching were the aim of this project and, while I do not underestimate that this was a challenging process for some teachers, I hope that it was practicable at the very least and didn't cause undue stress and difficulty.

It is important to acknowledge the potential limitations of an approach that calls on teachers to address aspects of their teaching which have been rated as needing most improvement (in the eyes of the students completing the ratings). There is an assumption on my part (in the capacity of intervention developer) that teachers have the capacity to

effect change in the areas identified, that their colleagues are able to help them make improvements, and that there is sufficient will to make the changes required to improve subsequent ratings. Such an assumption may not be correct, so it is worth acknowledging it as a potential limitation of the intervention design.

The provision of small and specific steps for an individual to follow ('channel factors') is suggested to impact on whether or not a person takes action (Levanthal, Singer, & Jones, 1965). This theory underpinned the construction of the consultative feedback intervention's peer coach meeting, and is a strategy which has achieved recent popularity with those who favour the use of behavioural science to influence policies designed to change behaviour[50] (BIT, 2012).

In Step 5 of the peer-coached meeting, teachers were requested to input a 'Mission Text' into the website. This text, recording the goal and implementation intentions that are noted to be of importance as actors in the ultimate achievement of goals (Guskey, 1988; Hoell et al., 1997; Lodi-Smith & Roberts, 2007), provided some rich insights which are explored in greater detail in Chapter 7. An example is given here, however, for illustrative purposes:

> "The area I am going to focus on improving over the course of next term are 'selling' my subject, so all students are aware of their learning and its academic value. Acknowledging that the lowest score (apart from workload and difficulty) comes from the 'breath of coverage' category, I am going to implement the following suggested teaching strategies from the secondary SEEQ notes: 1, 14 and 16. I

[50] The best example of this is the Behavioural Insights Team (BIT) (http://www.behaviouralinsights.co.uk/ accessed 3.7.14), an organisation created at Prime Minister's David Cameron's request and which, subsequently has been 'spun out' of government. It works with government departments in the UK and overseas on the development of strategies based on the behavioural science and economics literatures to influence citizen behaviour at scale. The BIT is a big supporter of randomised controlled trials as the most effective means by which 'what works' might be known.

> believe these will be the most useful for the students following this course at the current time."

The intention underpinning the engagement of teachers in discussing their data and subsequent teaching strategies with a colleague was to gain a different, supportive perspective on the SETs, but also to create an environment in which a 'contract' was publicly agreed, and goals set. The Mission Text was designed to form some kind of record that would represent a contract between the teacher and their future self: something that they could look back on in the following term, against which their progress could be compared when reviewing the quantitative SET data.

In addition to the Mission Texts, teachers were also asked to fill in a text box entitled 'Reflection Texts'. The aim of this text input was to capture the subjective impressions of teachers, in order to understand more about the experience of having their teaching evaluated, and of undergoing a feedback consultation process with their colleague. Again, full discussion of this is provided in Chapter 7, but an example is given here for illustrative purposes:

> "It is beneficial to be part of a project where students can, anonymously, provide critical feedback on their experience of my subject at post-16. Understanding the 'highs and lows' of student evaluation can help inform my own planning and practice, so that I can improve my own standards and transmit this to the student's learning environment. I would also welcome a brief qualitative analysis of the data, in writing (by e-mail!) to discuss with my faculty colleagues so as to address issues that may be common across the school in the post-16 experience. This should help to drive up standards of learning and teaching at this level. Thank you again for providing me with the opportunity to take part."

Clearly, this kind of rich narrative data is able to illuminate the feelings and experiences of the teacher concerned, but also provides valuable process feedback which contributes to the evaluation of the intervention.

6.14 The control arm condition

Those schools allocated to the control arm form what would be better termed 'comparison' clusters or 'active controls', rather than simply 'control groups'. A pure control group would have had 'business as usual', enabling a stark comparison to be made between the presence of the intervention and its absence. As is explained in the chapter on Methodology, this was deemed inappropriate for this study, as engaging sufficient numbers of teachers was predicted in the early design stages to be challenging.

The active control school teachers received feedback only in the form of the multidimensional chart. No notes were provided to assist with interpreting the data, nor were there any other resources to specify suggestions for the improvement of teaching in subsequent terms.

6.15 Collection of outcome data

One feature of this PhD project which was tremendously exciting was having access to the outcome (Advanced Subsidiary (AS) Level and Advanced (A) Level) data for students in participating cluster schools. As has already been mentioned, a feature of a school's registration to use the Advanced Level Information (ALIS) system provided by the Centre for Evaluation and Monitoring (CEM) at Durham University is the consent given for anonymised data derived from the ALIS system to be used for research purposes. From a purely administrative point of view, the huge benefit here is that schools have a reason not associated with - and arguably more potent than - this PhD project to comply with the requests of CEM to upload pupils' AS and A Level outcome data. For this reason, the missing data from pupils in schools which remained in the project (31 of them) are minimal and can be attributed to those students who ceased courses during the academic year or failed to complete the requirements for a grade to be awarded.

6.16 Conclusion

Having described in full the intervention and control conditions and the steps taken to collect the SecondarySEEQ data, the discussion now moves on to the analysis and interpretation of the findings.

Chapter 7: Main project analysis

This chapter sets out the analyses of data concerned with the validation of the instrument, as we well as those concerned with the impact of the intervention. Once analyses are presented, interpretations of the data are offered, prior to a summary discussion being presented in the subsequent and final chapter.

Of the 1241 schools which were assessed for eligibility in for this trial, 1204 were excluded (0 schools failed to meet the inclusion criteria due to the fact that they were all schools using the ALIS assessment and, therefore, also had students studying Advanced and Advanced Subsidiary Level qualifications; 38 schools declined to participate; 1,167 did not respond to recruitment communications), leaving 36 schools to be randomised. 18 schools were randomised to receive the treatment (the peer-coached feedback intervention for teachers) and 18 schools were randomised to the active control group (receiving a graphical representation of their student evaluation of teaching feedback only). Two schools in the treatment group did not receive the allocated intervention, and three schools in the active control group did not receive their allocated intervention (all schools in both arms did not receive their intervention as a result of not responding to communications post-randomisation). For the analytic portion of this thesis, all 36 schools randomised are included for Intention To Treat analysis, and sub-group analyses also conducted.

One of the important features of this study which should be mentioned early in the analytic section is that it was, ultimately, underpowered. Additionally, the active control arm created the very real increased possibility that any effect the intervention may have would be either very hard to detect, or simply undetectable. It is acknowledged that this is a result

of both a design flaw (not having a pure control group) and retention problems. Of the 36 schools recruited, 31 remained in the trial for the duration of the study, although data from all 36 are presented here for the purposes of Intention To Treat (ITT) analysis.

The design of this PhD study was intended to provide the framework and tools to answer two questions:

1. Is the SecondarySEEQ instrument used in this PhD study a valid means with which to capture measurements of teaching effectiveness?
2. Does the presence of a termly, one-hour, peer-coached teacher consultation meeting based on SecondarySEEQ student evaluation of teaching data (treatment) create a larger effect on attainment than the absence of it (control)?

In order to answer these questions clearly, this chapter has been divided into two, with the first part dedicated to answering the first question about the validation of the SecondarySEEQ instrument, and the latter part devoted to analysis and interpretation of the impact of the intervention on the outcome of interest.

7.1 Sample

Prior to answering the questions mentioned above, it is useful to understand more about the sample of schools participating in the main project. As such, descriptive statistics are given below.

Number of schools randomised	36
Number of teachers	249

Number of classes	481
Number of students	3002
Number of A/AS Level subjects	56
Average student GCSE score	6.64 (SD 0.86)
Average A Level points prediction	96.74 (SD 19.42)
Average AS points prediction	43.08 (SD 6.36)
Average age of students	17.2 years
Number of independent schools	15
Number of non-independent schools	21
Number of rural schools	1
Number of urban schools	35

The sample contained 15 independent schools and 21 state schools; of these one school is classed as rural, while the others are urban schools.

Along with acknowledging that the sample of schools is not representative of UK schools (the proportion of independent schools is far higher than the national figure), it should also be kept in mind that this sample was recruited from those schools already using one or more of CEM's assessment and monitoring systems. Such activity is suggestive of schools

already invested in the provision of feedback intended to support school improvement and, as such, is suggestive of bias which supports the view that the sample is not representative.

7.2 Validation of the instrument

In validating the SecondarySEEQ instrument for use with A Level students, the exhortation of Messick (1980) to adopt a construct-driven approach (as opposed to a task-driven one) to be adopted in the creation of performance-measuring instruments to begin by asking what "complex of knowledge, skills or other attributes" (p.16) should be measured. Thereafter, the identification of the actions or behaviours (the operationalisation, as represented by the latent variables) which would demonstrate such constructs should be completed, along with the suggestion of the kinds of activities that might enable such actions or behaviours to be witnessed and, subsequently, measured by means of the student survey instrument. Such focus, Messick contends, diminishes the possible inclusion of "construct-irrelevant" variance which may, ultimately, adversely affect the interpretation and use of scores.

7.3 A neo-Messickian approach

While Messick (1980) gave the foundations for this part of the study, the process of validating the SecondarySEEQ instrument for use in the novel context of UK secondary schools was, ultimately, conducted using the neo-Messickian framework for the evaluation of testing policy presented by Newton and Shaw (2014) (Figure 7.1). In operation, the technical evaluation questions (those concerned in the framework with the technical

quality of the mechanism) refer to the capacity of the SecondarySEEQ instrument to measure that which it intends to measure (something which Newton and Shaw say is truly in the spirit of Messick (1980)), while the social evaluation section refers to the value to society afforded by the mechanism (again, in this case, the instrument).

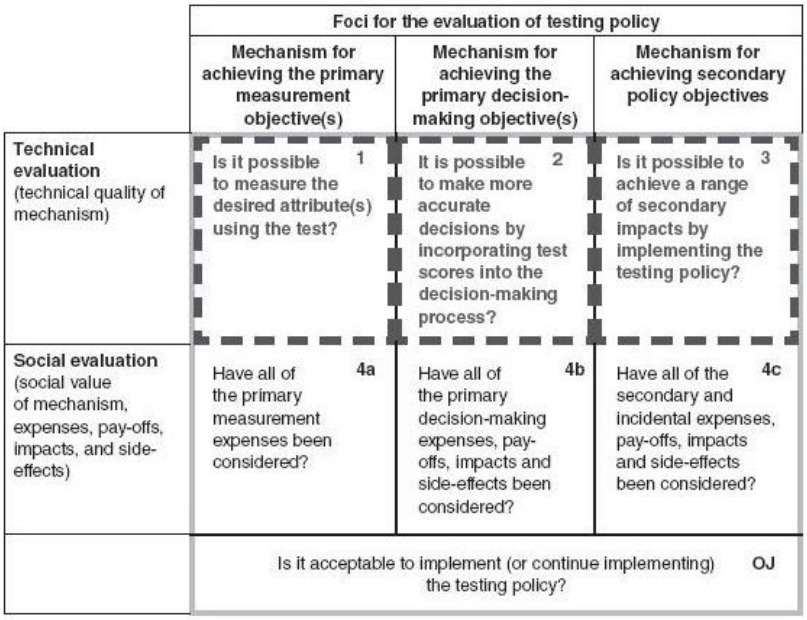


Figure 7.1: Key questions for the evaluation of testing policy under a neo-Messickian framework (Newton & Shaw, 2014)

This framework is adopted here for two reasons. Firstly, it clarifies some of the key aspects of measurement quality which must be considered in a testing validation process, while simultaneously avoiding some of the confusion caused by the multiplicity of meanings applied to the term 'validity'. Secondly, it offers a more balanced approach to validation: namely, it sees the value to society (society here is seen as schools, regions and education systems, as well as 'society at large') as being of equal import to the technical quality of the instrument. This thesis firmly situates its stance on the evaluation of the SecondarySEEQ in the work of Newton and Shaw (2014).

In operation, the application of the framework works as follows. The evaluation of the technical quality of the mechanism is enacted through a multilevel confirmatory factor analysis and regression analysis; the object of this part of the analysis is to understand whether or not it is possible (and the degree to which it is possible) to measure the desired attribute (quality of teaching) using the SecondarySEEQ instrument. The second part of the framework suggests that an evaluation of the instrument's application takes into account the costs and benefits to users and others affected by its use: this will be addressed by looking at the responses of teachers gathered in the 'mission' and 'reflection' texts (the means by which I was able to understand how teachers perceived the use and usefulness of the SecondarySEEQ and the system used to deliver it).

7.4 A priori foundations

In validating the SecondarySEEQ instrument for use with A Level students, the process used as a touchstone the a priori knowledge from the data available on its forebear, the SEEQ (Marsh, 1982). As such, it is crucial to note that the process undertaken was chiefly a confirmatory one rather than an exploratory one, using the theory established by Marsh and others as the foundation. Nonetheless, exploratory approaches were used in the first instance, as it could not – and should not - be automatically assumed that the performance of the instrument in such a novel context would match its predecessor.

7.5 Structural equation modelling

The structural equation modelling (SEM) approach adopted in this study is focused on two parts: the measurement model and the structural (regression) model. The former is a multivariate regression model which describes the relationships between the observed dependent variables (student's responses to the items on the SecondarySEEQ instrument, which act as factor indicators) and a set of continuous latent variables (the identified factors). The latter (structural) model describes the relationships (especially those with an implied direction, as seen in the regression analyses) among the factors, the relationships between the observed variables, and those relationships which exist between factors and other observed variables (those not classed as SecondarySEEQ items). Linear regression is used to describe the relationships.

The information provided by the SEEQ (Marsh, 1982) on the restrictions to be tested in the measurement model enables a more effective definition of latent variables, grounding the process in existing SET theory and the experience of manifold analyses of SEEQ SET data (Marsh, 2007c). Ultimately, by joining knowledge of the SEEQ's structure with the investigation of the SecondarySEEQ's structure, it was intended to specify a parsimonious model (Asparouhov & Muthén, 2009).

Using the information provided on appropriate techniques for modelling SET data (Marsh, Morin, Parker, & Kaur, 2014), the exploratory structural equation modelling (ESEM) approach was initially adopted in specifying an exploratory model to use with the SeecondarySEEQ data. Combining, as Marsh et al. (2014) indicate, the best aspects of confirmatory factor analysis (CFA), structural equation modelling (SEM) and exploratory factor analysis (EFA), ESEM's development as a tool has been significantly assisted by the parallel development of software packages such as Mplus (Muthén & Muthén, 1998 -

2014). Unfortunately, no software exists to cope with multilevel exploratory structural equation modelling, rendering this analysis unfeasible.

Any model including multigroup SEM can be described as a multilevel SEM, due to the hierarchical nature of the data (clustered individually and within groups). The most basic structure – that used in this analysis – is a two-level model with the lower level representing individual students, and the upper level representing students grouped in schools. The main aim of using such multilevel models is to summarise the within-group variability at the individual level, and the between-group variability at the group level.

7.6 Goodness of fit

Having established that multilevel structural equation modelling (specifically, multilevel confirmatory factor analysis) is the most appropriate form of analysis for the SecondarySEEQ data and their associated trial outcome data (standardised value-added data), a brief discussion of how the models will be evaluated is useful here. With the objective of identifying a parsimonious model kept in mind, attention is now turned to the goodness of fit indices chosen for use in this thesis.

Although there are no agreed guidelines for the acceptance of model fit, Schermelleh-Engel, Moosbrugger, and Müller (2003) suggest that an approach using the following steps is appropriate:

1. Establish that the model is identified;
2. Establish that there is convergence;
3. Establish that all parameter estimates are within permissible values;

4. Establish that the standard errors of the parameter estimates have a reasonable size;
5. Check the standardised residuals for patterns in the residuals matrix for suggestions of poor fit.

Moreover, with an absence of strict numerical consensus over what constitutes 'good' model fit, the indices used in this study will be considered simultaneously, thereby creating a triangulation of multiple information points and allowing the most robust judgement of the extent to which the model is consistent with the empirical data.

In triangulating goodness-of-fit indices, the problem outlined by Hayduk, Cummings, Boadu, Pazderka-Robinson, and Boulianne (2007) – that cut-off criteria are subject to misuse by those reporting them – is encountered. In this thesis, this potential problem is acknowledged and addressed in two ways: firstly, a range of appropriate fit indices (absolute and comparative) is presented and discussed; secondly, cut-off criteria are referenced, but are discussed within the context of the sample of data under investigation, and are used with caution.

Assessment of model fit to the SecondarySEEQ survey data was undertaken in light of work by Hu and Bentler (1999) and Schermelleh-Engel et al. (2003), focusing on seven statistics:

- Chi-square;
- Root Mean Square Error of Approximation (RMSEA);
- Standardised Root Mean Square Residual (SRMR);
- Tucker-Lewis Index (TLI);
- Comparative Fit Index (CFI);

- Akaike Information Criterion (AIC);
- Bayesian Information Criterion (BIC).

Below, each index is described briefly; discussion of the interpretation of each is given in subsequent paragraphs.

7.7 Chi-square

The Chi-square test is the only goodness-of-fit measure which has an associated significance test (while all other measures are purely descriptive). It is used to evaluate whether the population covariance matrix is equal to that implied in the structural equation model. Where the value of this statistic is high in relation to the number of degrees of freedom, the conclusion can be drawn that the matrices mentioned previously differ significantly from one another. As will be seen below, the Chi-square value for the best-fitting model of the SecondarySEEQ data is high, but this does not necessarily challenge the assertion of good fit; as will be explained later, the value can be artificially inflated when large sample sizes (as is present in the current study's dataset) are present. This point reinforces the need for an array of goodness-of-fit indices to be taken into account when judging goodness-of-fit (Hu & Bentler, 1999).

7.8 Root Mean Square Error of Approximation (RMSEA)

This statistic is a measure of approximate fit of the model to the data; the null hypothesis of exact fit mentioned above is here replaced with a null hypothesis of 'close fit'. Therefore, RMSEA is a statistic which is concerned with the discrepancy due to approximation.

7.9 Standardised Root Mean Square Residual (SRMR)

The Standardised Root Mean Square Residual (SRMR) is an absolute measure of fit; a value of zero indicates exact fit of the model to the data; values less than .08 are generally considered indicative of good fit (Hu & Bentler, 1999). The SRMR is defined as the standardised difference between the observed correlation and the predicted correlation.

7.10 Tucker-Lewis Index (TLI)

The Tucker-Lewis Index is an incremental fit index; the value of it depends on the average size of the correlations in the data. Therefore, low correlations between variables in the model will produce TLI values which are not very high.

7.11 Comparative Fit Index (CFI)

The Comparative Fit Index is another incremental measure of model fit; it is highly correlated with the TLI.

7.12 Akaike Information Criterion (AIC)

The AIC is a useful statistic for model comparison, as it adjusts the Chi-square statistic for the number of estimated parameters; it cannot be interpreted in isolation, only relatively. Of those compared, the model with the lowest AIC is considered the best-fitting on this measure.

7.13 Bayesian Information Criterion (BIC)

The Bayesian Information Criterion (BIC) is another comparative fit measure; lower values indicate better model fit (in comparison with other models). It is also worth noting that the BIC prefers a parsimonious model, so this is a useful point of reference in the particular objective of the model-fitting exercise undertaken in this doctoral study.

7.14 Interpreting goodness-of-fit indices

Barrett (2007) goes on to suggest that statements such as an "approximate fit" with RMSEA of 0.05 or 0.07, or a CFI of 0.90 (rather than 0.95) are devoid of meaning and that use of this term "approximate fit" is, by comparison with what is viewed by the author as the "more demanding" field of latent variable item response theory (in which a model either fits or does not fit), is redundant[51]. So the question that should be asked, the author proposes, is not whether or not there is relative goodness of fit, rather does the model fit the data or not, using the Chi-square statistic as the key means by which to answer. In the absence of such information, it is recommended that a claim of "empirical adequacy" be made. This approach is problematic, however, due to the impact of large sample sizes on (the inflation of) Chi-square values. Perhaps the better question to ask is: Is there a parsimonious model that fits the data better than an alternative model? With model comparison being the *Via Regia* in empirical research (rather than fit of a specific model to the data), this seems like a more pragmatic and appropriate approach to take (Hox, 2010).

Clearly, while the notion of 'approximate model fit' and rules of thumb (such as those popularly associated with Cohen's *d* (Cohen, 1992) have the potential to mislead,

[51] This is an interesting position to take, and a problematic one when model fit in item response theory is similarly arbitrary.

acknowledging the interpretations of certain value ranges for the statistics cited as representing a measure of goodness of fit is useful at this juncture, for the sake of clarity, particularly in light of more recent research in the area. As mentioned previously, the approach taken in this thesis is to adjudge model fit with cautious reference to a triangulated set of fit indices.

RMSEA values <0.05 are deemed to indicate a "close" fit to the data, while those <0.08 suggest a "reasonable" fit (Marsh, Hau, & Wen, 2004), although Marsh et al. (2009) state that more work is required fully to understand the application of such guidelines (ones which, it would seem, would make Barrett shiver) in ESEM studies. Furthermore, values of 0.90 and 0.95 on the TLI and CFI statistics infer, respectively, "acceptable" and "excellent" fit to the data (Marsh et al., 2009). SRMR values of less than .08 are considered by Hu and Bentler (1999) to represent good fit. Finally, in comparing AIC and BIC values for different model, lower values indicate better fit.

Acknowledging the warnings of Barrett (2007), this paper adopts the language of Marsh and colleagues in adjudging the fit of models tested with SecondarySEEQ data. Moreover, the advice of Marsh et al. (2009) to exercise subjective professional judgement in the determination of a "best" model (Marsh, Balla, & McDonald, 1988) is taken seriously.

Before detailing the specific process undertaken to fit models to the SecondarySEEQ data, the final model is presented below. Thereafter, the narrative to describe the iterative process of model fit (along with the respective models) is provided.

The final CFA model specified the following observed dependent variables:

- LAV1, LAV2, LAV3, LAV4
- TE1, TE2, TE3, TE4

- OC1, OC2, OC3, OC4
- GI1, GI2, GI3, GI4
- TSR1, TSR2, TSR3, TSR4
- BC1, BC2, BC3, BC4
- EG1, EG2, EG3
- H1, H2

It also specified the following continuous latent variables:

Within-level factors:

- F1W: LAV1, LAV2, LAV3, LAV4
- F2W: TE1, TE2, TE3, TE4
- F3W: OC1, OC2, OC3, OC4
- F4W: GI1, GI2, GI3, GI4
- F5W: TSR1, TSR2, TSR3, TSR4
- F6W: BC1, BC2, BC3, BC4
- F7W: EG1, EG2, EG3
- F8W: H1, H2

Between-level factors:

- F1B: LAV1, LAV2, LAV3, LAV4, TE1, TE2, TE3, TE4, OC1, OC2, OC3, OC4
- F4B: GI1, GI2, GI3, GI4
- F5B: TSR1, TSR2, TSR3, TSR4
- F6B: BC1, BC2, BC3, BC4
- F7B: EG1, EG2, EG3
- F8B: H1, H2

With the sample size of the SecondarySEEQ (n=3002 students), guarded caution must be used in interpreting the Chi-square statistic, as larger sample sizes are recognised as an inflationary factor. As Barrett (2007) points out, the greater the sample size, the higher the risk that the model will fail due to the increased sensitivity of the test to even the most marginal of discrepancies. Nonetheless, accepting this inflated value, the Chi-square statistic provides a useful starting point in assessing model fit (Marsh et al., 2009); Barrett (2007), indeed, contends that it is the only 'substantive' test of SEM fit.

Barrett (2007) points out that, as witnessed with the guidance Cohen (1992) offered on the interpretation of effect sizes, a highly-cited paper[52] by Hu and Bentler (1999) has become something of a key text for researchers applying threshold values for fit indices in SEM. Encouragingly, the author also points out that subsequent work by Beauducel and Wittmann (2005), Fan and SIvo (2007), and Marsh et al. (2004) has questioned the usefulness of what is, essentially, a reductive heuristic the ultimate (unintended) function of which is to generate greater misunderstanding than it does clarity.

7.15 Fitting a model to the SecondarySEEQ data

To start the process of model fit and judgement, six models were tested. Beginning with single-level ESEM, models with seven, eight, nine, ten, eleven and twelve dimensions were specified and run in Mplus.

The *a priori* factor structure of the original SEEQ (Marsh, 1982) comprised nine dimensions, so it was interesting to see that the ESEM model apparently fitting the SecondarySEEQ data best under the conditions imposed by the fit statistics also comprised nine dimensions (in bold in Table 7.1 below):

[52] Google Scholar reports 20, 637 citations on 29/5/14

Dim	AIC	BIC	Chi-square	RMSEA**	CFI	TLI	SRMR
12	n. conv.	n. conv.					
11*	268241.615	270452.222	578.865	0.23 (0.21 - 0.25)	0.992	0.979	0.008
10*	268608.571	270685.779	1037.94	0.31 (0.30 - 0.33)	0.983	0.96	0.011
9	268975.652	270913.109	1268.555	0.033 (0.032 - 0.035)	0.979	0.955	0.012
8	269569.906	271361.260	1817.682	0.039 (0.037 - 0.041)	0.968	0.939	0.015
7	270457.178	272096.076	2609.486	0.045 (0.044 - 0.047)	0.952	0.918	0.019

Table 7.1: initial ESEM (not multilevel) goodness of fit statistics

* denotes Heywood case[53]

** Confidence intervals for RMSEA are reported in brackets

[53] A Heywood case is a negative estimate of variance or correlation estimate greater than 1 (an impossible value in the population). Kolenikov and Bollen (2012) discuss the various plausible causes of Heywood cases, concluding that sampling fluctuations and structural misspecification of the model are often the likely drivers.

The twelve-model solution returned a negative Chi-square value and, as indicated in the table, there was no convergence to be reported by the AIC and BIC.

Using the nomenclature described by Marsh et al. (2009), the RMSEA of the seven-, eight- and nine-dimension models indicates a "close" fit to the data, while the CFI and TLI indicate goodness of fit which are considered "acceptable" and "excellent" (depending on the model). The nine-dimension model, however, indicates values on both the AIC and BIC which are lowest of all non-Heywood case dimensions, and which perform best on the other indicators discussed above. As such, this nine-dimension model was selected for closer inspection, specifically with regard to factor cross-loadings. One of the great strengths is that it allows items to cross-load; consequently, the 'subjective professional judgement' recommended by Marsh et al. (2009) had to be exercised by examining what the item cross-loadings were saying. This process, detailed below, was undertaken with the notion of construct-driven instrument generation firmly in mind.

7.16 Cross-loadings on the nine-dimension ESEM (single-level) model

Inspection of the loadings of items in the SecondarySEEQ instrument revealed that, with a nine-dimension solution – initially offering support for the *a priori* factor structure of the SEEQ – there were factors in which the cross-loadings necessitated a revision of the structure and the analysis of the fit of this revised model.

Allowing the items to load freely offered some interesting insights into the data. Table 7.2 below shows the factor structure as described by the permitted cross-loadings. Of particular interest is that the first factor (F1) has nine items (all of which have in common

the notion of teacher input – enthusiasm and clarity, for instance): the four a priori Learning and Academic Value (LAV) items, a single Teacher Enthusiasm (TE) item and all four Organisation and Clarity (OC) items. Moreover, the TE4 item which appears in F1 cross-loads onto F2 (its a priori target factor), and LAV1 and LAV4 load on F9 along with the original OR items.

Table 7.2: item cross-loadings in the nine-dimension ESEM model[54]

F1	LAV1	0.393	This term, I found this teacher's lessons intellectually challenging and stimulating.
	LAV2	0.455	This term, I learned some valuable things in this teacher's class
	LAV3	0.654	My interest in this subject increased this term as a result of this teacher's lessons.
	LAV4	0.578	I learned and understood what my teacher has taught me this term
	TE4	0.575	My teacher's style of teaching held my interest during lessons this term
	OC1	0.626	My teacher's explanations were clear in my lessons this term
	OC2	0.321	This term, the resources and other materials my teacher gave the class were well-prepared and carefully explained
	OC3	0.339	This term, I knew where lessons were going because my teacher explained the lesson objectives clearly, then taught in a way that helped me to achieve thei
	OC4	0.298	My teacher's lessons enabled me to take notes this term
F2	TE1	0.516	This term, my teacher was enthusiastic about teaching my class
	TE2	0.824	This term, my teacher was dynamic and energetic when teaching my class.
	TE3	0.356	My teacher enhanced their teaching with the use of humour this term.
	TE4	0.175	My teacher's style of teaching held my interest during lessons this term
F3	GI1	0.635	This term, my teacher encouraged students to take part in class discussions
	GI2	0.720	My teacher invited students to share their own ideas and knowledge in my class this term
	GI3	0.353	This term, I was encouraged to ask questions and my teacher gave me meaningful answers
	GI4	0.508	I was encouraged to express my own ideas and / or question my teacher this term
F4	TE3	0.309	My teacher enhanced their teaching with the use of humour this term.
	TE4	0.207	My teacher's style of teaching held my interest during lessons this term

[54] LAV = Learning and Academic Value; TE = Teacher Enthusiasm; OC = Organisation and Clarity; GI = Group Interaction; TSR = Teacher-Student Relationship; BC = Breadth of Coverage; EG = Exams and Grading; H = Homework; OR = Overall Rating

F5	TSR1	0.504	This term, my teacher was friendly towards individual students in my class
	TSR2	0.553	This term, my teacher had a genuine interest in the individual students in my class
	TSR3	0.634	My teacher made students feel welcome in seeking help / advice in or outside of lessons this term
	TSR4	0.464	My teacher was adequately accessible to students outside of lessons this term.
F6	BC1	0.440	My teacher contrasted the implications of various theories this term
	BC2	0.500	This term, my teacher presented the background or origin of ideas / concepts developed in class
	BC3	0.548	This term, my teacher presented points of view other than his / her own when appropriate
	BC4	0.548	This term, my teacher adequately discussed current developments in the subject
F7	EG1	0.427	The feedback my teacher gave on my assessed work this term was valuable
	EG2	0.757	The ways in which my teacher assessed my work this term were fair and appropriate
	EG3	0.256	The assessed work and examinations I did this term tested what I was taught in lessons
F8	H1	0.691	Homework set by this teacher this term was valuable
	H1	0.691	This term, the homework set by this teacher contributed to my appreciation and understanding of the topic(s) taught.
F9	OR1	0.261	Overall, how does the workload in this teacher's lessons compare with other teachers' lessons in this school / college?
	OR2	0.587	Overall, how does the level of difficulty in this teacher's lessons compare with other teachers' lessons in this school / college?
	LAV1	0.286	This term, I found this teacher's lessons intellectually challenging and stimulating.
	LAV4	0.275	I learned and understood what my teacher has taught me this term

Clearly, there are items which do not perform well in certain factors. Items such as TE4 on F2, and OR1, LAV1 and LAV4 on F9 should be removed from the model if it were to be used for further analysis. Because the ESEM model was fitted only to generate a preliminary understanding of the structure of the data (a confirmatory factor analysis (CFA) would need to be the final model, due to the confirmatory nature of the analysis), no adjustments were made at this stage. Discussion of the CFA models tested will be presented later in this chapter.

An interesting feature of the items which load on F1 is that, semantically, they can be grouped into two categories: teacher inputs and student outcomes. TE4, OC1, OC2 and OC3 all refer to some form of teacher input (the style of teaching; explanations; resources), whereas LAV1, LAV2, LAV3, LAV4 and OC4 are indicative of student-level outcomes (intellectual challenge; learning; growing interest; understanding). Clearly, the strength of the loadings varies considerably across the items, but for this element of the analysis, all items were retained, although there was concern about the way OR1 and OR2 were behaving (discussion of this is given later in this chapter).

7.17 Respecification of the model

In recognition of the differential foci of the items loading on F1, the model was respecified as an eight-dimension CFA (single-level) with the following structure:

F1 LAV1, LAV2, LAV3, LAV4, OC4, TE4, OC1, OC2, OC3

F2 TE1, TE2, TE3

F3 GI1, GI2, GI3, GI4

F4 TSR1, TSR2, TSR3, TSR4

F5 BC1, BC2, BC3, BC4

F6 EG1, EG2, EG3

F7 H1, H2

F8 OR1, OR2

The fit of this model to the data is detailed in Table 7.3 below.

Dim	AIC	BIC	Chi-Square	RMSEA*	CFI	TLI	SRMR
8	272886.42	273655.05	4512.985	0.049 (0.48 - 0.50)	0.917	0.904	0.039

Table 7.3: eight-dimension ESEM (not multilevel) goodness of fit statistics

*Confidence intervals for RMSEA in brackets

Clearly, this model does not compare favourably with the nine-dimension model discussed previously, with all fit statistics showing poorer values, except for the Chi-square (which Barrett (2007) would argue is a positive finding of interest).

Recognising that the OR factor is one which, conceptually, is different to the other dimensions (in that it represents a comparative judgement of teaching), and that it is one which had caused concern amongst teachers from the very beginning[55], it was decided to test a seven-dimension model (which was identical to the eight-dimension model mentioned above, only with the removal of the F8 factor – the one composed of OR1 and OR2, the comparative items) would be tested. The new F1 and F2 factors (with loadings as specified above) were included as they had in the eight-dimension model; all other factors retained their a priori items. The goodness of this seven-dimension model is seen in Table 7.4 below.

[55] Pilot school teachers expressed some concern about the items in this factor as they were deemed to offer an undesirable comparison of individual teachers with colleagues.

Dim	AIC	BIC	Chi-Square	RMSEA*	CFI	TLI	SRMR
7	253930.24	254616.29	3552.069	0.46 (0.45 - 0.47)	0.933	0.923	0.035

Table 7.4: seven-dimension CFA model (single level)

The removal of F8 in this model had an interesting effect on the goodness of fit of the data to the model. It should be noted, however, that this part of the process was grounded with the caution that simply re-running models until such time as the fit indices suggest the best model is an approach that does not suitably exercise justifiable professional, subjective judgement based on substantive a priori theory. In this instance, the justification for the removal of the OR factor began with the concerns raised by students and staff involved in the pilot project, as well as my own concerns over the utility of a comparative rating for the purpose of diagnostic feedback designed to support class-focused teacher development.

As stated earlier, the newly-generated F1 appeared to consist of two conceptually different parts: those items relating to the inputs of the teacher, and those items relating to the student-level outcomes of teaching. Allowing the model testing process to be guided by the constructs under investigation, it was decided to divide the F1 factor into two new factors as follows:

- LAV1, LAV2, LAV3, LAV4, OC4
- TE4, OC1, OC2, OC3

As stated, the rationale for this decision to divide the items between two new factors in this manner was rooted in the indications provided by the cross-loadings, the substantive content of the items themselves, and my own judgement. The results of this revised model (known hereafter as Model 7.1) are presented below.

The model specified the following observed dependent variables:

- LAV1, LAV2, LAV3, LAV4
- TE1, TE2, TE3, TE4
- OC1, OC2, OC3, OC4
- GI1, GI2, GI3, GI4
- TSR1, TSR2, TSR3, TSR4
- BC1, BC2, BC3, BC4
- EG1, EG2, EG3
- H1, H2

It also specified the following continuous latent variables:

Within-level factors:

- F1W: LAV1, LAV2, LAV3, LAV4
- F2W: TE1, TE2, TE3, TE4
- F3W: OC1, OC2, OC3, OC4
- F4W: GI1, GI2, GI3, GI4
- F5W: TSR1, TSR2, TSR3, TSR4
- F6W: BC1, BC2, BC3, BC4
- F7W: EG1, EG2, EG3
- F8W: H1, H2

The path diagram presented in Figure 7.2 below gives a physical representation of the above within-level factors.

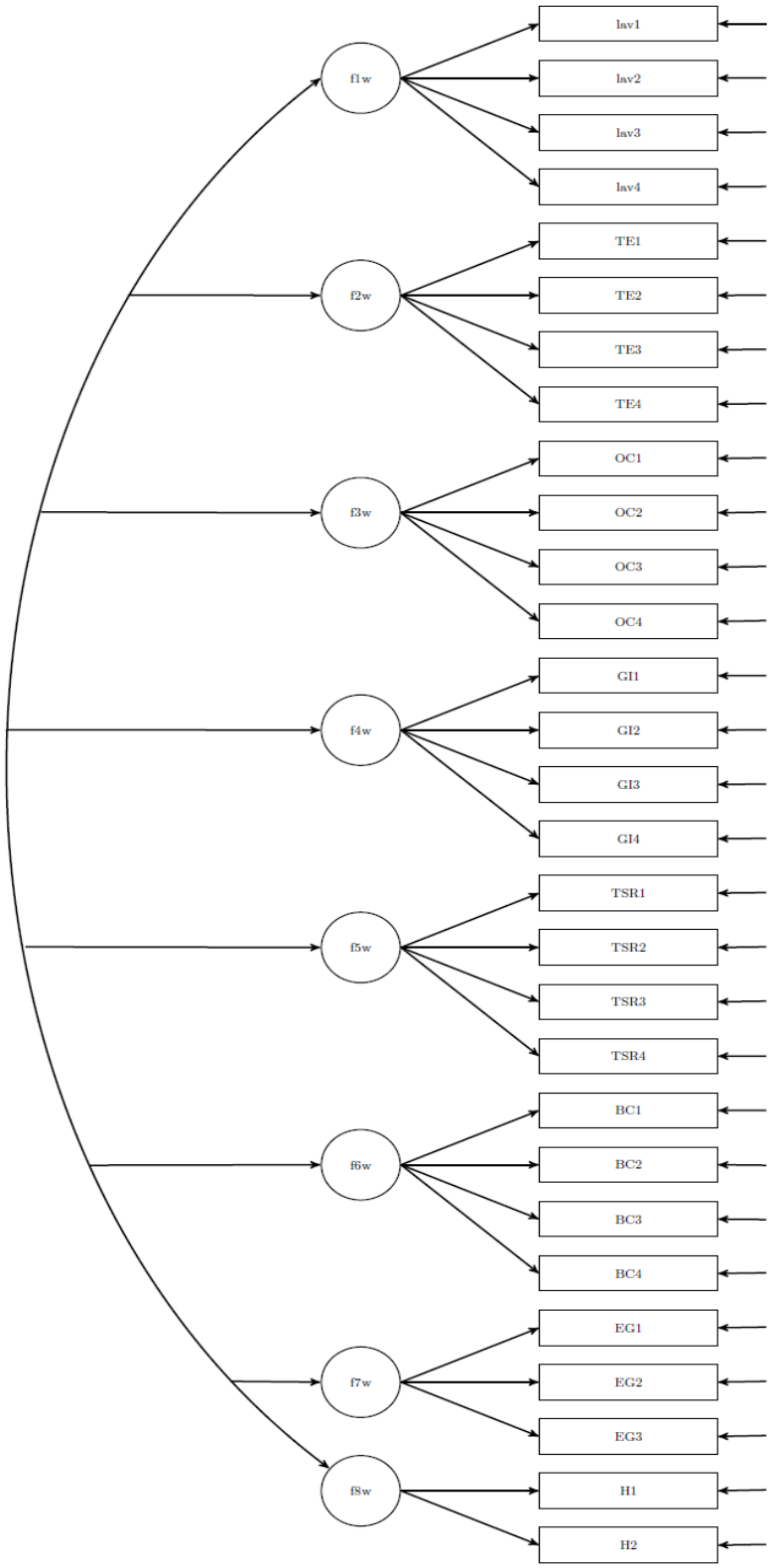


Figure 7.2: CFA within-level path diagram

The between-level factors are presented below and, subsequently, shown in the Figure 7.3 path diagram.

Between-level factors:

- F1B: LAV1, LAV2, LAV3, LAV4, TE1, TE2, TE3, TE4, OC1, OC2, OC3, OC4
- F4B: GI1, GI2, GI3, GI4
- F5B: TSR1, TSR2, TSR3, TSR4
- F6B: BC1, BC2, BC3, BC4
- F7B: EG1, EG2, EG3
- F8B: H1, H2

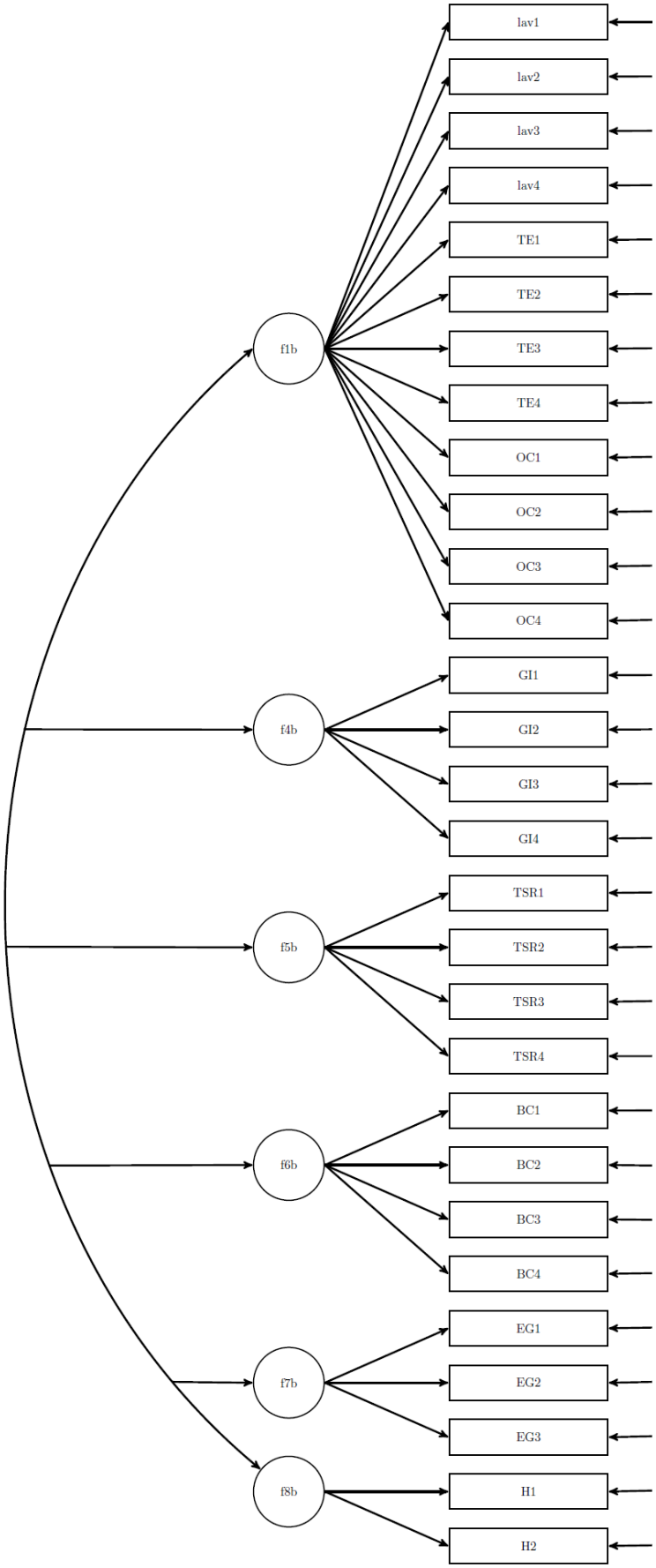


Figure 7.3: CFA between-level path diagram

The intraclass correlations (ICCs) for the Y variables in the final multilevel CFA model are presented in Table 7.5 below:

Variable	ICC
LAV1	0.290
LAV2	0.286
LAV3	0.272
LAV4	0.283
TE1	0.294
TE2	0.289
TE3	0.356
TE4	0.321
OC1	0.273
OC2	0.255
OC3	0.302
OC4	0.310
GI1	0.304
GI2	0.305
GI3	0.293
GI4	0.323
TSR1	0.287
TSR2	0.307
TSR3	0.274
TSR4	0.295
BC1	0.287
BC2	0.306
BC3	0.309
BC4	0.314
EG1	0.292
EG2	0.285
EG3	0.266
H1	0.321
H2	0.341

Table 7.5: intraclass correlation coefficients for the final multilevel CFA model (Model 7.1)

Having presented the relationships between the items and their factors, attention now turns to model fit. The model fit information below shows statistics for both the within-level and between-level analyses (with, obviously, only the SRMR value differing). Again, the Chi-square statistic should be treated with caution due to the inflationary effect of a

large sample size. It is important to note at this juncture, however, that model fit of single and multilevel analyses can't be compared, as the multilevel model includes additional manifest variables.

W/B*	AIC	BIC	Chi-Square	RMSEA	CFI	TLI	SRMR
W	362047.2	363314.3	5259.142	0.032	0.933	0.923	0.062
B	362047.2	363314.3	5259.142	0.032	0.933	0.923	0.107

7.18 Discussion of the two-level confirmatory factor analysis (Model 7.1)

The multilevel CFA model was the most parsimonious of all those tested in the analysis of SecondarySEEQ data. One of the most interesting features of this model is that it reflects the a priori model of the original SEEQ (albeit without the comparative items in the final factor specified by Marsh (1982)). Specifically, it is pleasantly surprising that an instrument designed for use with a higher education population – the items of which were changed only minimally to be more relevant to the secondary education population of interest in this study – retains the same structure when used in a novel context. It is feasible to suggest that, given these findings, the SecondarySEEQ data add weight to the assertions that student evaluations of teaching are reliable, valid and useful measures of teaching quality (Marsh & Roche, 1999).

The ICC values presented in the final model indicate a hierarchical structure to the data (students are nested within classes and individual observations are, therefore, not completely independent); this necessitates that the analysis takes the structure into account. The analysis uses the individual student as the cluster variable (on average, each

student in the data set provided an average of two evaluations); the ICC data presented in Table 7.5 show the variance in students' responses to the SecondarySEEQ items and, therefore, the degree to which they are able to discriminate between different teachers' teaching (ICC values range from .266 to .356). Muthen (1997) suggests that higher ICC values (around 0.5) are often associated with student test scores, whereas lower values are more related to attitudinal survey measures, of which the SecondarySEEQ must be considered one as it seeks information about student perceptions of teaching.

Interpreting relatively low ICC values such as these leads to the inference that students in this study were able to identify differences in the teaching they received from different teachers, though it was hoped that lower ICC values would have been identified. As a part of the analysis of the SecondarySEEQ data, this finding is of real significance as it lends support to the argument that the instrument is valid for the purpose of collecting subjective views of students about the teaching they receive: the simple fact that students appear able to see differences in the teaching they have received points to their facility for making judgements.

Moreover, inspection of the structural coefficients and their standard errors helps advance an understanding of how well each factor is measured and, thereby, an understanding of the reliability of the instrument. Assuming that a structural coefficient estimate represents the correlation between an item and a factor, assessing factor determinacy by looking at both coefficient estimates and the standard error associated with each is useful; in an ideal world, the correlations would be relatively high and the standard errors as low as possible. In the standardised model results for the SecondarySEEQ data, structural coefficients at the

within level range from .500 to .838, and standard errors at the same level from .008 to .025.

	Estimate	S.E.	Est./S.E.	P-Value (two-tailed)
Within Level				
F1W BY				
LAV1	0.581	0.024	24.366	0.000
LAV2	0.697	0.019	35.777	0.000
LAV3	0.701	0.017	41.396	0.000
LAV4	0.606	0.023	26.558	0.000
F2W BY				
TE1	0.836	0.009	90.289	0.000
TE2	0.838	0.008	106.709	0.000
TE3	0.707	0.014	50.233	0.000
TE4	0.735	0.022	33.252	0.000
F3W BY				
OC1	0.694	0.020	34.691	0.000
OC2	0.655	0.019	34.137	0.000
OC3	0.630	0.020	31.788	0.000
OC4	0.500	0.025	20.320	0.000
F4W BY				
GI1	0.705	0.025	28.359	0.000
GI2	0.745	0.024	31.002	0.000
GI3	0.690	0.025	27.818	0.000
GI4	0.735	0.016	45.191	0.000
F5W BY				
TSR1	0.745	0.015	51.266	0.000

	TSR2	0.805	0.012	64.601	0.000
	TSR3	0.723	0.019	38.510	0.000
	TSR4	0.553	0.027	20.756	0.000
F6W	BY				
	BC1	0.599	0.020	29.932	0.000
	BC2	0.648	0.019	33.390	0.000
	BC3	0.684	0.018	37.799	0.000
	BC4	0.600	0.021	28.277	0.000
F7W	BY				
	EG1	0.753	0.015	49.514	0.000
	EG2	0.782	0.014	54.200	0.000
	EG3	0.627	0.022	28.673	0.000
F8W	BY				
	H1	0.778	0.015	50.729	0.000
	H2	0.759	0.017	45.799	0.000
F2W	WITH				
	F1W	0.813	0.022	36.182	0.000
F3W	WITH				
	F1W	0.833	0.019	43.248	0.000
	F2W	0.641	0.034	18.699	0.000
F4W	WITH				
	F1W	0.686	0.029	23.670	0.000
	F2W	0.687	0.028	24.798	0.000
	F3W	0.657	0.032	20.751	0.000
F5W	WITH				
	F1W	0.725	0.025	28.502	0.000
	F2W	0.771	0.016	48.677	0.000

F3W	0.710	0.027	26.129	0.000
F4W	0.720	0.029	25.179	0.000
F6W WITH				
F1W	0.703	0.027	26.017	0.000
F2W	0.633	0.029	21.487	0.000
F3W	0.730	0.025	29.288	0.000
F4W	0.750	0.023	32.607	0.000
F5W	0.658	0.026	25.178	0.000
F7W WITH				
F1W	0.704	0.028	25.576	0.000
F2W	0.576	0.031	18.778	0.000
F3W	0.736	0.025	29.802	0.000
F4W	0.584	0.033	17.729	0.000
F5W	0.702	0.024	28.937	0.000
F6W	0.635	0.030	21.502	0.000
F8W WITH				
F1W	0.658	0.028	23.461	0.000
F2W	0.501	0.036	13.845	0.000
F3W	0.637	0.030	21.407	0.000
F4W	0.473	0.036	13.201	0.000
F5W	0.551	0.034	16.005	0.000
F6W	0.569	0.032	17.884	0.000
F7W	0.713	0.026	27.600	0.000
Between Level				
F1B BY				
LAV1	0.734	0.041	18.000	0.000
LAV2	0.842	0.027	31.426	0.000

LAV3	0.881	0.022	39.672	0.000
LAV4	0.781	0.036	21.810	0.000
TE1	0.598	0.152	3.927	0.000
TE2	0.667	0.147	4.550	0.000
TE3	0.506	0.084	6.060	0.000
TE4	0.981	0.020	49.863	0.000
OC1	0.926	0.020	46.850	0.000
OC2	0.821	0.035	23.761	0.000
OC3	0.804	0.032	25.399	0.000
OC4	0.702	0.037	19.156	0.000
F4B BY				
GI1	0.700	0.076	9.234	0.000
GI2	0.822	0.055	14.814	0.000
GI3	0.934	0.040	23.438	0.000
GI4	0.955	0.024	39.085	0.000
F5B BY				
TSR1	0.600	0.107	5.615	0.000
TSR2	0.785	0.068	11.495	0.000
TSR3	0.969	0.025	38.812	0.000
TSR4	0.911	0.035	25.877	0.000
F6B BY				
BC1	0.885	0.025	35.753	0.000
BC2	0.954	0.023	42.367	0.000
BC3	0.852	0.032	26.575	0.000
BC4	0.771	0.040	19.454	0.000

F7B	BY				
	EG1	0.882	0.024	36.573	0.000
	EG2	0.912	0.027	33.513	0.000
	EG3	0.846	0.031	27.027	0.000
F8B	BY				
	H1	0.968	0.018	52.839	0.000
	H2	0.976	0.017	59.103	0.000
F4B	WITH				
	F1B	0.778	0.055	14.187	0.000
F5B	WITH				
	F1B	0.806	0.057	14.204	0.000
	F4B	0.714	0.083	8.554	0.000
F6B	WITH				
	F1B	0.775	0.038	20.587	0.000
	F4B	0.751	0.045	16.726	0.000
	F5B	0.619	0.065	9.500	0.000
F7B	WITH				
	F1B	0.888	0.030	29.313	0.000
	F4B	0.724	0.064	11.292	0.000
	F5B	0.793	0.053	15.087	0.000
	F6B	0.667	0.052	12.859	0.000

F8B	WITH				
	F1B	0.786	0.033	23.956	0.000
	F4B	0.657	0.062	10.638	0.000
	F5B	0.698	0.061	11.346	0.000
	F6B	0.639	0.048	13.406	0.000
	F7B	0.765	0.040	19.081	0.000

At the within level, there is evidence of the a priori foundations of the original SEEQ instrument coming through in the structure of the data; the individual student perceptions of teaching effectiveness revealed here replicates the structure that has been found in other -mainly undergraduate – instances.

The between level, however, is structured differently, with Learning and Academic Value (LAV), Teacher Enthusiasm (TE) and Organisation and Clarity (OC) clustering together (although the estimates and standard errors for the TE1, TE2 and TE3 items show a weaker relationship than is seen, on average, across other items and factors). It is interesting to consider how, conceptually, these three aspects of teaching seem to align and represent something which may be called 'teacher input'. Here, it is hypothesised that teacher input is a characteristic which is shared among students in a class; it is difficult for the previously separate factors which comprise it to be disentangled at the class level. Moreover, when it is viewed in contrast with the remaining factors (Group Interaction, Teacher-Student Relationship, Breadth of Coverage, Exams and Grading, Homework), 'teacher input' may be seen as something which is in the domain of the teacher, whereas the other factors seem to depend on student interaction to a greater extent.

These results are very encouraging in the discussion of the SecondarySEEQ instrument's validity and reliability for the purpose of measuring students' subjective perceptions of teaching quality. This discussion will be taken up further in the final chapter.

One coefficient commonly used by researchers to indicate the reliability of a scale is Cronbach's alpha (Cronbach, 1951). While there is much written about the usefulness of the alpha statistic for comprehending the internal consistency of a scale, there is also a body of opinion which suggests it is flawed (Raykov, 1997; Raykov & Marcoulides, 2012; Sijtsma, 2009). Specifically, concern is raised that the statistic is, in many cases, a gross underestimate and that is not based on the usual assumptions about measurement error, and that inspection of the literature on alpha reveals that there is no clarity over the relationship between the statistic and the internal structure of an instrument (Sijtsma, 2009). Due to the problems associated with Cronbach's alpha mentioned here, the statistic is not used to advance the discussion of the SecondarySEEQ's reliability.

7.19 Analysis and interpretation of the impact of the intervention

This study was designed with the recognition that a state of equipoise is the ideal for an evaluator, but with equal recognition that my role as both intervention developer and evaluator would create conflict and the potential for bias. Throughout the project, however, I have maintained that I am most interested in the construction of the highest-quality evaluation of which I was capable; at the crux of this was a dispassionate view of the intervention itself. Clearly, it would be foolish to suggest that I wasn't keen to see the intervention have a positive impact; as a developer, I had a vested interest in its success. But I feel it is important to be clear that I would rather present accurate findings of no

impact or negative impact, or conclusions which recognise the limitations of the study and, in acknowledging them, state that only very limited assertions can be made, than present a shoddy evaluation which reports (incorrectly) a positive impact. My development as a researcher and as an evaluator is more important than the results of this evaluation.

7.20 Analysis of the intervention's impact on the outcome

One of the advantages of evaluating the impact of an intervention using a randomised controlled trial design is that the analysis is, in principle at least, straightforward. One of the most important factors in the analysis of cluster RCT data is to 'analyse how you randomise' (Torgerson & Torgerson, 2013); this is a guideline adhered to in the analysis that follows.

The outcome of interest in this trial is standardised value-added (StdVA). This outcome controls for baseline differences in individual student ability; it is derived by regressing the predicted AS or A Level outcome against the actual achieved outcome. The residual is a measure of progress familiarly known just as 'value-added' and is provided as standard in the feedback sent to schools subscribing to the Advanced Level Information System (ALIS) from CEM.

In this cluster randomised controlled trial, the unit of randomisation was the school; each cluster's average GCSE score was used in the stratified randomisation procedure carried out by an independent statistician. Logically, ITT analysis would, at first sight, indicate that data from all student in all schools randomised should be included; while this has the appearance of analytical purism, it is illogical and reductive in the context of this study's

aims. As such, the decision was taken to conduct ITT analysis on data only from those students who contributed SecondarySEEQ scores to volunteer participant teachers, and to present effect size estimates with confidence intervals to indicate a plausible range of values for the estimate. Prior to presenting data from this portion of the analysis, however, I turn attention to the somewhat problematic issue of reporting confidence intervals, in order to justify their use in this thesis.

7.21 A note on effect sizes and confidence intervals

Effect sizes

The effect sizes calculated in this thesis – using Hedges' *g* (Hedges, 2007) - are representations of the standardised mean difference between the outcomes (standardised value-added residuals) for students in the intervention and control condition groups (Coe, 2002), as indicated in Equation 1, below:

$$\text{Effect size} = \frac{[\text{Mean of the intervention condition}] - [\text{Mean of the control condition}]}{\text{Standard deviation}}$$

Equation 1: effect size calculation

Student-level value-added residuals for students in each cluster were collected by the Centre for Evaluation and Monitoring, Durham University as part of the ALIS system, described previously and means generated for both the treatment and control groups. The pooled standard deviation (defined as the average of the standards deviations of the control and intervention groups) (Coe, 2002; Hedges & Olkin, 1985) was used in the calculation of effect for the SecondarySEEQ trial. The assumption central to this decision is

that the standard deviations for both the control and the intervention groups differ only as a consequence of sampling variation, as is the case in the SecondarySEEQ trial. As such the use of the pooled standard deviation is justified in this case.

Confidence intervals

As with so many aspects of research, there are highly contrasting views put forth by academics regarding the use of confidence intervals (CIs) in the reporting of effect sizes. The argument that confidence intervals enable the reader to understand what a plausible range of values for an effect size would look like (with a defined level of confidence which, in the social sciences, is most often 95%) is one commonly propounded. Those who support the use of CIs suggest that, in particular, where two or more interventions produce positive effect sizes, CIs enable comparisons that are more meaningful and refined than those simply based on a point estimate. Two effect sizes of, for instance, 0.3 appear to be identical without CIs, but one may appear far more attractive than the other when an understanding of the plausible range of values is generated. The argument for the use of CIs is, therefore, a pragmatic one, and one which seeks to engender a deeper understanding than that achieved simply by reporting the effect size estimate alone.

Opponents of the inclusion of confidence intervals in effect size reporting (of whom Gorard (2013) is one whose argument is powerful and persuasive) claim that they do not perform the function intended (Carver, 1978) and can lead to mistakes of interpretation (Falk & Greenbaum, 1995); indeed, Watts (1991) claims that confidence intervals are almost universally misinterpreted.

Why are confidence intervals, seen by some as a pragmatic approach to the communication of error and plausible outcome values, viewed as such highly problematic statistics? The

crux of the problem lies in the notion that the probabilities reported by confidence intervals are based on assumptions about the sample which are untrue (the parameter estimates are based on 'untrue' models). Specifically, one argument is that confidence intervals do not take into account the quality of the sample drawn, or the attrition present in the sample (Lipsey et al., 2012) and are, thereby, flawed. A second allied – and powerful – argument suggests that, because confidence intervals assume complete random samples of a kind never actually found in the quotidian reality of research, they are inappropriate as a means of reporting probabilities (Berk & Freedman, 2001). Moreover, the further away from the point estimate one looks on a confidence interval line, the less likely it is that the true value is found, though this is something rarely represented by the line itself (most confidence intervals appear as continuous lines of equal width). It has to be concluded, therefore, that this feature of CIs is yet another reason for treating them with caution at the very least.

Accounting for all of the above, the decision was taken to report effect sizes with confidence intervals in this thesis, but also to provide a commentary offering plausible interpretations of them, and recognition of the limitations inherent in them. This may seem like a contrary decision, given the lengths I have gone to in arguing against their inclusion, but the fact remains that they still represent a useful mode of representing effect size estimates as accurately and fairly as possible.

7.22 The impact of the intervention

In the first instance, ITT analysis of data from all 36 schools was carried out. Analysis of clustered standardised value-added data revealed that the intervention did not have an

impact; the confidence intervals around the estimate, however, show the imprecision of this finding (Table 7.6):

ES	SE	Upper CI	Lower CI
-0.08	0.33	0.57	-0.73

Table 7.6: effect size and confidence intervals

The logic model behind this analysis is faulty (all standardised value-added score data from all students in each randomised school were included): correct conclusions are masked by the proportions of students who were in a randomised school, but whose teachers did not volunteer to participate in the trial. To 'analyse as you randomise' at the cluster level while ignoring the presence of data unrelated to the trial is illogical, but data from such an analysis are presented above to provide a defence to this position (Xiao, Kasim, & Higgins, 2016).

Additionally, complier average causal effect (CACE) analysis (Angrist et al., 1996; Raudenbush & Bloom, 2015) is used to refine the analysis by looking at impact data only for students in the classes of teachers who submitted reflection texts (detailed later in this chapter) online (these texts are used as an indicator of teachers for whom there is clear evidence of compliance with the treatment condition). This analysis offers a useful insight into the effect of the intervention on those who received the assigned treatment.

7.23 Impact analysis of students with participating teachers

A set of regression analyses was run in order to answer key questions which would allow the research question associated with impact to be answered. The question to be answered was:

1. What effect did the peer-coached intervention have on student progress as measured by standardised value-added scores?

There was no effect of the intervention on student progress as measured by standardised value-added scores for all students (in both year groups), as revealed in the table below.

ES	SE	Upper CI	Lower CI
0.081	0.072	0.222	-0.060

Table 7.7: Effect size and confidence intervals

The mean intervention effect (0.081) is very small but positive; however, when 95% confidence intervals are incorporated into the analysis, the conclusion must be drawn that there is no effect of the intervention.

Such a finding is disappointing, but not surprising. There are two factors which, upon reflection, appear to have contributed to the null effect: the lightness of the intervention itself (each peer-coached meeting lasted approximately one hour and relied upon teachers actively changing their practice in light of SET data presented to them); and the presence of an active control group. The effect of the second point here should not be underestimated: teachers in the control arm receiving SET feedback could quite easily begin to interpret and act upon the data of their own accord, thereby producing an effect.

It is also useful to put the very small mean effect size of the intervention into the context of progress made by other students in the 1,945 schools using the ALIS system during the years 2009 – 2015. The mean standardised value-added score for all students in this period (during which the teachers of the students concerned did not receive the treatment intervention) was -0.009 (sd=0.975); in essence, most students made the expected progress that was predicted by ALIS from their average GCSE scores. In the evaluation under consideration in this thesis, the intervention had an effect of 0.081, though the lower confidence interval includes the average standardised value-added score for all other pupils in the much larger, six-year sample, thereby rendering any difference undetectable.

Sub-group analyses do not offer insight or conclusions that differ to that presented above. Isolating the Advanced Level and the Advanced Subsidiary Level data reveals two pictures which are very similar: very small effect sizes with confidence intervals passing through zero.

To gain a fuller understanding of the effect the intervention has on those who take it up (as opposed to those who were offered but did not comply with their allocated condition), it is useful to look at data pertaining only to those participants who complied with their assigned treatment condition; the design of the trial and the use of a website to deliver the intervention and active control conditions means that compliance data (in the form of website metrics for logging on and inputting data, described below) are available for each participant.

Accounting for participant compliance – using Complier Average Causal Effect (CACE) analysis (Angrist et al., 1996) – provides insight into the effect not simply of the *offer* of the intervention (the role of the ITT analysis), but of the *take-up* of it. For the purpose of this

thesis, calculating the difference between the treatment group and control group means in the subsample of those participants identified as complying with their assigned treatment (as indicated by website use data detailed below) enables the calculation of the CACE statistic.

7.24 Complier Average Causal Effect (CACE) analysis

While the use of ITT analysis is appropriate as the primary analysis in this study, its limitations must be recognised and addressed in order to ascertain as accurate an inference about treatment effects as possible. As an indicator of the effect of *offering* the SecondarySEEQ intervention, ITT analysis is appropriate because it retains and uses data for all assigned participants in their original randomised groups; thus, it offers the least biased estimate of impact, albeit a conservative one.

ITT analysis preserves the independence of randomised assignment, but its dilution by non-compliance creates the need for analyses which identify the average impact for those teachers who *took up* the offer of the treatment as designed, the so-called local average treatment effect (LATE), or complier average causal effect (CACE).

Per protocol analysis (using data only from subjects who actually complied with their assigned treatment condition) is not appropriate as it fails to acknowledge that neither the treatment group nor the control group can automatically be considered homogeneous in terms of participants' take-up of (compliance with) their assigned treatment condition. Equally, on-treatment analysis (analysing data by treatment received rather than treatment allocated at random) is also inappropriate; both methods assume that the

probability of complying with treatment allocation is random. As Stuart and Jo (2015) put it, those who comply with treatment in the intervention group are a "self-selected subgroup of the full treatment [intervention] group"; as such, comparisons with the entire control group, or with participants in both groups who comply with their treatment condition are likely to generate biased results due to selection bias. For the above reasons, a CACE analysis approach is justified.

Complier Average Causal Effect (CACE) analysis (Gerber & Green, 2012) is a type of analysis which offers a useful tool for understanding the SecondarySEEQ data and, more broadly, RCT data in education trials due its focus on treatment effects for those individuals (in this case, teachers) who "receive a meaningful dose of treatment services" (Schochet & Chiang, 2009) (defined below in the context of this trial). CACE keeps the initial randomisation intact and, thereby, offers a more robust form of analysis than per-protocol or on-treatment analysis.

There are considered to be four subpopulations of participant when using a CACE analysis approach: compliers, always takers, never takers, defiers (Angrist et al., 1996). Compliers are those who take the treatment to which they were assigned; always takers are those who would receive the treatment irrespective of their assignment (this is not possible in the SecondarySEEQ study due to the online delivery of the intervention); never takers are those who do not take up the offer of their assigned treatment; defiers are those who are "induced to avoid the treatment by assignment to it, and induced to take the treatment by assignment to the control group" (Angrist et al., 1996). In this study, an assumption that there are no defiers is held.

Because the treatment (the online peer-coached meeting guide) could only be experienced through the SecondarySEEQ website (logging on to view feedback and entering text to the website), data are available to indicate which participants are deemed 'compliers' and which 'non-compliers'; in this trial, the assumptions required for CACE analysis to be used (detailed below) are satisfied; in particular the exclusion restriction (the causal effect of random assignment on the STDVA outcome for both the always taker and never taker subpopulations is zero) and the monotonicity (no defiers) assumptions.

Assumptions

CACE analysis assumes that participants assigned to the control group have the same probability of non-compliance as those assigned to the treatment condition, and that simply being offered the intervention has no impact on the STDVA outcome.

The following assumptions also apply to the use of CACE analysis (Hernán & Robins, 2006; Schochet & Chiang, 2009), each of which is deemed to be satisfied in the SecondarySEEQ trial:

- The outcomes for each participant are independent of outcomes for others involved in the trial (Stable Unit of Treatment Value Assumption, or SUTVA);
- No participant defied their assigned condition (monotonicity);
- The compliance rate is not zero: the offer of the intervention to those in the intervention arm generates compliers;
- Assignment is truly random (ignorable assignment);
- There is no effect of randomisation in the absence of treatment; in other words, there is a mediating effect of treatment receipt (exclusion restriction).

Identifying compliance

Pragmatically, the reflection and mission texts (detailed later in this chapter) submitted to the SecondarySEEQ website were used as the indicator for compliance. To indicate compliance with assignment to the intervention (limitations of this method are discussed later), the following condition had to be met:

- At least one peer-coached meeting reflection or mission text was inputted to the SecondarySEEQ website.

The two resulting latent classes of participants were coded as a binary variable (0 = non-compliance; 1 = compliance). While observable, this is not a completely flawless indicator of compliance, but it serves the purpose well: teachers in the intervention group who took the time to write up their thoughts following their peer-coached consultative feedback sessions showed evidence of compliance with their assigned treatment.

Calculation

Because there are two types of teacher identified in the SecondarySEEQ data for the purpose of this analysis (compliers and non-compliers), the effect of the treatment on those who received a 'meaningful dose' of it can be expressed as the average effects for these two groups of teachers. In order to do this, the proportion of teachers who complied with their assignment (those who recorded something in the mission text or reflection text section) must be ascertained so that the CACE statistic can be calculated in the following way:

CACE = ITT /proportion of compliers

In essence, the causal effect of treatment assignment on the outcome (ITT) is divided by the causal effect of treatment assignment on the treatment actually received. The proportion of compliers is calculated by subtracting the proportion of people who are always takers (those who, even in the control group would receive the treatment) from the proportion of people who are always takers *or* compliers. Because there are no 'always takers', it is only the value for compliers that is used in the calculation. This calculation respects the original randomisation process, and allows a comparison of the average outcome of the compliers in the intervention arm with the average outcome of the comparable group of 'would-be' compliers in the control arm: the CACE statistic.

CACE analysis data for this trial are based on 2872 standardised value-added scores; these students were nested in classes of 94 treatment arm teachers' classes. Treating compliance dichotomously (as defined above), the fraction of teachers in the SecondarySEEQ study who are compliers is 72%. Using this information, the following calculation estimated the causal effect on those who complied with their treatment assignment by dividing the ITT estimate by the proportion of compliers:

-0.08 / 0.72 = -0.111 (SE 0.018)

Table 7.8: Complier Average Causal Effect (CACE)

Discussion

Because CACE analysis divides the ITT estimate by a value that is between 0 and 1, its value is greater than the ITT estimate. By adjusting the ITT by the proportion of assigned participants who ultimately received the peer-coached intervention, a more refined understanding of the impact may be drawn in as unbiased a manner as possible.

Compliance in this RCT was imperfect, therefore the treatment actually received by participants cannot be ignored in the analysis. Using randomisation as the instrumental variable, the CACE estimate provides a more robust value for the effect of treatment receipt and a clearer understanding of the mediated causal relationship between random assignment and outcome.

Limitations

Drawing on objective evidence of compliance (as above) reduces the uncertainty around understanding the conditions which lead to specified effects, but the CACE analysis used in this thesis is also quite a 'blunt instrument', because it sets a single criterion for compliance or non-compliance. Amongst the group of teachers participating in this trial, there will have been more nuanced types of compliance which, with the inclusion of more precise fidelity-checking data, could have been specified and incorporated into the analyses.

7.25 Mission and reflection texts

Despite there not being a detectable positive impact of the intervention, there is much to be encouraged by from the teachers' own vignettes collected in the mission and reflection texts. Garnered through the SecondarySEEQ website, teachers in the treatment arm were asked to input to the website the objectives and strategies that they decided upon in their peer-coached meeting. Not everyone completed this exercise every time (teachers from 13 of the 15 treatment arm schools chose to complete the mission text sections), therefore generalising findings from these data is impossible, but having an insight into the thinking of teachers in this situation is possible, and it is illuminating.

The full set of texts is found in Appendix 7a, but the overall composition of them reveals teachers' thinking about specific strategies to address specific issues manifest in their feedback. The mission texts show reference to the materials provided, as well as to teachers' own professional judgement, and demonstrate a sense of them being demanding of themselves as much as of their students. While the function of the mission and reflection texts was, originally, a tool to help encourage goal-commitment, they have become an interesting record of teachers' thinking *about* their teaching at points during the process *of* teaching. As an example of this kind of thinking, one teacher wrote the following:

"My intention is to be more consistent in the use of some kind of plenary, getting them to reflect on what they have covered, whether they are aware of what they need to know / be able to do, and how confident they feel regarding their abilities in these respects as a consequence of the lesson. I can then get feedback from them on this, in order to build the following lesson on the progress made by the end of the current one."

This illustrates well the kind of cognitive activation the feedback was intended to generate; in considering the goals and strategies for implementation, the teacher is enacting the cognitive process the consultative feedback intervention was designed to encourage.

The reflection texts (found in Appendix 7b) also provide an insight into teachers' thinking, but this time about the process of having their teaching evaluated, rather than actions to take in light of the evaluation data. As one senior leader wrote in a reflection, apart from the problem of students having to rate multiple teachers on separate occasions (something which cause frustration) the process was "straightforward and the results, as all teachers have said to me, have been enlightening." This last comment highlights a theme that emerged through the reflections: the exercise of having ones teaching evaluated and the

evaluations themselves focused attention on teaching, thinking about improving teaching, and seeing if improvements had an impact on students. As such, the reflection texts represent an interesting set of findings in themselves. They also give great feedback to me as a developer: problems with website access and saving information to it[56] provide excellent foundations for any future work done in this area. Moreover, an excellent criticism of the support materials provided to teachers was presented by one teacher in the comment "Theory [is] perhaps a bit dated in some cases." I think this is a fair assessment, as some of the strategies used came from relatively old source material, albeit material with sound foundations in research and practice.

[56] The website had a time-out function for security purposes. Unfortunately, the site did not recognise typing as an activity which was valid for the purpose of continuing use and would, therefore, lock when the teacher eventually submitted their piece of text, leading to it being deleted. This was hugely frustrating for several people; once it was identified as a problem, new instructions were given to teachers to type their thoughts into a Word document, then to paste them into the website when ready.

Chapter 8: Discussion and summary

Introduction

This chapter discusses findings from the feasibility and main studies; this is followed by summary conclusions. The chapter highlights the applications and limitations of the findings presented in Chapter 7 (addressing the important topic of generalisability) and identifies implications relating both to future methodological and substantive work in the field of student evaluation of teaching. Finally, the chapter assesses the contribution made by this thesis to the field of study, and makes recommendations for future research.

8.1 Summary of results

The main facts and findings from both the feasibility study and main trial projects are summarised below.

The feasibility study

- The feasibility study involved ten volunteer teachers from one urban comprehensive school.
- There was no control group in this study.
- The purposes of the feasibility study were:
    - to explore the feasibility of student evaluation of teaching in a UK secondary school;
    - to gather teachers' views on the instrument proposed for study and use in the main trial;

    - to trial the online survey instrument delivery system;
    - to trial a consultative feedback intervention with teachers.
- 161 students responded to the SecondarySEEQ instrument; 57% of these students were in Year 12 (studying for A1 qualifications) and 43% were in Year 13 (studying for A2 qualifications).
- The average length of time students took to complete the survey was 6 minutes, 53 seconds.
- The subjects represented by teachers participating in the feasibility study were:
    - Biology
    - Chemistry
    - English
    - ICT
    - Modern Foreign Languages
    - Mathematics
    - Physics
- Students completed the survey online and teachers were presented with their feedback by me (taking the role of peer-coach); an hour-long meeting then took place in which interpretations of the data were discussed.
- There was strong support from the participating teachers for the mission of using SETs to improve teaching.
- Focus group 'talk aloud' exercises conducted with students revealed a sound understanding of the meanings of items used in the instrument, as well as

support for the principle of improving teaching which lay behind the whole exercise.

The main study

- The main study was a cluster randomised controlled trial of a consultative feedback intervention for teachers 250 teachers in 36 schools.
- Random allocation at the cluster (school) level led to a balanced design in which 18 schools were in the treatment arm, and 18 in the active control arm.
- Average GCSE scores were the baseline measure used for randomisation.
- Randomisation was conducted by an independent statistician.
- Five schools dropped out before the intervention commenced (2 in the treatment arm, 3 in the active control arm).
- Multilevel models showed that the intervention (a termly peer-coached consultative feedback meeting in which the topics of discussion were the SET results and actions to take in light of them) had no detectable effect on the outcome of interest (student progress as measured by standardised value-added scores).
- Analysis of data from the rest of the ALIS cohort over a period of 6 years demonstrated that the progress made by those students in the treatment arm was not different to the average progress made by all students.

8.2 Discussion

Both the feasibility study findings and those from the main study offer several points of real interest, both from methodological and substantive points of view, but also from the pragmatic and functional aspects of innovating ways to help teachers improve their practice.

Accepting that there will always be bias present – and particularly so when volunteers participate in studies – the fact that students and teachers involved in both the feasibility and main studies showed enthusiasm and support for the endeavour gave weight to the claim that there is an interest in garnering from students' feedback about the teaching they receive. From the early focus groups (Appendix 3b) to the mission texts and reflection texts captured through the SecondarySEEQ website (Appendices 7a and 7b), I saw clear evidence of a desire (from students) to give feedback and to receive feedback (from teachers). Capturing this interest, I built trusted relationships with schools and their teachers in order to gather data and to create an appropriate climate for the study to take place. Such a 'climate' needed to encourage a desire on the parts of students and teachers to give and receive feedback – and to act upon it in the case of those in the treatment arm – while maintaining a low level of threat to teachers' self-esteem and confidence. This was achieved successfully.

8.3 Validity of the SecondarySEEQ instrument

One of the foci of this thesis was the validation of the SecondarySEEQ instrument for use in a novel context and for the purpose of providing constructive feedback to teachers. In respect of this, the multilevel confirmatory factor analysis presented in Chapter 7 indicate that the a priori foundations of the instruments forebear (the SEEQ)

are replicated and, thereby, it is reasonable to conclude that the construct of 'effective teaching' is measured by the set of latent variables represented by the items and their associated factors. Yet, as Marsh (2007c) himself points out, SETs are challenging to validate as they represent only a single measure of teaching effectiveness and this, in isolation, is insufficient for the purpose. If the aim is to validate the SecondarySEEQ for the purpose of collecting information about the effectiveness of teaching, a wider range of criteria than simply student learning should be included (according to Marsh (1987)); student progress, changes in student behaviours, teacher self-evaluations, evaluations by others (such as line managers) all fall into this category. Alongside the standardised value added (STDVA) measurements used in this project, one of the most powerful sources of evidence is the set of narratives provided by teachers in the mission and reflection texts submitted to the SecondarySEEQ website (and presented in full in Appendices 7a and 7b). Multilevel CFA data show that the instrument's items perform generally well, and that students within classes have a degree of consensus about the effectiveness of the teaching they received, but the affective narratives submitted by teachers add weight to the argument that the SecondarySEEQ instrument is a valid measure of teaching effectiveness, but only for the purpose prescribed in this study (informing a peer-coached feedback consultation discussion with a colleague). To go any further than this – to the realm of making relative judgements about quality of teaching so as to rank teachers, promote teachers, or to fire them – is not supported by the analysis of the data presented in this thesis.

By providing insight into the SecondarySEEQ's dimensionality, it can be concluded that the hypothesised structure of the scale – that it would reflect the a priori structure of

the SEEQ – is identified in the multilevel confirmatory factor analysis, thereby lending support to the validity of the instrument. This is a plausible interpretation derived from three key sources of evidence: the a priori work on SETs and, particularly the SEEQ; the confirmatory factor analysis; and the narrative data collected by way of the mission and reflection texts.

8.4 Discussion of impact evaluation findings

The lack of a detectable effect size from the intervention may seem disappointing, but I do not view it as such: null and negative effects are, in my opinion, as important a finding as any other and must be reported transparently. My aim was to conduct a study of the highest quality of which I was capable and I believe I have succeeded in this endeavour; reporting the findings of the study and providing an accurate commentary on this were the goals of this thesis.

There is a fine line to be walked when trialling interventions in schools. To be efficient and effective means to use only those resources (chiefly time and money) which are needed in order to achieve a desired effect. In planning the intervention studied in this thesis, I think I erred too much on the side of cautious efficiency (the expectation of teachers committing around one or two hours per term was, in reality, too small a commitment for sustained changes in teaching practice to take hold and have an impact on student attainment), rather than the needs of effective implementation. In walking that fine line, I had to ensure that I could find and retain volunteer teachers; a key component of this was the limited extra burden on time and resources in school, though it may also have been a key limitation. Nonetheless, acknowledging this limitation offers

a point of departure for future work, work which should seek to build on and augment the findings of the SecondarySEEQ project.

In order to understand more fully impact of the intervention more fully, attention is now turned to a discussion of the limitations of the study.

8.5 Limitations

In discussing the limitations of the findings in the main study, it is important – as per point 20 ('Limitations') of the Consort 2010 guidelines to address sources of potential bias, imprecision and multiplicity of analyses, where appropriate (Schulz et al., 2010). Earlier in this thesis (Chapter 5), I discussed potential sources of bias with a view to using that exercise as a means of mitigating bias. Here, I go a step further by discussing actual sources of bias and how these impact on the study's findings. I also address the imprecision of the findings, and the multiplicity of analyses.

Bias is unavoidable and, even with the 'gold standard' of experimental design (the RCT), assuming that bias has been eradicated is illogical. While a randomised design can control for differences in observable and unobservable variables, it cannot remove all bias completely. In this study, there was attrition (five clusters: two from the treatment arm, three from the control arm) and, as such, this is a potential source of bias. This was counteracted by the use of Intention To Treat (ITT) and complier average causal effect (CACE) analyses, the intention being to retain the integrity of the trial by analysing data from those clusters that were randomised, as well as data for those teachers for whom there was evidence of compliance with assigned condition.

There is no evidence in any of the quantitative or qualitative data that there was contamination of the active control or treatment conditions; precautions were taken to prevent this and I conclude that these were successful in their aim.

The effect size findings (reported with confidence intervals in Chapter 7) represent the best estimate of the intervention's effect, but nothing more than this. Confidence intervals representing the plausible range of values for the estimate aid interpretation, but increase imprecision by virtue of their representing a normal distribution of values. It has to be accepted, therefore, that there is a compromise to the reporting of these data which limits the precision of the findings (Gorard, 2014). Nonetheless, this is the most appropriate means by which to report the findings in this thesis.

Finally, this study only looked at very short-term effects on student attainment (over one academic year); the intensity and duration of the intervention limited the study considerably, so future work in this area should be undertaken over a longer period.

8.6 Generalisability

Due to the recruitment of volunteer schools and teachers to the SecondarySEEQ project, the sample involved in this doctoral study was not representative of all schools and colleges offering Advanced Level and Advanced Subsidiary Level qualifications. Clusters randomised were all involved in the ALIS project and volunteered to participate in this PhD study; these two factors mark out the schools involved as different to the rest of the population of schools and colleges. At the teacher level within clusters, only volunteers participated, a feature which compounds the problem of drawing generalizable conclusions. Additionally, when one considers the fact that the main study

was powered to detect a minimum detectable effect size of 0.36, the security of the findings and, therefore, the ability to draw generalizable inferences from them, comes very much into question.

The lack of a detectable effect size may be attributable to the study being underpowered, but it is also wise to consider the active control group as a problematic feature. The decision to present control arm teachers with a feedback graph was taken as a way of trying to encourage participation and minimise attrition; the decision was a sensible one but came at a cost. A 'business as usual' control group would have created a better point of comparison and possibly made a detectable effect size more recognisable. That said, if the data from all ALIS schools in the 2009 – 2015 period provide an indication of progress without the presence of the intervention under examination in this thesis, it may be concluded that the active control treatment had little or no effect on student progress. This adds weight to the notion that the intervention evaluated was not sufficiently intensive to have a detectable effect, a finding that is very useful for future work.

The final limitation is the intensity of the intervention. As has already been mentioned, the tension between providing an efficient but also effective intervention led to a one hour per term (three hours in total) peer-coached consultative feedback meeting being held; on reflection, this was most likely too limited in its intensity to have anything other than a minimal effect.

8.7 Methodological implications for future work

When I began this study, the Education Endowment Foundation was in its infancy; the number of randomised controlled trials being conducted in education in England was far smaller than it is now. As education systems become more interested in measuring impact, the development of improved methodology for applied settings is necessary. It is in light of this context that the methodological implications for future work are here considered.

Implication 1: the use of randomised experiments in education should be a systematic part of and evidence-based school system in the UK

I have learned that doing randomised controlled trials in schools is feasible; I would add to this that it is desirable, especially if we are to concern ourselves with understanding the impact of interventions on valued student outcomes. If causal inference is the aim, randomisation must feature in the research design. That said, RCTs can be a very blunt instrument, and in their raw form can promulgate the sense of research being 'done to' participants. In my study, I set out with a 'by teachers, for teachers' mentality, and the collaborative aspect of this was absolutely crucial to my work; while teachers played no part in the design of the trial, I communicated with them and listened to them very much from the point of view of one learning how to make things as easy, attractive, social and timely (Service et al., 2014) for participating teachers.

Implication 2: collaborating with teachers and school leaders on study design and intervention delivery is an important component of school-based research

Since beginning this PhD, there have been interesting methodological developments from Education Endowment Foundation-funded trials to which I wish to draw attention

and, in doing so, refer back to my own work and synthesise several points to make one recommendation. See, Gorard, and Siddiqui (2015) make recommendations about improved communications with research subjects, as well as the provision of training for developers and others involved in the delivery of an intervention (to ensure fidelity); Siddiqui, Gorard, and See (2015) make specific mention of the role teachers can play in trial design and implementation. Notably, the latter authors state that teachers are capable of conducting evaluations of their own practice, when provided with suitable guidance.

I feel fortunate that my own background in teaching was, I believe, instrumental in gaining the trust and compliance of teachers involved in this project. For those researchers working in education who do not have such a background, there needs to be better guidance on working effectively with schools if they are to make the most of the opportunity to do so.

In uniting a background in teaching with training as a researcher, it as become clear that there is much to do in refining the methods used to generate useful, high-quality evidence in an education system. While there is clearly a place for the large, well-funded evaluations of kind funded by the Education Endowment Foundation, there is also a gap to be filled by aggregating small-scale 'DIY' trials of the kind advocated by Coe, Kime, Nevill, and Coleman (2013)

Implication 3: closer working relationships between developers and evaluators may lead to better intervention development

In the SecondarySEEQ project, I was both developer and evaluator, a combination of positions which should not, ordinarily, be recommended. Despite the potential conflict of interest, I was very clear from the outset that the quality of the evaluation was more important than the finding of the evaluation; as such, the developer in me had to listen to my inner evaluator and resist the temptations to tinker with the intervention. While I would not suggest that all developers be trained as evaluators, I think there is great merit in there being a clearer understanding between the two and, potentially, some form of training for developers working under the conditions of an evaluation.

Planning and implementation are two significant factors in the success of both an evaluation and the development of an intervention. I would argue that, if any education system is to maximise the time, effort and money put in to such activities, greater understanding by evaluators and developers of the exigencies of each other's work is needed. If, for instance, developers are more aware of the need for good outcome measures, or evaluators are more aware of the commercial and practical implications associated with randomised trials, closer and more effective working relationships may be possible.

8.8 Substantive implications for future work

The measurement of students' perceptions of teaching quality is challenging and imprecise, but it is also possible and useful. In light of the work undertaken during the course of this doctoral study, there are three main substantive implications for future work around student evaluation of teaching:

Implication 1: work should be done to develop the SecondarySEEQ items

The items used in the SecondarySEEQ instrument were changed from their original SEEQ wording only minimally[57]. Their wording has its origins in higher education and, chiefly, universities in the USA; more work to refine existing items and create new ones specifically for UK schools would be useful. Increasingly, teachers are interested in using surveys to collect feedback from students (Fauth et al., 2014), but there is often little known about the psychometric qualities of such instruments. All of this indicates that there is a warrant for work which seeks to provide good tools to help teachers gather the information they need.

Implication 2: work should be done to create interventions which use SET data to improve teaching

Alongside work to increase the precision of a student ratings instrument designed for high school students, work should also be done to develop interventions based on SET feedback. By building a peer-coached consultative feedback intervention which is more intense, and which aligns with the best available evidence on teacher learning (S. E. Higgins et al., 2015; Timperley, 2008), the potential for a greater effect on student attainment will increase.

Implication 3: a multiple measures approach to evaluating teaching quality similar to the MET Project should be trialled in the UK

Following in the footsteps of the MET Project (Kane et al., 2013), it would be useful to incorporate student evaluation of teaching data, student progress data (value-added)

[57] Herb Marsh was keen for the original SEEQ items to remain intact wherever possible, and the time available for trialling and creating new items was not sufficient for significant changes to be made. Amendments were made to make the context of A1 and A2 study clearer to students using the instrument.

and classroom observations by trained raters to assess the technical capacity and social value (Newton & Shaw, 2014) of a suite of multidimensional measures of teaching effectiveness. In light of the relative weaknesses inherent in all of these measures, caution is urged in the uses to which such data are put. Used as a starting point for constructive, development-focused conversations amongst teachers, feedback from these multiple measures could provide a powerful tool; this would be my recommendation. What would not be recommended at this point, however, would be use of the data for high-stakes accountability purposes: ranking teachers; performance-related pay; recruitment and retention and such like. While these latter functions may be desirable to some – specifically those working in education policy – there is no warrant to use them to make high-stakes judgements about individual teachers.

8.9 The contribution of this thesis to the field of student evaluation of teaching research

This doctoral study makes two key contributions to the field of student evaluation of teaching: the validation of an instrument designed to collect high school students' perceptions of teaching; and the design and evaluation of an intervention which brings together both diverse theories in a pragmatic application.

From a theoretical perspective, this thesis has built on the work of Remmers and Brandenburg (1927), Marsh (1982), Fauth et al. (2014) and others, but has also brought together goal theory (Locke & Latham, 1990), feedback (Hattie & Timperley, 2007; Penny, 2004), solution-focused brief therapy (Corcoran & Pillai, 2007; Kim, 2007) and teacher professional development (Timperley et al., 2007) to study a pragmatic, feasible in-school intervention. As has been suggested in previous chapters,

there is much promise to be realised in programmes which offer feedback to teachers based on high-quality data derived from student perceptions; in order for student evaluation of teaching truly to be a fruitful and powerful tool for the development of teaching, future research must continue to build on the foundations laid both *before* this thesis and *in* this thesis.

Appendices

Appendix 1a

Student Evaluation of Teaching: can it raise attainment in secondary schools? A cluster randomised controlled trial

Trial Protocol

Summary

Age range of students	16 - 18 (Years 12 and 13)
Number of schools	36
Number of teachers	250
Design	Cluster Randomised Controlled Trial (CRCT)
Primary outcome	AS Level and A Level

Significance and rationale

The use of questionnaires for student evaluation of teaching (SET) in universities is widespread; plentiful and high-quality evidence indicates that, when appropriate instruments are used to collect SET data, and ratings are fed-back to teachers (generally using a consultative feedback approach), SETs provide a valid and useful multidimensional measure of teaching quality the use of which can substantially enhance future SET ratings and attainment (Marsh, 2007c). In their combination of meta-analyses of SET instrument use, Wright and Jenkins-Guarnieri (2012) identify that the impact of SET raptings data is often increased when some form of feedback consultation is incorporated; the most recent of these meta-analyses (Penny & Coe, 2004) estimates consultative SET feedback (that which is supported by dialogue with someone acting in the capacity of consultant) to have a *d* of 0.69. Previous meta-analyses (P. A. Cohen, 1980; L'Hommedieu et al., 1990) also found positive effects, though these were smaller (Cohen recorded 0.38 and L'Hommedieu et al recorded

0.342). It is worth noting that these previous studies acknowledged that the actual effect may have been larger, had they not been attenuated by validity issues.

Despite the strength of the evidence and the range of higher education contexts in which consultative feedback based on SET data has been used, there are no existing studies which have attempted to adapt this intervention for, and evaluate it in, secondary school settings. This doctoral study will attempt to do exactly that by trialling and evaluating the well-defined Student Evaluation of Educational Quality (SEEQ) instrument (Marsh, 1982). Support for such a venture is drawn from S. E. Higgins et al. (2013); the *Sutton Trust / EEF Teaching and Learning Toolkit* identifies feedback as among the most promising strategies for supporting learning, and (Hattie & Timperley, 2007) find feedback to be among the most significant documented influences on achievement, with an average *d* of 0.79 (where the average effect size for all interventions was 0.40).

Research plan

Research questions

The primary research question is: can student evaluation of teaching feedback coupled with a consultative feedback intervention raise the attainment of AS and A Level students?

The secondary research question is: is an adapted version of the Marsh (1982) Student Evaluation of Educational Quality (SEEQ) instrument a valid tool for collecting SET data in the novel context of UK secondary schools?

Design

This project will use a cluster randomised controlled trial (CRCT) design with a single distance randomisation performed. The trial will be designed and implemented with the guidelines of CONSORT (2010) (Schulz et al., 2010); reporting will, therefore, be in line with these standards.

The trial will use an active control group. While it is acknowledged that a control group receiving 'business as usual' would offer a greater opportunity to detect a difference between the two conditions, this is not feasible with this trial. The incentive of receiving some form of feedback (as the control group will) is needed in order to recruit and retain those schools randomised to the control condition.

The intervention

The intervention is a consultative feedback package which uses student evaluation of teaching feedback as the basis of three termly discussions held between a teacher and a colleague (their nominated peer coach). SET feedback will be delivered online to teachers via a secure website.

Teachers in the control arm of the trial will receive feedback in the form of a simple line graph. Those in the intervention arm will receive a chart, a set of downloadable teaching strategies to help inform how they respond to specific dimensions on which their performance is low (relative to the other dimensions), and a guide to the consultation meeting to be held with the peer coach. This guide takes the form of a video which is designed to give information in a uniform manner to all participant teachers, while also focusing the discussions with peer coaches in a constructive

manner. The video asks a set of questions for the teacher to answer and encourages them to discuss their answer and seek feedback from their peer coach. The meeting guide suggests that the peer coach meeting should last around one hour.

Inclusion criteria

Schools already registered with the Centre for Evaluation and Monitoring (CEM) to use the A Level Information System (ALIS) system in the 2012-2013 academic year are eligible for inclusion. A convenience sample will be drawn from this population by means of self-selection: all schools participating will do so voluntarily.

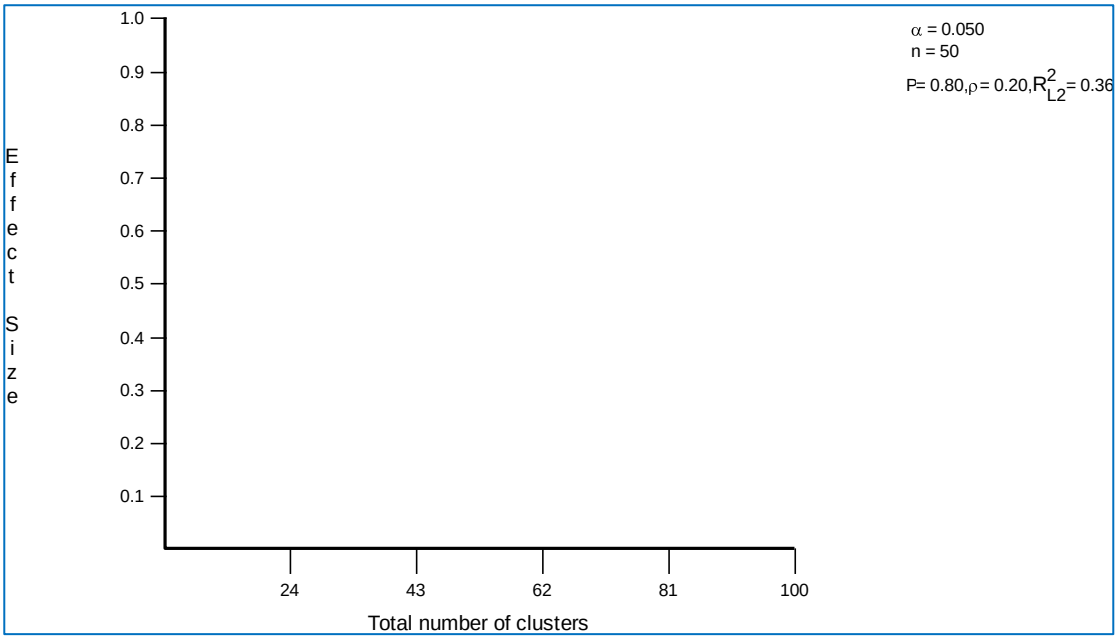
Randomisation

Stratified blocking will be used in a distance randomisation procedure executed by Dr Adetayo Kasim (Research Statistician at the Wolfson Institute for Health and Wellbeing). All schools will be given a numeric identifier to prevent Dr Kasim from knowing their identity, and to reduce bias in subsequent data handling.

Sample size calculation

It is estimated that an adequately-powered trial would require a sample of 48 schools. Assuming random allocation of schools with fifty individuals in each, and an ICC of 0.2, the availability of covariates with a multiple correlation of 0.6 with the primary outcome measure, a Type-1 error rate (alpha) of 0.05 and power set to 0.8, a 0.3 effect size could be detected with 48 schools.

The graph below (created using Raudenbush's Optimal Design Software[58]) shows the relationship between the minimum detectable effect size (MDES) and the number of schools with the above assumptions:



Analysis

The impact of the intervention will be analysed using an effect size analysis (Hedges *g* (Hedges, 2007)) which uses a pooled standard deviation in the following calculation:

$$\text{Effect size} = \frac{[\text{Mean of the intervention condition}] - [\text{Mean of the control condition}]}{\text{Standard deviation}}$$

The primary outcome will be student attainment as measured by CEM's ALIS value-added scores at Advanced Subsidiary (AS) Level and Advanced Level.

The main analysis combining pre-test and post-test data will use Intention To Treat (ITT) analysis: schools' AS and A Level data will be analysed as they were randomised,

[58] http://hlmsoft.net/od/ accessed 02/11/11

thereby reducing the bias created by non-random attrition. Both unadjusted and adjusted results will be presented, the latter being useful due to the presence of an important predictor of AS and A Level outcomes. The adjusted analysis should result in a smaller standard error and smaller confidence intervals.

Multilevel models will be created using Mplus (Muthén & Muthén, 1998 - 2014) in order to account for the clustering of the data.

Using compliance data available for both the treatment and control groups (drawn from SecondarySEEQ website metrics relating to online activity) Complier Average Causal Effect (CACE) analysis (Gerber & Green, 2012) will be used to provide an unbiased estimate of the treatment effect on those teachers who complied with their assigned condition. Satisfying the five assumptions required for a valid causal interpretation of the CACE statistic (Angrist et al., 1996), the following calculation will be used:

CACE = ITT/proportion of compliers

Additional analyses of implementation and take-up will be conducted using data about logins, page views and downloads from the secure website that will deliver the feedback to teachers.

Process evaluation

The purpose of the process evaluation is to develop an understanding of the experience teachers have of participation in a process that sees their students rate

their teaching. Teachers will be asked to submit 'reflection texts' to their secure website, detailing their experiences and perceptions.

Developer

The intervention developer is Stuart Kime, a PhD candidate in the School of Education at Durham University.

Trial personnel

The project will be conducted in its entirety by Stuart Kime, a PhD candidate in the School of Education at Durham University.

Data protection statement

This trial will adhere to Durham University's Data Protection Policy, which is available at https://www.dur.ac.uk/data.protection/policy/

Month	Activity
• December 2011 - February 2012	• Review SET literature and research design • Confirm data collection arrangements
• March 2012	• Begin recruitment
• May 2012	• Pilot trial
• September 2012	• End recruitment
• October 2012	• Randomisation • Send website logins to schools • Teacher self-efficacy survey 1

• November 2012	• SET data collection phase 1 • Feedback delivery phase 1
• March 2013	• SET data collection phase 2 • Feedback delivery phase 2
• June 2013	• SET data collection phase 3 • Feedback delivery phase 3 • Teacher self-efficacy survey 2
• December 2013	• Collect post-test data
• January 2014	• Impact analysis
• March 2014	• Process evaluation

Risks

Risk	Assessment	Countermeasures
Not enough schools recruited to power the study as per the design	Likelihood: high Impact: high	• Ensure a frank but attractive offering is made to schools. Subsequent communications must be clear and encourage retention

Difficulties accessing the secure website	Likelihood: moderate Impact: high	• Provide clear instructions to teachers • Provide responsive email and telephone support • Work with web and database developers to mitigate likelihood and impact
Poor responses to data collection phases 2 and 3	Likelihood: moderate Impact: high	• Remind teachers of upcoming deadlines for data collection • Remind school coordinators to be active in encouraging teachers to arrange for students to complete evaluations
Illness of the researcher	Likelihood: moderate Impact: high	• Automate as many processes as possible • Work with database developers to create back-up plans
Poor fidelity to the planned intervention	Likelihood: moderate Impact: high	• Maintain regular contact with teachers in the intervention arm and

		advise on actions to be taken
IT systems failure	Likelihood: low Impact: catastrophic	• CEM systems are well-maintained to commercial standards • Close links will be maintained with CEM IT staff

References

Cohen, P. A. (1980). Effectiveness of Student-Rating Feedback for Improving College Instruction: A Meta-Analysis of Findings. *Research in Higher education, 13*(4), 321 - 341.

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research, 77*(1), 81-112. doi: 10.3102/003465430298487

Higgins, S., Katsipataki, M., Kokotsaki, D., Coleman, R., Major, L.E., & Coe, R. (2013). The Sutton Trust-Education Endowment Foundation Teaching and Learning Toolkit. In E. E. Foundation (Ed.). London.

L'Hommedieu, R. , Menges, R. J. , & Brinko, K. T. (1990). Methodological Explanations for the Modest Effects of Feedback From Student Ratings. *Journal of Educational Psychology, 82*(2), 232 - 241.

Marsh, H. W. (1982). SEEQ: A reliable, valid, and useful instrument for collecting students' evaluations of university teaching. *British Journal of Educational Psychology, 52*, 77-95.

Marsh, H. W. (2007). Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases and Usefulness. In R. C. Perry & J. C. Smart (Eds.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*: Springer Netherlands.

Penny, A., & Coe, R. (2004). Effectiveness of Consultation on Student Ratings Feedback: A Meta-Analysis. *Review of Educational Research, 74*(2), 215-253.

Schulz, K. F., Altman, D. G., & Moher, D. for the CONSORT Group. (2010). CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *Annals of Internal Medicine, 152*.

Wright, Stephen L., & Jenkins-Guarnieri, Michael A. (2012). Student evaluations of teaching: combining the meta-analyses and demonstrating further evidence for effective use. *Assessment & Evaluation in Higher Education, 37*(6), 683-699. doi: 10.1080/02602938.2011.563279

Appendix 2a

Marsh's Student Evaluation of Educational Quality instrument (Marsh, 1982)

University of Western Sydney

Student Feedback on Teaching (SFT)
SEEQ Survey[1]

Student evaluation is one of the methods used for improving the quality of teaching at the university. This survey will provide this staff member with valuable feedback about teaching effectiveness. Your name is NOT required and all information is confidential. Please complete as accurately and honestly as possible. You should base your responses on this staff member's teaching in this unit.

STAFF MEMBER'S NAME:

UNIT / CLASS: DATE:

INSTRUCTIONS
- Use a blue/black pen or 2B pencil
- Do NOT use red pen or felt tip pen
- Erase mistakes fully
- Make no stray marks

Please MARK LIKE THIS ONLY:

Please indicate the EXTENT of your agreement / disagreement with the following statements as descriptions of this unit by using this scale:

Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree
1	2	3	4	5	6	7	8	9

NOTE: Leave blank any items that do not apply for this staff member's teaching.

LEARNING / ACADEMIC VALUE
- You found the class intellectually challenging and stimulating.
- You have learned something which you considered valuable.
- Your interest in the subject has increased as a consequence of this class.
- You have learned and understood the subject materials in this class.

STAFF MEMBER'S ENTHUSIASM
- Staff member was enthusiastic about teaching the class.
- Staff member was dynamic and energetic in conducting the class.
- Staff member enhanced presentations with the use of humour.
- Staff member's style of presentation held your interest during class.

ORGANISATION / CLARITY
- Staff member's explanations were clear.
- Class materials were well prepared and carefully explained.
- Proposed objectives agreed with those actually taught so you knew where the class was going.
- Staff member gave presentations that facilitated taking notes.

GROUP INTERACTION
- Students were encouraged to participate in class discussions.
- Students were invited to share their ideas and knowledge.
- Students were encouraged to ask questions and were given meaningful answers.
- Students were encouraged to express their own ideas and / or question the staff member.

INDIVIDUAL RAPPORT
- Staff member was friendly towards individual students.
- Staff member had a genuine interest in individual students.
- Staff member made students feel welcome in seeking help / advice in or outside of class.
- Staff member was adequately accessible to students during office hours or after class.

BREADTH OF COVERAGE
- Staff member contrasted the implications of various theories.
- Staff member presented the background or origin of ideas / concepts developed in class.
- Staff member presented points of view other than his / her own when appropriate.
- Staff member adequately discussed current developments in the field.

EXAMINATIONS / GRADING
- Feedback on assessments / graded material was valuable.
- Methods of assessing student work were fair and appropriate.
- Assessments / Examinations tested units content as emphasised by staff member.

ASSIGNMENTS / READINGS
- Required readings / texts were valuable.
- Readings, assignments etc. contributed to appreciation and understanding of the unit.

OVERALL RATING
(1= Very Poor 3= Poor 5= Average 7= Good 9= Very Good)
- Overall, how does the class compare with other classes at this institution?
- Overall, how does this staff member compare with other staff members at this institution?

1 Students' Evaluation of Educational Quality SEEQ ©2002 H.W. Marsh

Student Feedback on Teaching

University of Western Sydney

Teacher

Unit ..... Date ...../...../.....

Open Ended Comments

Your teacher will not see your written comments until after your exam results have been finalised. If you have a concern for your anonymity, please print your comments.

Please indicate the important characteristics of this lecturer/ class that have [illegible] your overall learning experience.

1.

2.

3.

Please indicate the particular areas of this lecturer/class that you feel are important for him/her to improve.

1.

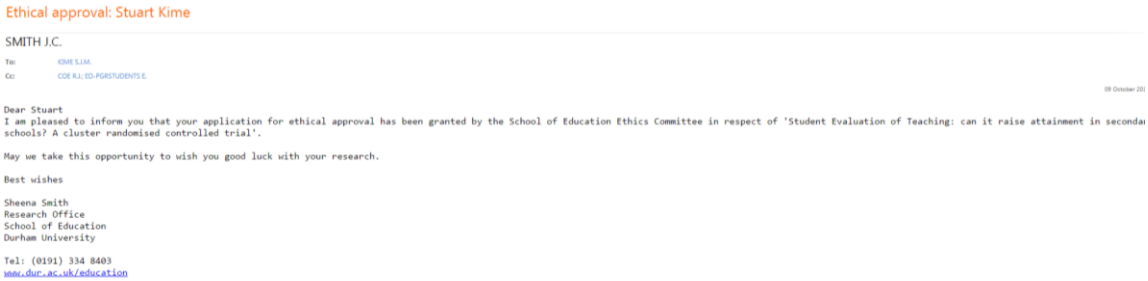
2.

3.

NOTE: Please feel free to write on the back of this sheet if you wish to clarify any of your responses or to make additional comments. Thank you.

UWS Educational Development Centre

Appendix 3a Ethics approval



Ethical approval: Stuart Kime

SMITH J.C.

To: KIME S.J.M.
Cc: COE R.J.; ED-PGRSTUDENTS E.

09 October 201

Dear Stuart
I am pleased to inform you that your application for ethical approval has been granted by the School of Education Ethics Committee in respect of 'Student Evaluation of Teaching: can it raise attainment in secondar schools? A cluster randomised controlled trial'.

May we take this opportunity to wish you good luck with your research.

Best wishes

Sheena Smith
Research Office
School of Education
Durham University

Tel: (0191) 334 8403
www.dur.ac.uk/education

Appendix 3b

Note: interviewer's text is in *italics*.

This is a 'talk aloud' exercise. Talking aloud definitely encourages you to say whatever you want to say about any of these things. Having an opinion is good. To begin with here is a bit of background to this project. This is something that has been in trial in Universities as a survey of students providing feedback to their teachers on their teaching. What I am going to show you now is the set of statements that the students have been given and with each of those statements they are asked to strongly disagree, disagree, be neutral about, agree or disagree about. All I am going to do is to ask you what you think each of these statement means. And then I may ask you why you said so. That's basically it.

1. *I found this teacher's lessons intellectually challenging and stimulating. What does it mean?*

Student 1: It means that you are feeling pushed by the teacher's way of teaching. The questions aren't too easy and straightaway but also you can answer the questions when given extra information if needed.

Anybody wants to add to that?

Student 2: It means you find it interesting and engaging and encouraging and you would like to keep on learning it.

Student 3: You are engaging more with the syllabus like Student 2 said but not in a didactic way but you are interacting with the teacher and you get more knowledge than is necessary just on the syllabus.

2. *This term I have learned some valuable things in this teacher's classes.*

Student 4: I think it means that this one is learning about the subject, learning all the things you like to know about the exam and also means learning things may be about yourself. For example, how you learn you might have learned more independently or ways of learning differently which you might have not known how to do before.

Student 1: Again I think it is just things outside of the syllabus which is more interesting for you to acquire knowledge around the subject.

3. *My interest in the subject has increased this term as a result of this teacher's lessons.*

Student 2: It means that the teacher is an interesting person and an interesting teacher and has a good method of teaching and the students value the teacher as a whole. So they feel without the teacher they would be worse off.

So is it something like setting that teacher apart in some way rather than saying that if your interest is increased as a result of that teacher's lessons then that teacher is a good teacher?

Student 2: Yes.

Student: I think it is about whether your teacher is engaging with your interests so that you can form more interest in your subject in general. It shows if the teacher is exploring your lines of thought rather than just having a prescribed sort of thing.

And is that important?

Student: Well for my subject it depends on the kind of essay arising out of those thoughts and how we can develop certain lines of enquiry.

Any other thoughts on that?

4. *I have learned and understood what my teacher has taught me this term.*

Student 2: I think it means the clarity of your teacher is quite good in general and you have understood it and learnt it at the time it was taught to you but not necessarily you remember what you were taught. Or that you feel confident about taking?

Is there a difference between learning and understanding?

Student 4: I think understanding is different from learning. Say for example if you learn the method it is good but if you understand it, it means you can do it without having to go through the?? Of what you learnt through a set method.

Student 5: Another thing in understanding is that you can explain ?? when you actually understand it. You can teach …

So if you have learned it, you know it, and if you understand it, you can teach it.

5. *This time my teacher was enthusiastic about teaching my class.*

Student 1: If a teacher isn't enthusiastic, then you don't end up grasping as much as when a teacher is enthusiastic and has passion about their subject because they all tend to sort of take up small tangents and give you more interesting bits of side information. Whereas a teacher who just teaches the syllabus can really de-moralize students.

So what is the value on those tangents there?

Student 1: You can remember them and apply them to further questions in maths and ?? subjects. They are generally interesting really and keeps you interested. They are useful in pub quizzes.

Anything else?

Student 2: I think it means whether you value your if your teacher enjoys teaching you and think if the teacher is enthusiastic and approachable as well. You feel you can ask something and talk to them all about problems you might have or things like that.

Student 3: Well you look for things like if your teacher is enthusiastic or engaging in debate about the subject so that you can learn a bit more. However, there is also the fact that they are not just willing to just say or just you have to know lessons(?) and hat s just you need to know.

6. *My teacher was energetic and dynamic while teaching my class.*

Student 4: I don't think this necessarily means that the teacher is jumping up the walls while teaching you. Basically, they are enthusiastic and passionate about their subject. They are not just standing there dictating you like a PowerPoint, they are actually giving you activities that stimulate you and help you learn more independently as opposed to just lecturing.

Student 2: I disagree with that. I would take that to be that the teacher was physically moving around a lot and may be changing his tone and voice or moving hands a lot or pacing about- something to do with movement.

When you read this, does it suggest somebody who is just enthusiastic and passionate about the subject or does it suggest something else such as…?

Student 3: I think you can manifest it in various ways. If you have a teacher who is genuinely interested in his subject and uses a lot of humour to help you learn more and you can engage in a debate about your subject. That is a good thing but if you have got someone who just bounces around the classroom it may be hard to follow which direction he is going. I don't think energetic and dynamic is necessarily the best thing

because [?? calmest environment] but I suppose it depends on the kind of teachers you have. If you have teachers who are really solid about the subject and are going to get that across then if they are dynamic and energetic that might help. Otherwise it is not particularly helpful.

Student 5: I think it depends on the various ways of teaching…. So one does not necessarily have to be a PowerPoint person or do research on the Internet???

7. *My teacher enhances his power of teaching with the use of humour this term.*

Student 1: I think that if the teacher uses humour to help people learn things quite often using anecdotes or things like that, they can make you remember things more clearly.

Student 3: I think it is really helpful if the teacher engages with the class and use a lot of humour as a method of doing that. It can be a good thing in the sense that you have got a good teacher student relationship going on. But you can also use humour to gloss over issues that might have greater connotations for different sorts of people. It is not necessarily the best thing. You might as well get more serious in terms of looking at it [?? or] thought process. If you are just using humour to explain, it might not be the best idea.

So you are saying that humour could be a quite useful tool only when used judiciously or appropriately and not to mask things?

I don't think it should be used to soften the blow when you are talking about sensitive issues or if you are talking about something you can engage in a debate with, it is not necessarily something that people might find morally and ethically upsetting them. Otherwise it is ok.

8. *My teacher's style of teaching held my interest during lessons this term.*

Student 2: I think this means the style of your teacher complements the way you learn.

So are you suggesting that the teacher should adapt their way of teaching to the individuals in front of them or changing it accordingly?

Student 2: I think it is considerable of a teacher to adapt the style of teaching to the preferences of the majority of their students. But then you will get teachers who are not naturally like that anyway.

So do you think it is natural for some teachers to adapt themselves like a chameleon to the preferences of students?

Student 2: Yes, there are teachers who can adapt and there are others who can't adapt but they may be appropriate in their way if teaching as well.

Student 3: I think it means that your teacher is engaging with the student and so it is more about the student and teacher conversation rather than making notes or reading from the book or handing out sheets. Holding interest is denoted by active and lively conversations rather than going out and finding things on one's own.

9. *My teacher's explanations were clear in my lessons.*

Student 1: I think it means that you have understood exactly what the teacher is trying to tell you and there is no two ways about it.

Student 2: And also the pace was neither too quick nor excessively slow. Although it is good to be a little slow if you are explaining things for the first time so that they have time to take it in and understand. Also the voice and intonation of the teacher is appropriate and they use appropriate language and technology in what they are teaching.

So you are saying it is about the practical thing about altering pitch and pace and volume and tone of what you are trying to say.

Yes.

Student 3: I think it is mainly an issue of time really because if you want to make something clear, you need to go on a sensibly slow pace so that the students understand every aspect of the issue. And if your teacher thinks that the students are now in University and should therefore know, and we don't need to explain in detail but just show a slide, then that is not really the best way to explain. They should adopt a style where the majority of people understand the major concepts of what they are trying to explain.

Interesting that you have used the word majority there, you have also used it before. It is a bit of a tangent, the majority part, but what about everybody? Everybody understanding and everybody being able to access these things, or should the teacher be concerned with only the majority and assume that there will always be some who …

Student 2: I think you can do both. When you are teaching a full class you cannot please everyone because you are not talking to just one person and to each one individually, but a group as one entity of people. So there the best thing to do is to keep in mind the majority. But after that it should be great if the teacher goes around individually and explains on a one to one basis…

Student 3: I agree with Student 2 said. I think teachers have to be pragmatic most of the time. If you try to reach everybody and ensure that they have had a crystal clear understanding, we would never go anywhere. And so I think it is a clear explanation of most people understanding what is being said. Those few who can't follow can always learn from the people around but if a teacher has a clear explanation that usually carries

out of the classroom as well. If you go and see the teacher to clear something up afterwards that probably is a part of defining clear explanation. It means that the teacher is available to explain further if you are not clear or confident about something that you could not discuss in front of other people.

Student 1: I think similar to what Student 3 said. Once the majority has understood what the teacher is trying to say and if they can understand it, then they can explain t to their friends who will be more likely to ask them than the teacher.

10. *This term the resources and other material my teacher gave were well prepared*

Student 3: I think this means that they fully gone through everything and understood the implications arising out of the reading so that you may engage with questions arising out of it rather than perhaps leaving people to ponder on things themselves. Although that is useful to a certain extent.

Student 1: I think it means that the teacher has gone through the resource and made sure that all of it or most of it applies to the syllabus rather than using all material from past syllabuses . That can be useful for outside reading but if it does not apply directly to the syllabus than some of it might not be worth learning as much.

11. *I knew how my lessons were going because my teacher explained the lesson objectives clearly and taught in a way that helped me to achieve them.*

Student 5: The teacher is actually teaching you what you're doing rather than just instructing you to do this.

Is it the preferable way around for you?

Student 5: Yes ???

Is that important to know because you can see what you are doing but not why you are doing it in the broader scheme. Is the Why important?

Student 5: If you are working towards an exam

Student 4: I think if I was given a task to do and I found that it was relevant while it was not actually so, I definitely would question why we were doing it? I think when I am given a task I need to know why I am doing it and whether it is relevant. So it is definitely important to know why.

Student 3: I would agree with what Student 4 says because if you had just carried on with a lesson because you were told to particularly if you are in a high pressured time or environment like the GCSE s or A Level where you don't have much time to waste and nobody explains why you were doing them, that could be dangerous. You could

under revise it or think it as an abstract concept that is kind of useless ti you which could be to the detriment of your success and overall learning.

12. *My teacher's lessons enabled me to take notes this term.*

Student 1: I think it means that the teacher taught at a slower pace and progressed through examples slow enough for you to be able to take then down and understand them rather than rushing through them with just one example so that you don't have the time to fully take it down and understand it.

Student 5: If you can note them down, it helps you later to revise.(?)

So it is not just about the clarity and speed which is ordering things appropriate

Student 5: Certainly understand things as it was (?)

Student 4: I agree with what Student 5 said because I find when someone is talking and if they don't have an order of what they are saying, then I find it very confusing to write down notes because I am one of these people who like things categorized and clear. But if it goes from one thing to another and that too very fast, then it is confusing. I would like things in a manageable pace.

Student 3: I think it has to be episodic in a way and have clear signs and examples of where they are going because if you are making notes, you need concluding sets which gives you key points of everything. Otherwise you are wasting lots of tome trying to manage confusing things. Sometimes they can explore a subject and ten come to the key points that you really do need to know, that is particularly good.

13. *This term my teacher encouraged students to take part in class discussions.*

Student 4: I think that it means that the teacher planned out a lesson on a debate form and they also ask the quitter students to voice their opinions. It also means that they raise issues that are deliberately controversial even if they do not believe that themselves. So that that people can form their opinion on that.

Student 5: They ask questions that are controversial ???

14. *My teacher invited students to share their own ideas and knowledge in my class this term.*

Student 2: I think the teacher used people's names and make sure nobody feels excluded in class and ensure that everybody has an opportunity and not just some people getting more attention than others. It is very much like treating everyone equally and respecting one's ideas and encouraging that.

Student 1: I think it means that the teacher instead of explaining themselves certain areas of the subject will ask the students their understanding of it so that the class has more than one way of approaching a matter at hand.

15. *This term I was encouraged to ask questions and my teacher gave me meaningful answers.*

Student 2: I think it means that every student is made aware that should they not understand a certain point they should not feel scared to ask for help if they don't understand something. Meaningful answers means that your teacher is not just rushing through explanations but persist till they realize that you have understood.

Student 3: I think it is when they realize that you are learning right and so they should make a deliberate pause and go back and check to make sure that the majority of the people are reasonably confident with what they have got. So the students realize they have time to answer questions.

Student 4: I think that when an issue is raised, the teacher must leave it open, encourage people to contribute to it their viewpoints.

16. *I was encouraged to express my own ideas and/or question my teacher this term.*

Student 2: I think it is about the teacher making sure that everyone is comfortable about what they are learning and how the teacher is teaching them.

Student 3: I think it means that your teacher is being able to teach you in a way so that they can stand back and encourage students to interact among themselves and then intervene to add bits of knowledge. It is basically learning from each other.

17. *This question is in two parts. This time my teacher was attentive to Individual students. This time my teacher was friendly towards individual students. Mention if the two parts mean the Student 1e thing or something different, which one you prefer.*

Student 1: I think my teacher is attentive towards individual student means that if a specific person n class is struggling then the teacher will know that the person is struggling and they will be able to keep an eye on them, explain things extra or just make sure they understand the lessons. Whereas my teacher was friendly towards individual students means that the teacher takes interest in a student as a person, can share a joke if appropriate and generally make lessons more fun.

Student 2: I think the question is confusing. Does individual students mean certain individual student or towards everyone individually. There is a big difference between being attentive and friendly. If the teacher is friendly with certain individuals and not

the whole class that is an extremely bad thing. So the way in which you interpret that semantically can make a big difference to the answer I will give on that.

So if I rephrase it into something like my teacher was attentive towards all students. Does that change things?

Student 2: Yes, that would imply things in a positive way. I think the word attentive can be interpreted both in good and bad ways whereas friendly is definitely a good way of explaining things. Attentive can have bad connotations like talking of things appropriate or tens towards discrimination. But it can also mean in a positive way.

So if you were to choose between either attentive or friendly, assuming that both means a good rapport and relationship with individuals, which of these two statements would be better.

Student 3. I think you can't explain the difference between attentive and friendly depending on the teacher's engagement with the individual person because if they are being attentive then that means the teacher understands any questions and issues the students might have and gives meaningful answers. But they can't do that unless they talked to you in an appropriate way so they are not getting angry or frustrated with opinions you may have. So I don't think you can draw a line between the two necessarily.

So who would you choose a teacher attentive towards you or a teacher friendly towards you?

Student 3: Attentive

Student 2: If they are friendly that implies they are paying attention towards you or wish to pay attention towards you.

Student 4: I think the two are completely different things.

Student 5: I think they are different things but if I were to choose, I would choose an attentive teacher because their job is to be your teacher and not be your friend.

Student 4: I Would say that as well.

Student 1: Yes, Similar. The teacher should be attentive but not overtly, so that you are not obviously singled out among others. I think they should be equally attentive towards everybody.

18. *This term my teacher had a genuine interest in individual students in my class.*

Student 3: I think the meaning of it depends on interpretation. It can be dangerous but if it is genuine interest, then it has got to be a universal thing among the class. If the

teacher is interested in the opinion of the students, he can add to those and incorporate those into building concepts. But if the interest concerns only certain individual students it can be detrimental to the class as a whole. So it can be difficult to interpret semantically.

Student 2: I think the teacher should always bear in mind that the students in class may not be very advanced in their knowledge but can appreciate things and still learn something.

So is it suggestive of a two-way process.

Student 2: Perhaps. The teacher should have the ability to appreciate this.

19. My teacher made students feel welcome in seeking help or advice in or outside of lessons this term.

Student 5: This means that the teacher will actually say the student that if he/she does not understand something they are welcome to ask more and the teacher would be happy to do that.

Student 4: If a student approaches a teacher with a question and the teacher says he is busy at that moment that can be very off-putting. It means that the teacher has to have a set time to ask students to come and see his/her. In that way students will feel welcome.

Student 1: A regular time every week can be very useful. Say for example if they say, every Wednesday lunch time I am in this room that is very reassuring for students.

Student 3: I think it is the recognition from teachers that learning does not stop inside the classroom in terms of your attempt to understand things. Sometimes students may feel awkward to express concerns before the class. So if a teacher thinks that explaining a thing in the syllabus is all that his/her job is about that is wrong because their job is actually more continuous than just in the classroom.

20. *My teacher was adequately accessible to students outside of lessons this term.*

Student 4: That is basically what we were just discussing that the teachers should make known that they are available at a regular set time and the students need not have to wander about looking for them.

Student 3: If the teachers recognize that the students are genuinely concerned about something they have gone through in class, they must be willing to share their concern in any way possible. Having a fixed time is a good idea, but the teacher should also be flexible to see a student if required.

Every topic has a life cycle of discussions. So is it not better to know that the teacher is available if you require more help in understanding a particular issue that may arise out of these discussions at any stage?

Student 1: I think in my subject, where once you have gone over the subject you will do exercises on it. When the exercises become harder towards the middle to later stages and many questions arise, the teacher should be more available so that they pick up topics combining various issues.

Student 3 : It is definitely at the application point the issues tend to arise the most because a student may know the concepts but the application may require much more than that.

21. *This term my teacher demonstrated how different theoretical ideas are linked together. Or this term my teacher contrasted the implications of various theories.*

Student 1 : I think the first statement shows how things progress through time where things develop from the most simple models to the most complex and up to date ones. The second statement implies how theories of the Student 1e time or phase can be edited contrasted or combined to arrive at the next level of theory.

Student 5: I think the first part ??? together to show how the ?? using parts of various chapters.

Student 2: The first part suggests the structural element of what you are learning, the method of taking in and understanding a particular situation. The second statement is the individual effects of the different ideas as against the structure. It is not necessary that the two parts have anything to do each other.

22. *This term my teacher explained the background or origin of concepts and ideas developed in class.*

Student 4: I think this means that the teacher gives a detailed background of a particular topic rather than rushing through explanations so that you are informed well.

Is that important.

Student 4 : Definitely important. In my subject, if my teacher did not give a background as to hoe things progressed towards the big event being discussed, I would really feel confused.

Student 3: I think it means the teacher should provide some contextual information about the birth of an idea so that you can evaluate it yourself.

Student 2: I think whether the background information of an idea or event is relevant or not depends on the subject. It can't be as important in history. Some subjects on the

other hand may focus only on the present or future and not be bothered about the past. As for my subjects (Maths, French, English Language, Ethics and Philosophy), knowing the background of a particular thing is not always very relevant

Student 4: I also agree with Student 2 that it really depends on the subject.

Student 1: I think in the Maths and Sciences (my subjects), the background of things tends to be less important. But if you can see where something has come from it can be useful nevertheless.

Student 3: There is a difference between Arts and Science Subjects. But for Arts Subjects it is more important to have contextual information so that you get various interpretations of things which enhance your understanding.

23. *This term my teacher presented points of view other than his/her own when appropriate*

Student 4: I think it is important to present various points of view. If your teacher is very politically aware, and gives points of view of the party they support, that can make you biased in favour of that party.

Student 1: I think this applies less in Sciences. If a teacher is particularly passionate about a subject or area they can gloss over other areas so that they have time for teaching their favourite section which is not always good.

Student 5: Especially in Maths, when the teacher might just not show only one way but all the ways to solve a problem so that you know which way works best for you.

Suppose, there is a set syllabus for you to follow in your A Levels, and then the teacher is giving you opportunities to learn things around peripheral lines, is the teacher's point of view relevant there at all?

Student 3: I think there is a disparity again between Arts and Science subjects because Maths and Sciences have set things which are true which might be otherwise in Arts Subjects. I think a teacher should go through all opinions. Their opinion is particularly important because you are going to consider it in a major way in a theory or event.

Student 2: I think a teacher's point of view is very important. However, they should also communicate to the students the fact that their opinion is not always the most important so that it leaves room for debate.

So you are fine with the idea of teacher's having their own opinion if they communicate it as being solely their belief?

Student 2: Yes. Even in science subjects, teachers can have their own methods in working mathematical or scientific problems out but that might not be the right way out or students.

24. *This term my teacher adequately discussed current developments in the subject.*

Student 3: I think it means that your teacher has used all resources from the syllabus and applied them so that you understand the subject well. If you are studying statistics, your teacher may be talking about Gambling so that you understand the application of what you are doing.

So is it important and relevant to discuss the current developments and recent discoveries in your subject?

Student 4: As a Language student, I think it is important that the teacher adequately discusses current developments in the Language. So for example if somebody I studying English, it is necessary to know how the language has changed and is changing over time. The talking vocabulary has changed from the formal correct usage.

Can anybody think of a subject where you would not have to know about current developments?

Student 3: Sometimes it can be detrimental if it is not handled in the right way. If you are doing Maths or Science, you are taught something and then somebody says it is not true but that you have to know it only for the syllabus that can be quite confusing and undermining.

Student 1: I think you need to know what has happened before you know what is happening. If you are interested in it you should definitely read up on the latest developments to further your interests.

Student 5: It also helps ??? if the teacher discusses that in class.

25. *The feedback my teacher gave on my assessed work for this term was valuable.*

Student 4: I think if the teacher gives you set things to improve on after going through the essay that is really useful. If the essay is graded but has comments, you would not know how to improve.

Student 2: I think if you get an essay back you need to have comments on how to improve further, why you didn't get top grade and what went wrong. If the teacher spends some time in how to help you overcome the problem and write them down and advice in a proper legible handwriting that is quite useful.

Student 3: The evaluation should be constructive rather than just saying good or bad so that you know ways of improving yourself and strengthening your skills to improve your marks.

26. *The ways in which my teacher assessed my work this term was fair and appropriate.*

Student 1: I think in the Maths and Science Subjects there is quite often a right or wrong answer, the teacher will find it difficult to do. .If the teacher can re-explain to you the places where things went wrong rather than just pointing out your mistakes that would be more helpful.

Student 5: Also the way it is marked with massive crosses can be done away with.

Student 3: I think it is necessary to put you in the correct conditions. They should mark your essays with any bias if their opinion does not match yours. It should be based on the quality of your work rather than the concepts of your argument.

Student 4: If a teacher has a set belief on something which you disagree with, they are not going to mark that as incorrect but appreciate the difference of opinion. It is also very important to have the Student 1e standardised marking for each individual to be fair in marking.

Student 2: I agree. When a teacher marks a student's work, he/she must maintain the stance of being neutral and not mark a student down because it is something different from what they think.

27. *The assessed work and examinations this term has tested what I have been taught in lessons.*

Student 2: It means that the exams and the lessons were on the Student 1e lines.

28. Homework set by this teacher this term was valuable.

Student 3: The homework makes you do exam style application of your work rather than just doing it for the sake of it.

Student 1: If the teacher gives you homework, it should be doable and not just mountains and mountains of questions. If you understand a particular type of question you need not redo similar types of questions over and over again. If you need more practise, you can choose the right amount to improve yourself.

Student 5: The attitude of the teacher should not be like I don't have to do it, so just do it for homework.

So it should be consolidation of learning that has already taken place.

Student 5: Yes, and also we should be able to go and ask the teacher for help.

29. *This term the homework set by this teacher contributed to my understanding of the topics taught.*

Student 2: I think that it is asking whether you think homework is stimulating, helping you to appreciate your understanding. Sometimes homework has the inbuilt connotation of being warned that you have do it etc. which is off putting and too binding.

Student 3: I think that in means you are appreciating the various interesting applications of something that is taught to you. When you are doing your homework, you can still use the concepts taught in class without the teacher's first person support.

30. *Overall, how does the lessons of this teacher compare to the other teachers in the school/college?*

Student 3: I do not like the question much because it is open to too much personal interpretation. You may not just like a teacher for may be a bad haircut which has no bearing on whether they can teach. You may just dislike the subject which also bears on the teacher and you may just not like their teaching.

Student 2: I think comparing this teacher to others and rating them is a concerning issue. The question does not aim to point out any inconsistency.

Student 1: I think the question rather answer will depend on the average level of teaching in the school. If the teacher has lots of poor teachers than average teachers the rating will vary likewise.

Student 4: I think it relates to the previous question. The previous question was judging their lessons but now it almost gets personal.

Is that a good thing?

Student 4: I think if names are involved that can be bad but as long as it is anonymous it would be ok.

Student 2: I don't think it is a good thing. The only thing you should rate your teacher upon is their ability to teach. Everything else is irrelevant.

Student 1: I think if people were to answer, there can e problems. If the entire class dislikes a particular teacher, may be something is wrong with the teacher, technique or something.

31. *If there is any other feedback you want to give your teacher about their teaching, please do so? You can enter text. Is it a good thing to have a free text entry in this?*

Student 4: The question is very self-explanatory. I think the free text entry is a good thing but it should not be spied on but moderate. If someone does not like a teacher they may use it in a negative way to express their dislike but there must be a set number of characters…

Student 3: I think it is better to ask if we can contribute any feedback towards improving the learning environment. You can highlight any direct issues you have but that can be professionally dangerous.

What if whatever was written there was only seen by the teacher?

Student 3: I think it is really difficult because you expect your teacher to take the best from what you say.

Student 2: I think the free text idea is a good one and I don't think you should have options because they tend to generalize things anyway. They can apply to any teacher so the real problems of a particular teacher cannot surface. If some bias or hatred goes into the system, it cannot be a constructive method.

Student 5: The free text allows you to express things you can't do only by ticking boxes.

Student 1: I think if you are asked to highlight what is good and bad on a list, which can be useful.

This is just a trial. If your teachers agree to take part in it, and everything goes well, we can take it on o next year. Realistically, there is an underlying question through all this. If you take a survey, how should that be for you.

Student 1: I feel quite often when surveys are given out I don't agree because ticking boxes cannot always give the right impression. Having more thought about each question is more important than just generalized questions. It also needs to be stressed on how important the survey is rather than just distributing sheets.

Student 3: if you have an issue with a teacher which is legitimate and based in their teaching and you don't want to see them directly, or reporting to the head of the department, the surveys can help as a voiced opinion.

Student 2: It is reassuring to know that the school appreciates us having a voice.

Thank you very much.

Appendix 3c

SecondarySEEQ Pilot Project

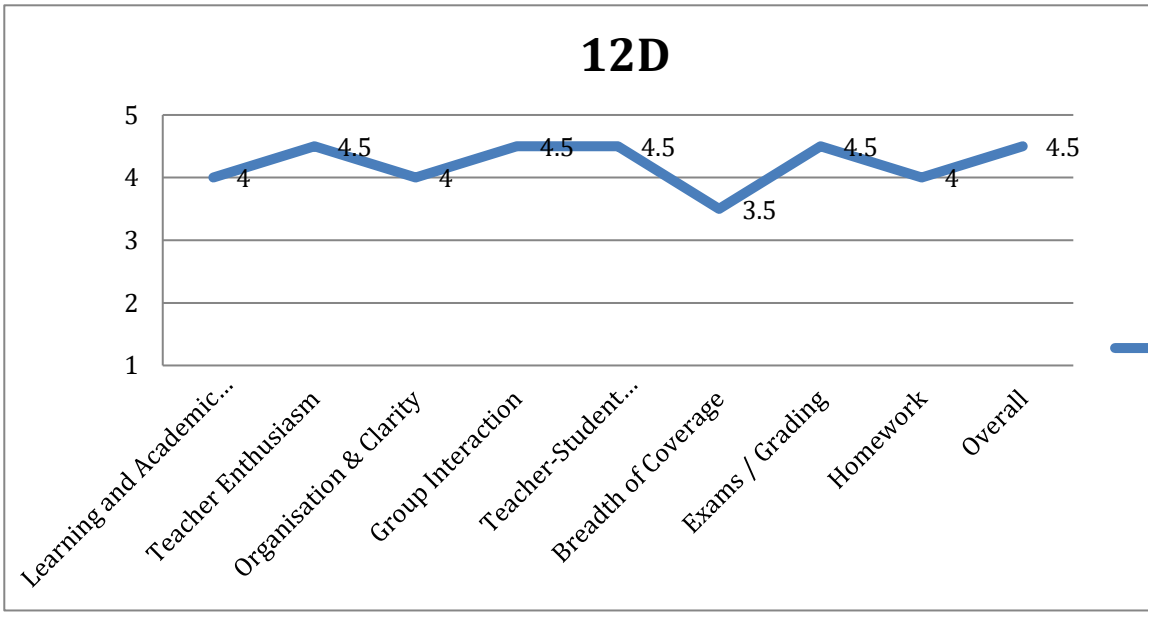
Spring Term 2012

Feedback for Wood

The graph(s) over the page show(s) the average ratings for each of the 9 factors your class(es) (named above) has given for your teaching during the Easter term 2012.

Once you have looked at the graph(s):

1. Discuss these data with Stuart
2. Question and discuss what you think these data mean



5 – Strongly agree

4 – Agree

3 – Neutral

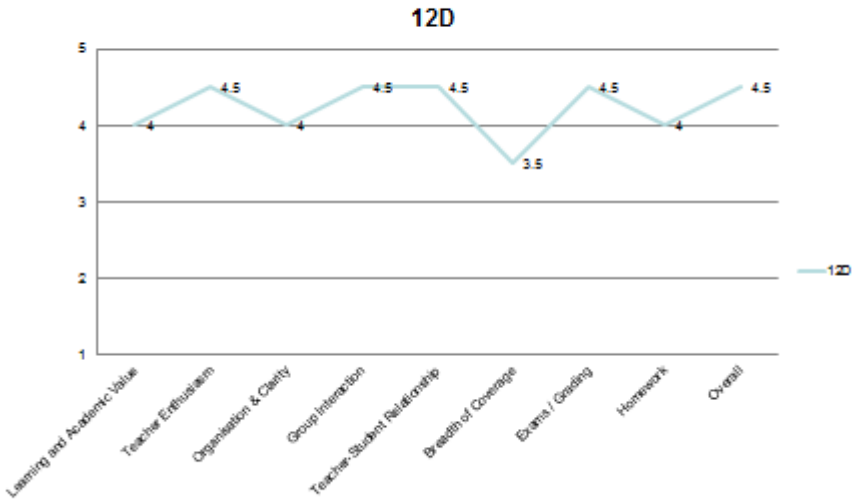
2 – Disagree

1 – Strongly disagree

SecondarySEEQ: Feasibility Study School Feedback: Teacher 1

Stuart Kime
Centre for Evaluation & Monitoring (CEM) and School of Education, Durham University

12D



Q.1 The process so far

- How would you describe the experience of having your students evaluate your teaching?

Q.2 Feedback sheets

- Do you think you can rely on these data as being an accurate evaluation of your teaching during the Easter term?

Q.3 The process so far

- Do you perceive that it is your teaching that has been evaluated, or that you as a person have been evaluated?

Q.4 Feedback sheets

- What uses could you make of your data?

Q.5 Feedback sheets

- What are your interpretations of these data?

Q.6 Feedback sheets

- Do you think these data are presented to you in the most effective way?
  - *If you think there are more effective ways of presenting the data, please indicate what these are.*

Q.7 Feedback sheets

- What will you / would you do with these data?

Good feedback causes thinking

"… we have each been asked several times by teachers, 'What makes for good feedback?'—a question to which, at first, we had no good answer. Over the course of two or three years, we have evolved a simple answer—good feedback causes thinking."

(Black & Wiliam, 2003)

SecondarySEEQ Strategies

- Involvement in editing existing strategies / suggesting new ones?
- Which area / areas would you like to contribute to?
- Would you be willing to take part in a 'teacher talking heads' video in which you would offer your thoughts (to other teachers) on the evaluation of teaching process?

Appendix 3d

Feasibility Project teacher interviews: key phrases

Initial reactions

- “a bit of a dip here” on homework
- “that’s pretty much expected” - unsurprising
- “you’re too busy trying to get the specification covered to do that sort of thing” - going off-piste is difficult (so BOC may go low?)
- “I’m a little bit surprised the 13s are higher than the 12s”
- “I’m not surprised about the homework being low…I set a lot of homework”
- “Fine. It’s fairly much where I hoped it would be.”
- “I’m interested in the O&C rating, because the kids are out of their seats…I’m planning to make them dynamic. But maybe they think it’s unplanned and spontaneous”
- “It’s only my second year of teaching A Level, so it’s useful to have this feedback”
- “I kind of expected the low O&C rating because I am kind of a floaty character”
- “Oh, that’ll do”
- “I’m kind of curious as to what goes into OR, as I think that the rest of them cover everything”
- “I don’t think I’m excellent so it would be distressing if they thought I was, but equally it would be bad if they thought I was rubbish”
- “It’s good. Everything that I would hope to accomplish in a lesson seem good, so that’s ideal”
- “I’m pretty nondescript on it…but this just proves what I thought was going on so it’s nice to have that”
- “I knew the 13s would be more favourable than the 12s…the 13s are more forthcoming and lively.
- “If I’d have drawn the graph myself if would have looked like that”
- “I must say there are only 3 children in the class, so statistically they’re not viable, but it’s good to know”
- “Does the number mean what they think of the subject or what they think of me?”

How would you describe the experience of having your students evaluate your teaching?

- “I was a bit apprehensive about it…”
- “You think you’re doing ok but you don’t really find out until someone tells you otherwise”
- “In general the process of getting them to fill out questionnaires and do it wasn’t a problem at all”
- “It’s satisfying…that’s ok to be where I am in their opinion”
- “a little bit nerve-wracking to have them evaluate me and see what the results would be”
- “it’s a rare experience to get to that kind of feedback”

- “I think we should all do it. I feel I’m constantly learning as a teacher. I find it really rewarding and not nerve-wracking at all. The children have the right to feed back”
- “Maybe I need to be more explicit about what we’re doing so that they don’t think I’m just a crazy woman”
- “Mildly nerve-wracking”
- “I really hoped they didn’t think I’m pants”
- “I think our students say what they think…the data will be reliable, I think”
- “Thought-provoking, but it isn’t anything I didn’t expect. It makes you think about what you can do to improve”
- “I was a little bit nervous”
- “I knew there was something going on there that I needed to sort out”

Do you think that you can rely on these data as being an accurate evaluation of your teaching during the Easter term?

- “I think that would be fair to say, but there’s only a small number of students”
- I think so, yes, but I imagine they answered for the whole year experience as opposed to a term…I’d imagine they have an impression of me as a teacher, not just of me in a term”
- “I would say that, when this was taken - this was just after the mock - there was a little bit of anxiety… So I don’t feel that this evaluates fully my teaching…I would expect that, hopefully, post-exam, these things would be better”
- “No, because they did it after the holidays and they always look back with rosy-eyed vision”
- “It was a bit odd looking back retrospectively…” Students found it hard to remember what had happened
- “Possibly”

Do you perceive that it is your teaching that has been evaluated or you as a person have been evaluated?

- “Probably both. I don’t think they’re mutually exclusive.”
- “Teaching, I would say. I feel they’re very much about the quality of the lessons”
- “I think the teacher persona I have has been evaluated, so I think that has what’s been evaluated”
- “Both. You can’t separate them. They tend to work for you, rather than the subject. They feel like they’re letting you down if they don’t work hard and that’s quite a good carrot and stick”
- “It must be my teaching they’re thinking about, but it’s probably about both”

What uses could you make of your data?

- “I think these 3.5s [the lowest ratings]…maybe I could do something about that.” (But *what* is the issue here)
- “Maybe I’d ask them” - to follow up and find out more about where improvements could be made

- “I’m not really sure… I don’t think I would change anything really”
- “I think the two key things are to make it explicit why we’re doing very creative things…I need to show them the structure underneath…and the homework. I give mainly reading homework.”
- On high BoC rating - “I think that’s the resources section that we’ve just done”
- “I want 5s”
- Discussion about departmental use of the data for feedback (a colleague had had ratings done also)
- “Looking at the weaker areas and trying to improve them”

What are your interpretations of these data? What do you think they mean?

- “The Year 12s agree with the opinion that I have of my teaching, but less so the Year 13s”
- “It’s all generally good, but there are dips in the data.”
- “BoC is a general teaching conundrum related to the curriculum…”
- “The 3.5s give me something to focus on and I think I would be arrogant to think that there is nothing that I could improve”
- “It’s ok, pretty good. She could probably be a bit better. What I thought anyway”
- “I’m pleased with it. Students are our strongest critics”
- “I think student feedback is essential”
- “It’s making us think of all sorts of things…it throws up lots of questions”
-

Do you think this is the most effective way of presenting these data? If not, how could it be improved?

- Have a range on each on - “it would be useful to see the spread”
- “these are averages so it would be good to see what individuals said without naming names”
- “It would be quite interesting to know the range…not the raw data”
- “Could you not do it as a bar chart (it’s categorical data)? But it could be more difficult it you have more than one class and you want to see them on one graph”
- “I don’t think you could get much clearer than that”

What will you / would you do with these data?

- “just make sure that what the Year 13s have commented on with the 3.5s…to not have that repeated with the Year 12s”
- “I’d quite like to go through my schemes for 12 and 13 to look at how I could do more group interaction”
- “I wouldn’t want to compromise the survey, but I’d like to get suggestions from the children about O&C and H”
- “It’s reinforced what I thought already so it’s useful to have that reinforcement”
- “If there was a dip somewhere, it would be something I could work on, but there isn’t”

- Comment about being able to see if there are ways to view the data holistically. To see school average / subject average etc. - with a view to coaching those who need development
- “I need to go back in with a spring in my step”

Comments

- Justification of responses - “they’re just going to have to lump it” (sometimes things just can’t be changed because of policies etc.)
- “These are just means…that would have been interesting to see the range of scores”
- It would be useful to know who said what
- One teacher commented about response of students to seeing the list of all teachers on the list - expressed dismay at the thought of having to do it for everyone!!

Appendix 6a

SecondarySEEQ final version for trialling in main project

SecondarySEEQ: X School Term 2

*1. Select your school name from this menu(*Required)	
Select one.	
○	X School

*2. What is your first name? Please check you have spelled your first name correctly.(*Required)

*3. What is your surname? Please check your surname is spelled correctly.(*Required)

*4. What is your date of birth?(*Required)	
*Date:	
*Month:	
*Year:	

*5. Are you male or female?(*Required)	
Select one.	
○	Male
○	Female

*6. Which year group are you in?(*Required)	
Select one.	
○	Year 12
○	Year 13

*7. Select the surname of the teacher whose lessons you are going to comment on.(*Required)
Select one.

○	A
○	B
○	C
○	D
○	E
○	F
○	G

*8. Select the class (the teaching of which you are rating in this survey) from the list below:(*Required)

Select one.

○	12 English (Mr A)
○	12AA Design & Technology (Mr B)
○	12AA History (Dr C)
○	12AB Physics (Mr D)
○	12AB Psychology (Mr E)
○	12AC Maths (Miss F)
○	12AC Psychology (Mr G)
○	12AD Geography (Miss G)
○	13 English (Mr F)
○	13AA History (Dr C)
○	13AA Psychology (Mr B)

○	13AB Geography (Miss D)
○	13AB Psychology (Mr A)
○	13AC History (Dr C)
○	13AC Physics (Mr D)
○	13AD Design & Technology (Mr E)
○	13AD Maths (Miss F)
○	13AD Physics (Mr G)

*9. Which subject are you taught in this class?(*Required)

Select one.

○	Accounting / Finance
○	Ancient History
○	Arabic
○	Archaeology
○	Art & Design (Applied)
○	Art and Design
○	Art and Design (3D Studies)
○	Art and Design (Critical & Contextual Studies)
○	Art and Design (Fine Art)
○	Art and Design (Graphics)
○	Art and Design (Photography)
○	Art and Design (Textiles)

○	Biblical Hebrew
○	Biology
○	Biology (Applied)
○	Biology (Human)
○	Business (Applied)
○	Business Studies
○	Chemistry
○	Chinese
○	Citizenship
○	Classical Civilisation
○	Classical Greek
○	Communication Studies
○	Computing
○	Critical Thinking
○	Dance
○	Design & Technology
○	Drama and Theatre Studies
○	Dutch
○	Economics
○	Economics and Business
○	Electronics
○	Engineering
○	English Language

○	English Language and Literature
○	English Literature
○	Environmental Studies
○	Ethics and Philosophy
○	Expressive Arts and Performance
○	Film Studies
○	French
○	General Studies
○	Geography
○	Geology
○	German
○	Government and Politics
○	Gujarati
○	Health and Social Care
○	History
○	History of Art
○	Home Economics
○	ICT (Applied)
○	Information and Communication Technology
○	Irish
○	Italian
○	Japanese
○	Latin

○	Law
○	Leisure Studies
○	Mathematics
○	Mathematics (Further)
○	Mathematics (Pure)
○	Mathematics (Statistics)
○	Mathematics (Use of)
○	Maths
○	Media / Film / TV Studies
○	Media: Communication and Production
○	Moving Image Arts
○	Music
○	Music Technology
○	Performing Arts
○	Persian
○	Philosophy
○	Physical Education / Sports
○	Physics
○	Portuguese
○	Psychology
○	Religious Studies
○	Russian
○	Science (Applied)

○	Science in Society
○	Science: Single Award
○	Sociology
○	Spanish
○	Travel and Tourism
○	Turkish
○	Urdu
○	Welsh (1st Language)
○	Welsh (2nd Language)
○	World Development

*10. This term, I found this teacher's lessons intellectually challenging and stimulating.(*Required)

Select one.

○	Strongly disagree
○	Disagree
○	Neutral
○	Agree
○	Strongly agree

*11. This term, I learned some valuable things in this teacher's class.(*Required)

Select one.

○	Strongly disagree
○	Disagree
○	Neutral
○	Agree
○	Strongly agree

*12. My interest in this subject increased this term as a result of this teacher's lessons.(*Required)

Select one.

○	Strongly disagree
○	Disagree
○	Neutral
○	Agree
○	Strongly agree

*13. I learned and understood what my teacher has taught me this term.(*Required)

Select one.

○	Strongly disagree
○	Disagree
○	Neutral
○	Agree

○	Strongly agree

*14. This term, my teacher was enthusiastic about teaching my class.(*Required)

Select one.

○	Strongly disagree
○	Disagree
○	Neutral
○	Agree
○	Strongly agree

*15. This term, my teacher was dynamic and energetic when teaching my class.(*Required)

Select one.

○	Strongly disagree
○	Disagree
○	Neutral
○	Agree
○	Strongly agree

*16. My teacher enhanced their teaching with the use of humour this term.(*Required)

Select one.

○	Strongly disagree

○	Disagree
○	Neutral
○	Agree
○	Strongly agree

*17. My teacher's style of teaching held my interest during lessons this term.(*Required)

Select one.

○	Strongly disagree
○	Disagree
○	Neutral
○	Agree
○	Strongly agree

*18. My teacher's explanations were clear in my lessons this term.(*Required)

Select one.

○	Strongly disagree
○	Disagree
○	Neutral
○	Agree
○	Strongly agree

*19. This term, the resources and other materials my teacher gave the class were well-prepared and carefully explained.(*Required)

Select one.

○	Strongly disagree
○	Disagree
○	Neutral
○	Agree
○	Strongly agree

*20. This term, I knew where lessons were going because my teacher explained the lesson objectives clearly, then taught in a way that helped me to achieve them.(*Required)

Select one.

○	Strongly disagree
○	Disagree
○	Neutral
○	Agree
○	Strongly agree

*21. My teacher's lessons enabled me to take notes this term.(*Required)

Select one.

○	Strongly disagree
○	Disagree

○	Neutral
○	Agree
○	Strongly agree

*22. This term, my teacher encouraged students to take part in class discussions.(*Required)
Select one.

○	Strongly disagree
○	Disagree
○	Neutral
○	Agree
○	Strongly agree

*23. My teacher invited students to share their own ideas and knowledge in my class this term.(*Required)
Select one.

○	Strongly disagree
○	Disagree
○	Neutral
○	Agree
○	Strongly agree

*24. This term, I was encouraged to ask questions and my teacher gave me meaningful answers.(*Required)

Select one.

○	Strongly disagree
○	Disagree
○	Neutral
○	Agree
○	Strongly agree

*25. I was encouraged to express my own ideas and / or question my teacher this term.(*Required)

Select one.

○	Strongly disagree
○	Disagree
○	Neutral
○	Agree
○	Strongly agree

*26. This term, my teacher was friendly towards individual students in my class.(*Required)

Select one.

○	Strongly disagree
○	Disagree
○	Neutral

○	Agree
○	Strongly agree

*27. This term, my teacher had a genuine interest in the individual students in my class.(*Required)

Select one.

○	Strongly disagree
○	Disagree
○	Neutral
○	Agree
○	Strongly agree

*28. My teacher made students feel welcome in seeking help / advice in or outside of lessons this term.(*Required)

Select one.

○	Strongly disagree
○	Disagree
○	Neutral
○	Agree
○	Strongly agree

*29. My teacher was adequately accessible to students outside of lessons this term.(*Required)

Select one.

○	Strongly disagree
○	Disagree
○	Neutral
○	Agree
○	Strongly agree

*30. My teacher contrasted the implications of various theories this term.(*Required)

Select one.

○	Strongly disagree
○	Disagree
○	Neutral
○	Agree
○	Strongly agree

*31. This term, my teacher presented the background or origin of ideas / concepts developed in class.(*Required)

Select one.

○	Strongly disagree
○	Disagree
○	Neutral

○	Agree
○	Strongly agree

*32. This term, my teacher presented points of view other than his / her own when appropriate.(*Required)

Select one.

○	Strongly disagree
○	Disagree
○	Neutral
○	Agree
○	Strongly agree

*33. This term, my teacher adequately discussed current developments in the subject.(*Required)

Select one.

○	Strongly disagree
○	Disagree
○	Neutral
○	Agree
○	Strongly agree

*34. The feedback my teacher gave on my assessed work this term was valuable.(*Required)

Select one.	
○	Strongly disagree
○	Disagree
○	Neutral
○	Agree
○	Strongly agree

*35. The ways in which my teacher assessed my work this term were fair and appropriate.(*Required)	
Select one.	
○	Strongly disagree
○	Disagree
○	Neutral
○	Agree
○	Strongly agree

*36. The assessed work and examinations I did this term tested what I was taught in lessons.(*Required)	
Select one.	
○	Strongly disagree
○	Disagree
○	Neutral

○	Agree
○	Strongly agree

*37. Homework set by this teacher this term was valuable.(*Required)

Select one.

○	Strongly disagree
○	Disagree
○	Neutral
○	Agree
○	Strongly agree

*38. This term, the homework set by this teacher contributed to my appreciation and understanding of the topic(s) taught.(*Required)

Select one.

○	Strongly disagree
○	Disagree
○	Neutral
○	Agree
○	Strongly agree

*39. Overall, how does the workload in this teacher's lessons compare with other teachers' lessons in this school / college?(*Required)

Select one.

○	Very low workload
○	Low workload
○	Average workload
○	High workload
○	Very high workload

*40. Overall, how does the level of difficulty in this teacher's lessons compare with other teachers' lessons in this school / college?(*Required)

Select one.

○	Very easy
○	Easy
○	Average difficulty
○	Difficult
○	Very difficult

Appendix 7a: text documents used to inform teachers about plausible strategies for use in improving subsequent SecondarySEEQ ratings

Teaching strategy ideas for Learning & Academic Value

In essence, this aspect of teaching denotes subjective feelings of success obtained through participation in a course and / or at the hands of a particular teacher. Students who are challenged and stimulated, who consider their learning through the course to have been worthwhile, whose interest in the subject was increased, who are conscious of having understood the subject-matter and who generally rate the course highly are clearly expressing feelings of accomplishment or challenging learning tasks.

The following ideas were suggested and used by outstanding teachers across a range of institutions and disciplines. We suggest that you select no more than 3 or 4 strategies which appear potentially the most profitable, then make a commitment to apply OR adapt them to improve your teaching effectiveness.

1. Plan your teaching around a conceptual framework on which students can hang major ideas and the factual information of a course.

Planning a clear framework which would be useful and meaningful to students is a distinct feature of the approach used by several very effective teachers across a broad spectrum of disciplines.

The framework may be a structure, a theme, a controversial issue, or a theory. It should be made salient to your students through repeated reference.

As one teacher of Biology points out, "To the uninitiated, our field looks like a mass of facts; by establishing a conceptual framework, I minimise the amount of rote memorization my students have to do."

Often the framework can be represented symbolically or graphically. Another Biologist, for example, begins each of a series of lessons by projecting the same outline of the human heart on the whiteboard. Details then given about the heart, in terms of structures and processes, change according to the specific topics to be covered in that day's lesson.

A teacher of Spanish identifies two or three major literary concepts (e.g., irony or tragedy) and applies them repeatedly in lessons, discussions and homework to reinforce student understanding.

2. Take into account what your students *want* to know.

Being aware of the students' expectations for a course is a critical prerequisite for obtaining their attention, interest and participation, according to a number of outstanding teachers in various disciplines. "It is important to be seen to be aiming to meet their needs, rather than simply following a checklist of things to be taught" says a teacher of Physics.

One teacher of MFL said that she feels it is important to recognise that, because different students need to know different things at different times, being attuned to where your students 'are' in their learning is vital.

3. Assign "minute responses" at the end of some lectures.

Several outstanding teachers have found that asking just two or three open-ended questions after some lessons is a very effective means of establishing what students understand. Some also found it useful in allowing students to see what they have achieved in a session.

The Physics teacher who invented this process says, "Several times during the term - about once a week - I write two questions on the board before the class starts:

1) What's the most significant thing you learned today?

2) What question is uppermost in your mind at the end of today's lesson?

"Then, I teach the lesson. One minute before the end of the period, I say to the class, 'Take out a piece of paper. You have one minute to answer these two questions.' My students sign the papers and pass them to the centre aisle. Now I read the papers. I find I can tell whether I am getting my points across. I can also tell what problems students

are having. I clarify difficult points next time we meet. I can identify students in trouble early in the term. If a student gives me off-the-wall responses, I catch up with him/her as soon as possible. The 'minute responses' process causes students to listen more actively. All the way along during the class session they are saying to themselves, 'Is this the most significant think I'm going to learn today!' Toward the end of the hour they have to wonder, 'Well, what question is uppermost in my mind at the end of this session.' Students' writing improves. Responses I get in the last week of the term are more articulate, and longer than those at the beginning."

A teacher of French said that saying 'Having had this lesson, write me a question to answer next lesson' is a useful way to close a lesson in a manner that links forward to the next session.

4. Touch base repeatedly with the fundamentals or basics.

"Students like to be challenged, but they need to feel confident and well-prepared to meet the new challenges, and that often means consciously retreating a little to reinforce the foundations on which new material is based," according to an outstanding teacher.

5. Stress the most enduring values or truths in your discipline.

"I stress the permanent values in literature, the emotional responses that a particular novel or collection of novels elicits from us all," says one teacher of English. "I try to get my students to understand why they respond to a given novel the way they do."

After a class has discussed how they feel about a novel - the common emotions it arouses - he tries to lead them to analyse, understand, and explain why nearly everyone feels the way they do. He poses questions such as: What must literature be like in order to get us to respond the way we do? Why does a particular novel affect everyone in the same way? "Behind all my questions is the search for a way of analysing and discussing literature that will explain the most with the fewest assumptions."

6. Confound yourself, and let your students "rescue" you occasionally.

Asking open-ended questions which the students can sense are mystifying to you personally is a great way to encourage relevant discussion and to model your enthusiasm for discovering the secrets of the subject.

A humanities teacher reports that "When I ask myself a question that initially seems puzzling to all of us, it lets my students know that I'm not omniscient, and that it's all right to ask questions or get confused occasionally. The class really responds to that."

A teacher of MFL says that he feels it important that students have the sense that they are engaged in a "legitimate intellectual process".

7. Get to know your students: Where they are "at"; and what they relate to.

Several outstanding teachers stressed that new learning must begin from what students are already familiar with. "Otherwise they quickly become confused, disinterested or anxious," a teacher explains.

A Physics teacher noted that students will work harder to solve a problem that appears to be relevant to them. Rather than restricting problems and issues to essay questions, she delivers many of her lessons around puzzles and quandaries.

One teacher who successfully engages students this way warns that it is important, even when deliberately trying to be provocative or "realistic", to choose topics that the students are "ready" to deal with in the context of the material being covered.

Getting to know what gets the class "fired up", or what they relate to is a strategy that several teachers have recommended. This requires establishing good rapport and making time to chat with students in non-teaching situations.

8. Encourage your students to form small study groups and send representatives to see you about difficulties their groups are having.

One Humanities teacher who does this says, "By encouraging them to form study groups, I am trying to help them get to know at least some of their fellow students and to take advantage of what they can learn from one another. Also, it seems to be easier for some students to come to me for assistance if they 'represent' a group, because the problems are then seen as common to many students not just the group's representative. Staff members can be very intimidating for some students, even those of us who try very hard not to be. While it is difficult for them to admit that they don't understand something, there is a certain comfort in knowing that some of their fellow students are in the same boat and that by joining forces they can help one another."

9. Schedule an individual appointment with each student.

An outstanding teacher stressed the importance of knowing and treating students as people, rather than simply as students. "This is central to making the material relevant, opening up discussion, and generally meeting their learning needs," she explained.

A Maths teacher felt that he was not being successful in generating class discussion. At the end of the third week, still unable to encourage class participation, he decided to pass around a sheet of paper with a list of 10 minute blocks of time when he would be available for individual appointments.

Each of his students was required to sign up for one of the 10 minute appointments. They were told that the chief purpose was for him to get to know his students better and to listen to any complaints or suggestions they might have.

"I found that this was a real ice-breaker," he explains. "Even though most of our discussions were mainly chit-chat, some of my students used the opportunity to indicate problems they were having in our lessons or to make suggestions about course improvements. Perhaps the chief benefit was that it gave me an opportunity to get to know my students better. As a result, they seemed to feel more comfortable asking and answering questions in class."

10. "Open with gusto" and "Finish strong."

One teacher feels there are distinct advantages to giving special thought to the beginning and ending each lesson.

"The opening should secure students' attention and give them the desired mental set. Get off to a good start. Do something to command attention from the outset. Put punch into your opening.

"Have some form of attention-getter.....a gadget or piece of software whose operation depends upon the principles of the day's lesson usually excites attention. Carefully planned questions or statements can also develop the curiosity necessary to get students' attention.

"Action is always an attention-getter. If you intend to use graphs or information on a board, keep them hidden from students' view as they come into the the classroom. Likewise for models and experiments. This will usually whet students' curiosity and make them eager to see what is going to happen.

"The ending is as important as the beginning. Don't let a class session fade into nonexistence. Make an impressive ending. For example, end with: a question for the class to cogitate and answer before next meeting; a quotation conveying the essential theme; a summary; a plenary."

Giving students a strong sense of having achieved something worthwhile or useful by the end of a session has been recommended by several outstanding lecturers as an effective way to end a lesson and motivate students.

11. Begin each lesson by outlining the objectives for the session.

An History teacher refers to this as his "battle plan". "At the beginning of the lesson, I give my students a battle plan so they know where the discussion is going and can follow it more easily," he says. "By laying out exactly what we are going to do, I eliminate a lot of student confusion. I don't want students spending the hour wondering, 'Why are we learning about that?' or 'What does that have to do with anything?' instead of concentrating on the topic in hand."

12. Pose paradoxes for your students to solve.

A Chemistry teacher emphasises conceptual understanding by challenging his students with apparent paradoxes. "Several times each term," he says, "I set up a demonstration to give a visual result that is at variance with what is described in the textbook. My students are then helped to explain the paradox."

"This kind of demonstration really gets my students thinking," he says. "Furthermore, many of my students tell me that they learn more from seeing than from reading. It gives them another way of understanding and helps them gain self-confidence that they do in fact understand."

13. Explicitly call attention to the most important ideas in each lesson.

"I began to emphasise the main points about ten years ago," says one Politics teacher, "when I discovered that you can't rely on students to intuitively know what the most important points are. You have to tell them."

Staff members in several disciplines stress the need to call students' attention to the most important ideas being presented. Some teachers announce the importance of an idea before presenting it, saying such things as "This is really important, so you have to be alert." Other teachers emphasise the most important ideas when summarizing, saying "The most important thing to remember here is..." or "This is so important that everyone of you should have it engraved on a gold plaque and hung over your bed!" as one ICT teacher puts it. "There is no point in my students having to guess what is important if I can tell them," he says.

14. Indicate the relative importance of ideas presented in your lesson.

A professor of Politics says, "I highlight major points by saying, 'This is more important than that.' For example, if I am giving a list of six contributing factors to some phenomenon or event, I deliberately identify which in my view are most important. I don't want students to go away thinking that everything I say is of equal weight or importance."

15. Explain or demonstrate to students why a particular point is important.

Several teachers believe that the best way to cue their students to the importance of an idea is to show them the role that idea plays in an overall understanding of the course material or in applications beyond the course.

"I think it is crucial for students to know why a concept is important," says one Biology teacher. "Just saying that it is important is not enough. You need to put the concept in some perspective, to show why it is important. Explaining why an idea is important not only gets your students' attention, it gives them a framework on which to hang the idea."

Another science teacher concurs. "I follow the introduction of a major concept with lots of specific examples, including anecdotes which show application of the concept in current professional practice," he explains. "You must show your students why it is important to know a particular concept if you expect them to master it."

16. Call on your students to paraphrase or summarise what you have just said.

"Asking your students if they understand gets you only so far," one History teacher explains. "Asking a student to summarise the main things to remember about X, and then asking other students to help out if she is having difficulty is a far better check on your students' understanding."

Asking questions of specific students has other benefits too. For example, because your students know that they may be called upon, they listen more attentively for the main ideas and that in turn helps them to organise their notes better.

17. Create opportunities for role-playing.

A Law teacher makes use of role playing to encourage his students to develop skills they will need in their careers. "I give my students copies of a legal report, for example. Then one half of the class is asked to assume the role of the authors of that report and

prepare an oral presentation for the client. The other half of the class is assigned to act as representatives of the client and to prepare questions to be asked of the lawyers.

"About a week later, during class time, I select certain students to actually enact these roles in front of the class. My students do not know ahead of time who will be called upon, so everyone has to be prepared. Those not called on join me in the role of the observer. When the students have enacted the meeting, the rest of us review each side's performance."

18. Assign provocative or controversial topics for papers.

"I find that the quality of the essays I get often depends on the quality of the assignment I give," says a Business teacher. He tries to give provocative titles for essays. For example, in a recent assignment he asked his students to respond to the question, "If you were working in a company that illegally pollutes the environment what would you do and why?" Giving provocative assignments challenges his students and makes for more interesting work.

One teacher who successfully engages students this way warns that it is important, even when deliberately trying to be provocative or "realistic", to choose topics that the students are "ready" to deal with in the context of the material being covered.

The importance of getting to know what gets the class "fired up" and what they relate to is emphasised by several lecturers in setting appropriately provocative assignments.

19. Give assignments which put your students in the role of another.

A History teacher reports that she used to give rather standard, sometimes dull, writing assignments.

"Most recently, however, I have asked my students to read a collection of the 18th century speeches on why Louis XVI should be killed and assigned them the task of writing their own speech as if they had been living during the French Revolution.

"Many students really are enthusiastic about this kind of assignment and do an incredibly good job. It helps them to identify with the issues of the time; in fact many of my students went to great lengths to research the authenticity of their own empathic interpretations."

20. Assign "thought problems" which are typical of the problems faced by professionals in the field.

A Biology teacher assigns weekly "thought problems" which are of the same type of questions professionals in the field are asked, such as, 'What is killing that tree?' not 'Name six factors which can kill trees."

21. Be true to yourself and don't try too hard.

A humanities teacher says it is important to remember that students will 'see through' her if she tries too hard to make an activity exciting or 'whizzy' simply for the sake of it. She felt it important that her students recognised they were being directed toward authentic activities that supported their learning.

22. Don't do it alone.

Of all of the aspects of teaching in this project, several teachers saw the Learning and Academic Value aspect as being the one most difficult to improve on your own. Most teachers, it was felt, go into the classroom trying to do their best and, if a low evaluation score is received for this area, it may be very difficult to see clearly how to improve on your own. Seek help from other colleagues by observing lessons or simply asking for tips.

These strategies are adapted with permission from a package of materials available in: Marsh, H. W., and Roche, L. A. (1994). The Use of Students' Evaluations of University Teaching to Improve Teaching Effectiveness. Canberra: Department of Employment, Education and Training. Further information on the Improving Academic Teaching Project can be found in Marsh, H. W., and Roche, L. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. American Educational Research Journal, 30, 217-251.

Many of the ideas presented in the original package of materials were adapted with permission from Davis, B.G., Wood, L., & Wilson, R. (1983). ABC's of Teaching with Excellence. Teaching Innovation and Evaluation Services, University of California.

The authors would also like to express their appreciation to the many teachers who contributed strategies and other suggestions during the project.

Teaching strategy ideas for Teacher Enthusiasm

A minimal condition for learning is that attention be aroused. Stimulus salience, that is, the extent to which a stimulus stands out against a background, is known to be crucial in evoking interest and attention. It is to be expected, therefore, that teachers who impress students with their enthusiasm, appropriate dynamism and energy, and who make judicious use of humour will increase student interest and command the class's attention. Moreover, teacher enthusiasm can vicariously induce enthusiasm for the subject in students. Students who rate their teacher highly are more likely to model their behaviour towards a subject upon that teacher. Thus, the latter's enthusiasm can be acquired by students. Students whose interest in and enthusiasm for a subject are aroused are likely to have enhanced achievement in learning the subject. This factor is especially relevant to the principle that learners must be motivated to learn.

The following ideas were suggested and used by outstanding teachers across a range of institutions and disciplines. We suggest that you select no more than 3 or 4 strategies which appear potentially the most profitable, then make a commitment to apply OR adapt them to improve your teaching effectiveness.

1. Look for ways to relate the course material as a story.

"I always try to tell some kind of story," says a Biology teacher. "My primary belief about communication is that it doesn't matter what you say if you can't get them to listen."

The idea of teaching using a story allows for incremental, structured learning; high-quality planning and resources will help the 'story' be told effectively.

2. Vary the pace and type of activities in lessons.

An English teacher believes that his wide variety of teaching strategies accounts for his high ratings on interesting style of presentation. "I read whatever I can find on teaching in my discipline," he says, "and I borrow shamelessly from other teachers when it comes to pedagogical strategies."

Some of the variations used by excellent teachers include: student panel discussions, guest speakers, PowerPoint or Prezi presentations, films, interactive games on PCs, photographs, whiteboard work with coloured pens, role-playing and simulations, and a wide variety of group discussion techniques.

3. Focus your lessons around a common object, event, or phenomenon which exemplifies the major concepts of the course.

A Biology teacher calls this his "potato lecture." "Biology is an empirical discipline; it depends on observation and investigation. I pass out potatoes to all students in the class and begin a dialogue about the kinds of things they can observe about their potato. I have to overcome their previous experiences," he explains. "Although potatoes are familiar objects to them, they don't have the foggiest idea what a potato is. I stress what you can get out of everyday experiences by asking the right questions. I ask them about their observations, help them ask questions and describe ways they could investigate answers." An interactive exercise around a common phenomenon tends to "break the ice" between a teacher and their students. "After the lesson, some students cook their potatoes and others plant them. Months later, I still get stopped by students who want to tell me how their potatoes are doing," he says.

4. Invite guest speakers to your lessons.

An English teacher sometimes invites professional actors to talk about their interpretation of a scene or a role from a play his students are studying. "It's very important to make clear to a guest what you expect of him or her in order to ensure that it is an educational experience for your students," he points out.

"I always take detailed notes during a guest's talk," says a Biology teacher. "In that way I am able to answer student questions about the material during later sessions and may learn something new myself!"

An Art teacher prepares his guest speakers well in advance so that they know exactly what is expected of them. "Practicing artists are asked to submit working drawings,

models, photos, and publications on one of their pieces so that my students can be well acquainted with their work beforehand," he says.

"Students are asked to submit a set of questions to a guest speaker beforehand about his/her work. Designated students are given responsibility to see that these questions are addressed to the speaker." He makes a point to confirm the time and place of the guest's presentation, provide a map, arrange for parking, and promptly send each speaker a thank-you letter.

5. Learn to vary the pitch or inflection of your voice.

If your students complain that you teach in a monotone, you may want to try one of the following: taking speech lessons, taking acting lessons, joining (or organizing) a poetry or drama reading group, or simply practicing reading aloud to yourself or members of your family. Each of these methods has been used effectively by one or more teachers involved surveyed.

One teacher combined speech lessons with Bible readings at his church with good effect. Another took an acting class during the summer, and not only improved his vocal delivery in the classroom but had enjoyed himself as well.

Several teachers recommended joining a poetry-reading group. "Reading poetry aloud can be particularly helpful because poetry requires greater vocal inflection for its meaning to become clear," one said. Reading plays aloud with friends or family can also be an enjoyable means of practicing vocal variety.

7. "Open with gusto" and "Finish strong."

One teacher feels there are distinct advantages to giving special thought to the beginning and ending each lesson.

"The opening should secure students' attention and give them the desired mental set. Get off to a good start. Do something to command attention from the outset. Put punch into your opening.

"Have some form of attention-getter.....a gadget or piece of software whose operation depends upon the principles of the day's lesson usually excites attention. Carefully planned questions or statements can also develop the curiosity necessary to get students' attention.

"Action is always an attention-getter. If you intend to use graphs or information on a board, keep them hidden from students' view as they come into the classroom. Likewise

for models and experiments. This will usually whet students' curiosity and make them eager to see what is going to happen.

"The ending is as important as the beginning. Don't let a class session fade into nonexistence. Make an impressive ending. For example, end with: a question for the class to cogitate and answer before next lesson; a quotation conveying the essential theme; a summary; a plenary."

11. Begin a lesson with an incident, example, or anecdote to get your students' attention.

An attention getter does not have to be "gung-ho" or "whizz-bang" - carefully planned questions or statements which are provocative, controversial or paradoxical can be quite effective.

A History teacher says that he often begins by reading aloud a short passage from a primary source or a story to illustrate his major theme or point in the lecture. "For example, I start out by stating that the Wizard of Oz is a parable for progressivism and read passages from it to illustrate my major point. I then get my students to help identify the different characters and what they represent. I usually end with a quotation that pulls together what I have been trying to say," he says. "Also whenever possible, I try to link the past with current events, to show how the topic is important for the present."

One teacher of MFL has a selection of defunct Weimar banknotes on his desk. He believes that 'personalising' lessons and the classroom environment with items that engage students' curiosity creates a starting-point for discussions outside the specification's confines.

14. Make diagnostic and practice videos.

Although videos of your teaching may not be adequate to diagnose fine points of pitch, inflection, articulation, and pronunciation, they can be used effectively to note whether you speak too slowly or too rapidly, whether you vary your tone and inflection sufficiently to hold your students' attention and communicate meaning, and whether you articulate clearly and/or forcefully enough to be heard and understood.

Videos can also be used to check the organization of your lesson and the clarity of your explanations. Repeated videoing will also allow you to monitor your improvement on any of these variables.

One Biology teacher has had his lectures videotaped many times. "The first time was a shattering experience," he says, "but it is the most effective kind of feedback you can get. I have found videotape invaluable for getting rid of annoying mannerisms, for

learning to vary the speed of my delivery and to put more expression and greater clarity into my explanations."

15. Use your students to monitor your presentation.

If you want an in-class reminder of when you are speaking too softly, too rapidly, or without sufficient articulation to be heard and understood, ask one or more of your students to give you a pre-determined signal if your speed of delivery or articulation makes it difficult to understand what you are saying.

19. Confound yourself, and let your students "rescue" you occasionally.

Asking open-ended questions which the students can sense are mystifying to you personally is a great way to encourage relevant discussion and to model your enthusiasm for discovering the secrets of the subject.

A humanities teacher reports that "When I ask myself a question that initially seems puzzling to all of us, it lets my students know that I'm not omniscient, and that it's all right to ask questions or get confused occasionally. The class really responds to that."

A teacher of MFL says that he feels it important that students have the sense that they are engaged in a "legitimate intellectual process".

21. Develop effective ways to encourage students to see you about their difficulties.

"Enthusiastic teaching is reflected not only in how you relate to and represent the subject matter, but in how you relate and respond to the students," says a Science teacher.

Teachers who were rated by their students as most welcoming and accessible in the Improving Academic Teaching Project used cartoons, scribble boards or notepads on doors as a means of "attracting" reticent students to see them, and of providing a more relaxed, informal atmosphere. Such strategies can also help to promote a higher level of enthusiasm for the subject.

22. Take care to communicate your genuine concern for students after class.

"Students can be very sensitive to non-verbal messages implying that you are not genuinely interested, and this can quickly turn them off seeking help or pursuing an interest in the subject," warns an outstanding Physics teacher. "Some teachers seem to

fear that any further encouragement of their students to see them outside lessons would leave them inundated. It is actually quite easy to learn how to avoid negative rejection of students' requests, without devoting your entire day to them."

She explained that she made a point of never making students feel unwelcome. If a student came to see her at an inappropriate time, she would maintain a positive attitude, saying, for example, "I'd love to see you -- how about 4:30?" rather than "I can't see you now, I'm busy -- try again later".

23. Take time out for yourself to maintain a positive attitude and avoid excessive stress.

An outstanding teacher emphasises the need to have a positive working environment to maintain the enthusiasm of staff. "Educators often take on too many responsibilities and endure stressful or unsupportive working environments, while still expecting themselves to perform with verve and enthusiasm in their teaching role. It is important to direct some attention towards yourself, and the quality of your work environment, in order to keep the energy in your teaching," she explains.

While some aspects of the work environment cannot readily be influenced at the individual level, there are a number of things that teachers can do to maintain a positive and enthusiastic approach to teaching and avoid excessive work-related stress. "Timetabling into your personal schedule a regular physical activity, social meeting or 'walkabout' is a simple, effective means of keeping control and relieving stress," she says.

24. Fall in love again...with your subject.

A teacher of French said that she increases her enthusiasm for teaching her subject by revisiting the aspects of her subject that she loves, those which prompted her to teach the subject in the first place. "Falling in love again with your subject is a great way of helping you retain a freshness about your teaching." She suggests re-reading favourite texts, going to see new plays or simply discussing the subject with like-minded people.

25. Take acting lessons.

Several teachers suggested acting classes as a means to increase evaluation ratings on the teacher enthusiasm aspect of the survey. They felt strongly that teaching is, in part, an act and that, as such, acting is a skill that can be taught and learned. Equally, taking acting lessons may well have the secondary effect of giving you new anecdotes to share in your lessons as well as enabling you to meet new people.

These strategies are adapted with permission from a package of materials available in: Marsh, H. W., and Roche, L. A. (1994). The Use of Students' Evaluations of University Teaching to Improve Teaching Effectiveness. Canberra: Department of Employment, Education and Training. Further information on the Improving Academic Teaching Project can be found in Marsh, H. W., and Roche, L. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. American Educational Research Journal, 30, 217-251.

Many of the ideas presented in the original package of materials were adapted with permission from Davis, B.G., Wood, L., & Wilson, R. (1983). ABC's of Teaching with Excellence. Teaching Innovation and Evaluation Services, University of California.

The authors would also like to express their appreciation to the many teachers who contributed strategies and other suggestions during the project.

Teaching strategy ideas for Group Interaction

Learning in institutionalised educational contexts is a social phenomenon. That is, except in rare cases of individual tuition, teaching is delivered to groups of students ranging from small to very large in size. This aspect of teaching refers to verbal interaction in classrooms in the form of questions and answers facilitating the expression and sharing of ideas and knowledge. Higher ratings on this aspect of teaching suggest that the motivational potential of social interaction with others in learning contexts is being capitalised upon and also that the classroom context is being exploited as a venue for activity in practising and testing ideas, and engaging in a feedback cycle. As such, the Group Interaction aspect has a strong basis in principles of teaching and learning.

The following ideas are suggested and used by outstanding teachers across a range of institutions and disciplines. We suggest that you select no more than 3 or 4 strategies

which appear potentially the most profitable, then make a commitment to apply OR adapt them to improve your teaching effectiveness.

1. Face openly any reservations or fears you may have about encouraging group discussion and interaction.

Many teachers feel uncomfortable about implementing or keeping control of class discussions.

"It is important to recognise potential problems and limitations involved in a discussion format, without rejecting it as an option," says an outstanding Economics teacher. "The decision regarding whether to encourage more discussion with and between students should be based on a careful search for ways to maximise the utility of discussion, and to minimise any limitations - even if that does sound like a cold economist's approach!"

Potential concerns include covering less material, getting "side-tracked", and causing confusion among students who experience difficulties in understanding the purpose, relevance or flow of the discussion.

The ideas presented below provide a number of solutions to these and other concerns.

2. Plan your sessions to incorporate time for discussion.

Allowing time for discussion or other group activities often means that fewer points can be addressed in a given session. Acknowledging this in advance, a number of teachers set homework to cover additional content.

"I was initially reluctant to assign material which I had not specifically covered in lessons, but I quickly found that when students knew in advance what was required, and how the discussion related to the content, they were highly motivated and productive, both during the session and in their homework," a Physics teacher revealed.

3. Identify discussion questions in advance.

Students are more inclined to participate when they know the focus or intent of a discussion. A preview of discussion topics can help your students to organise their thinking and prepare to express their views. Several teachers develop discussion questions in advance and distribute them to their students.

"In my course," one teacher explains, "I give my students a series of discussion questions on each week's reading homework. These are spelled out in my course plan which is handed out to students during the first week of class. All of my students are responsible

for all of the questions each week. These questions serve both as study prompts and stimuli for discussion."

4. Divide the class into smaller groups with a formalised structure.

One teacher divides his class into small groups of students. Each group is assigned a specific question or topic to discuss, selected from a list of questions prepared in advance. But, because students do not know beforehand which questions their group will be assigned, they must be prepared to discuss all of them.

The teacher assigns one student in each group to be the discussion leader, another to be the group's summariser, and a third to be the group's evaluator. Each group conducts its discussion in the manner it feels will be most effective. During the discussion, the teacher moves back and forth among the groups, noting any issues he may want to bring up or clarify at the end of the class.

After the groups have discussed their respective topics, the class is called back together and each group presents the results of that group's discussion, highlighting key terms or other information felt to be important. Each group's evaluator then provides some observations on how well the group functioned and makes suggestions as to how it might have functioned more effectively. During the course of the term, each student serves at least once as a group discussion leader, a summariser, and an evaluator.

5. Create an appropriate physical setting for the discussion, or use a fixed setting appropriately.

It is difficult for students to talk to people they cannot see. In a typical classroom, with fixed seats facing forward, students tend to direct their comments to the front of the room - to their teacher - rather than to other students. This arrangement encourages one-to-one dialogues rather than group discussion. If, on the other hand, students can see each other, they are more likely to interact with one another as well as with the teacher.

If there is no alternative to fixed seating, discussion may be best limited to pairs of students, who are then invited to report back to the teacher. The teacher can also move around the room to focus attention on a variety of contributors to a discussion.

A circle or U-shaped arrangement of chairs is the most useful for discussion. Teachers also find that if they sit with their students rather than stand in front of them or sit behind a table, it helps promote group discussion rather than student-teacher exchanges.

6. Get to know your students: Where they are "at"; and what they relate to.

Knowing your students is important for a number of reasons. Several outstanding teachers stressed that new learning must begin from what students are already familiar with. "Otherwise they quickly become confused, disinterested or anxious," one teacher explains. "Students will also open up more in class discussion if they feel a comfortable rapport with the teacher".

Getting to know what gets the class "fired up", or what they relate to is a strategy that several teachers have recommended as a means of generating a vibrant learning atmosphere. This requires establishing good rapport and making time to chat with students in non-teaching situations.

7. Confound yourself, and let your students "rescue" you occasionally.

Asking open-ended questions which the students can sense are mystifying to you personally is a great way to encourage relevant discussion and to model your enthusiasm for discovering the secrets of the subject.

A distinguished teacher reports that "When I ask myself a question that initially seems puzzling to all of us, it lets my students know that I'm not omniscient, and that it's all right to ask questions or get confused occasionally. The class really responds to that."

8. Explain the purpose of discussion.

To get your students involved in class discussion, it is helpful to explain the value of their participation and what they can expect to get out of the experience.

A Business teacher stresses the importance of explaining the benefits of discussion with students. "My students don't know how to participate in a seminar discussion so I make a point of telling them what skills they will acquire: how to speak and discuss their ideas, how to listen and respond to the ideas of others."

In seminar formats, especially, many teachers find that it is worthwhile taking some time to teach their students how to listen to others, how to paraphrase, how to involve other members of the group. "Students have to understand that in a seminar they share the responsibility for making the discussion a worthwhile experience for us all," says one teacher. "This is a new idea for most of them."

Several teachers have found that a clear explanation of the purpose or value of particular activities motivates the students to respond positively, and helps to focus their energies.

9. Force yourself to lengthen your "wait time" after questions and after answers.

One teacher points out that it is well worth waiting an extra few seconds before continuing to speak after asking a question or after a student contributes something to the class.

It takes a conscious effort to wait beyond the normal 'comfort zone', she warns, "but my experience has confirmed research which shows that both teachers and students make great mental use of such a pause - at least three seconds - to work through what is being said or asked, leading to better quality questions, answers and discussion in general."

"Students in Mathematics are often not quite confident enough to offer answers, or even ask for clarification, for fear of sounding ignorant," explained another teacher. "By pausing for an extended period, and looking around at the faces in the class, students are more likely to decide to 'bite the bullet'. Sometimes it gives them a chance to formulate their question to their satisfaction, sometimes they sense that I'm waiting to hear from them (which reinforces their confidence that their question is appropriate), and sometimes they sense that I won't go on until I've heard from someone. It can be a little uncomfortable for all of us occasionally, especially early on, but the gradual growth in confidence and competence shown by the students is definitely worth the effort".

10. Move around the room in a way which will promote discussion.

A Business teacher finds that the way he moves around the room alters the kinds of interaction he is able to generate among his students. "When a student asks a question, it is natural for a teacher to move toward that student," he points out. "However, this tends to exclude the other students and focuses the interaction on the teacher and the one participating student.

"In order to draw my other students into the discussion and to get them to address their comments to one another as well as to me, I find that it helps if I move away from the student who asks a question rather than toward him or her. This forces the student to project so that everyone is drawn into the conversation. It also makes it more likely that the student will address fellow students."

Moving around the room during group discussions also allows you to monitor the direction that groups are taking in their discussion.

11. Redirect your students' questions: basketball vs. Ping-Pong

Whenever you have reason to believe that there are students in your class who know the answer to a student's question, it is useful to redirect the question to one of those students or to the class as a whole. A Science teacher, for example, says that in discussions he tries hard not to answer his students' questions directly unless he doubts that anyone in the class would be in a position to give the correct response. Some teachers call this technique 'Basketball Questioning', a term used to refer to the question being 'bounced' around the room between students (instead of being answered by the teacher). Using this technique is the opposite of 'Ping-Pong Questioning', in which the interaction is one-way, between the student and the teacher only.

12. Call on students who might provide an interesting viewpoint.

"I call on students whom I think might have a different perspective or set of experiences relevant to a given topic or issue," says a Politics teacher. "I try to take advantage of the probability that students who like outdoor activities have different experiences and attitudes about environmental issues, or that female and male students view things such as childcare differently."

Several teachers stress the fact that by getting your students to talk about their experiences, you can greatly increase the amount of knowledge all your students take away from the course.

13. Encourage your students to write papers related to their backgrounds.

An English teacher encourages his students to make use of knowledge and skills developed in other courses in combination with those emphasised in his course. "I strongly encourage my students to write essays on interdisciplinary topics," he says. Examples include: a student taking Biology wrote a paper on "Shakespeare and Plants," and an Art student analysed the connection between the paintings of Watteau and imagery in Pope's "Rape of the Lock".

"If you can get your students to realise that they each bring different kinds of talent and expertise to the course and encourage them to apply these, that goes a long way toward motivating them to do their best work.

14. Encourage your students to make presentations in class.

"Sometimes my students come up after class and pose an interesting question or make an insightful comment," says one teacher. "Often I encourage those students to pursue the topic in more detail and then make a brief presentation to the class. When possible,

I try to get several students with complementary interests to work together on a project of this kind."

15. Assign your students specific leadership responsibilities.

"I find this very effective in getting my students to take responsibility for class discussions," notes an Art teacher. Students select topics for which they will serve as discussion leaders. The number of student leaders per topic depends on the size of the class (usually from one to three students per topic). Each student, either alone or with other students, leads a discussion two or three times per term.

"The leaders' task is to prepare a set of three to six discussion questions about the material they have read or viewed. These discussion questions are handed out to the rest of the class the week before the topic is covered. If there is more than one leader for a topic, the leaders assume responsibility for facilitating the discussion."

16. Call on your students to paraphrase or summarise what you have just said.

"Asking your students if they understand gets you only so far," one History teacher explains. "Asking a student to summarise the main things to remember about X, and then asking other students to help out if she is having difficulty is a far better check on your students' understanding."

Asking questions of specific students has other benefits too. For example, because your students know that they may be called upon, they listen more attentively for the main ideas and that in turn helps them to organise their notes better.

These strategies are adapted with permission from a package of materials available in: Marsh, H. W., and Roche, L. A. (1994). The Use of Students' Evaluations of University Teaching to Improve Teaching Effectiveness. Canberra: Department of Employment, Education and Training. Further information on the Improving Academic Teaching Project can be found in Marsh, H. W., and Roche, L. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. American Educational Research Journal, 30, 217-251.

Many of the ideas presented in the original package of materials were adapted with permission from Davis, B.G., Wood, L., & Wilson, R. (1983). ABC's of Teaching with Excellence. Teaching Innovation and Evaluation Services, University of California.

The authors would also like to express their appreciation to the many teachers who contributed strategies and other suggestions during the project.

Teaching strategy ideas for Teacher-student Relationship

Opportunities to provide for individual differences in capacity and to take account of learners' present knowledge and attitudes to education depend heavily upon individual contacts with teachers. Furthermore, individual tuition and guidance are available to the extent that teachers are interested in and accessible to individual students. Students who feel welcome also have greater access to motivationally significant opportunities such as face-to-face reinforcement and encouragement.

The following ideas are suggested and used by outstanding teachers across a range of institutions and disciplines. We suggest that you select no more than 3 or 4 strategies which appear potentially the most profitable, then make a commitment to apply OR adapt them to improve your teaching effectiveness.

1. Acknowledge the importance of getting to know your students: where they are "at"; and what they relate to.

Knowing your students is important for a number of reasons. Several outstanding teachers stressed that new learning must begin from what students are already familiar with. "Otherwise they quickly become confused, disinterested or anxious," one teacher explains. "Students will also open up more in class discussion if they feel a comfortable rapport with the teacher".

Getting to know what gets the class "fired up", or what they relate to is a strategy that several teachers have recommended as a means of generating a vibrant learning atmosphere. This requires establishing good rapport and making time to chat with students in non-teaching situations.

2. Arrange an individual appointment with each student.

One outstanding teacher stressed the importance of getting to know students as people, rather than simply as students. "This is central to making the material relevant, opening up discussion, and generally meeting their learning needs," she explained.

A Maths teacher felt that he was not being successful in generating class discussion. At the end of the third week, still unable to encourage class participation, he decided to pass around a sheet of paper with a list of 10 minute blocks of time when he would be available for individual appointments.

Each of his students was required to sign up for one of the 10 minute sessions. They were told that the chief purpose was for him to get to know his students better and to listen to any suggestions they might have or concerns about the course.

"I found that this was a real ice-breaker," he explains. "Even though most of our discussions were mainly chit-chat, some of my students used the opportunity to indicate problems they were having in the course or to make suggestions about course improvements. Perhaps the chief benefit was that it gave me an opportunity to get to know my students. As a result, they seemed to feel more comfortable asking and answering questions in class."

3. Schedule specific topics to cover in 'open surgery' sessions.

"I find it useful to identify in advance a specific topic for my 'open surgery' sessions," says a languages teacher. "I encourage my students who are having difficulty in that area to come for help." Based on past experiences she knows which concepts and ideas cause problems and she schedules these sessions to provide further elaboration and discussion on these topics.

"This way, if one of my students misses a lesson or doesn't fully understand the topic, he or she has an opportunity to catch up with me.

Another teacher uses one extra-curricular period a week in a similar fashion, although the specific topics are not necessarily ones covered by the course. "Sometimes they are enrichment topics; sometimes they are remedial, focusing on exam technique, perhaps," he says.

As an added bonus, students and teacher get to know one another in a small informal setting.

4. Take care to communicate your genuine concern for students after class.

"Students can be very sensitive to non-verbal messages implying that you are not genuinely interested, and this can quickly turn them off a course," warns an outstanding Physics teacher. "Some teachers seem to fear that any further encouragement of their students to stop by the staff room would leave them inundated. It is actually quite easy to learn how to avoid negative rejection of students' requests, without devoting your entire day to them."

She explained that she made a point of never making students feel unwelcome. If a student dropped came at an inappropriate time, she would maintain a positive attitude, saying, for example, "I'd love to see you -- how about 4:30?" rather than "I can't see you now, I'm busy -- try again later".

5. Have your students fill out a brief questionnaire about their backgrounds and interests.

Typically, such questionnaires ask for information about students' other subjects, extra-curricular experiences, and career plans. Teachers use this information to understand their students and to call on those whose experiences may give them a different perspective.

A Politics teacher, for example, asks students to describe their most memorable experience that relates to the subject matter of the course. A History teacher asks about students' travel experiences or knowledge of the peoples of the area being studied through relations, friends, or through reading.

6. Develop effective ways to encourage students to see you about their difficulties.

Teachers who were rated by their students as most welcoming and accessible in the Improving Academic Teaching Project used cartoons, scribble boards or notepads on doors as a means of "attracting" reticent students to see them, and of providing a more relaxed, informal atmosphere. Such strategies can also help to promote a higher level of enthusiasm for the subject.

7. Meet regularly with your students who do poorly on initial assessment tasks.

One teacher gives assessments early enough in the course to allow him to identify those students who may be having difficulty before it becomes too problematic. After the first assessment, he asks each of his students who did not pass to talk with him about the exam results. In these meetings he tries to discover each individual student's problem.

"I ask questions such as 'Did I misread anything you gave as an answer?' or 'What problems did you have in taking the exam?'"

He concludes each meeting by telling his students that he is certain that they can do better and striking a bargain with them. "Usually, I tell them that I won't record the first mark and that I look forward to seeing improvements," he says, "on condition that they agree to meet with me every week or two to go over homework assignments and to get additional help.

"About nine or ten students take advantage of this help each term," he explains. "Giving students a second chance, I find, is a powerful motivator."

8. Require all your students who do below pass level work on assignments or quizzes to see you.

One teacher writes a note, "Please see me" to students who score below the pass mark on his weekly tests.

"It's important to find out why students score low," he explains. "If they are having difficulty understanding the material, I offer to help them. If it's a question of motivation or a student placing less priority on their studies in my subject, I help them to make improvements where necessary. It helps me as a teacher to know the reasons for the poor performance. Showing concern is also a powerful motivator for some students: they begin to do better."

A Science teacher concurs. "I ask to see students who get less than 50% on the biweekly quizzes," he says. "In a way, I play parent with them; I 'sit on' them a little. I think I understand better now than when I began teaching the need some students have for external motivation."

9. Provide a relaxing informal atmosphere when appropriate.

"I bring coffee and biscuits periodically to my lessons," says a teacher. "This helps relax my students and lends a congenial tone to the discussions."

"I find that this simple act seems to make the sessions more interactive. People tend to discuss issues over coffee and biscuits more readily than in a fixed formal classroom setting."

One teacher of Business and Economics set up what he called 'Businomics Cafe' during break times. He supplied coffee, biscuits and copies of the most recent Financial Times and Economist publications as a source of discussion material.

10. Hand out brief excerpts or abstracts of contextual material to fill in cultural gaps in your students' knowledge.

One Religious Studies teacher distributes such handouts fairly often. "I don't expect my students to know a great deal about the geography, religions, and literature of the Near Eastern countries and cultures we are studying," he says. "However, more and more, I find that I cannot make many assumptions about what my students know about Western culture either."

"This poses some difficulty for me as well as for my students, because I believe one of the best ways to teach something 'foreign' is by analogy to something that is familiar. Yet, in order to do that, I find it is increasingly necessary to provide information on the so-called familiar Western examples as well." Among the contextual materials he distributes in excerpted or abstracted form are fables, fairy tales, nursery rhymes, biblical stories or quotations, and Greek mythology.

11. Give a diagnostic test at the beginning of the term.

One Biology teacher frequently gives a diagnostic test covering knowledge and skills needed to access the course or an individual unit. The test, which is given in the first week, is not given a grade. "Its sole purpose is to help me identify those of my students who need extra help so I can begin working with them early in the unit or course.

"Students need to recognise their weaknesses and begin to correct them if they are to succeed on the course. But they have to be given the means for correcting deficiencies. I meet with students whose preparation is inadequate and help them to plan and prepare on as regular a basis as possible."

12. Make special efforts to integrate your weaker students into the class through small group work.

One foreign language teacher divides his students into small groups. "I pose a question to each group," he explains. "One student in each group gives the answer orally; a second student corrects the first student, if necessary; and the third student writes the answer on the board. Each student has a role, and these roles are rotated throughout the quarter.

"Initially I assign my weaker students to do the boardwork," he says, "although I am careful not to do this in an obvious way." This allows the weaker students to participate, but in a way which will reinforce their own learning without holding back the others. "Also I often ask a better student to help out if a weaker student is having difficulty responding. Then I have the second student repeat the question to the first student to

give him another shot at it. Peer teaching can be extremely effective," he says, "especially when a class takes responsibility for its weaker members."

These strategies are adapted with permission from a package of materials available in: Marsh, H. W., and Roche, L. A. (1994). The Use of Students' Evaluations of University Teaching to Improve Teaching Effectiveness. Canberra: Department of Employment, Education and Training. Further information on the Improving Academic Teaching Project can be found in Marsh, H. W., and Roche, L. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. American Educational Research Journal, 30, 217-251.

Many of the ideas presented in the original package of materials were adapted with permission from Davis, B.G., Wood, L., & Wilson, R. (1983). ABC's of Teaching with Excellence. Teaching Innovation and Evaluation Services, University of California.

The authors would also like to express their appreciation to the many teachers who contributed strategies and other suggestions during the project.

Teaching strategy ideas for Breadth of Coverage

This aspect of teaching reflects students' responses to items concerning the contrasting of implications of various theories, the provision of the backgrounds of ideas and concepts, the presentation of different points of view and the discussion of current developments. These all have to do with substantive qualities of teaching. Each would seem to have the potential to increase student knowledge and understanding through facilitating generalisation beyond the confines of the specific situation, to clarify the material to be learned and its meaningfulness to the learner.

The following ideas are suggested and used by outstanding teachers across a range of institutions and disciplines. We suggest that you select no more than 3 or 4 strategies which appear potentially the most profitable, then make a commitment to apply OR adapt them to improve your teaching effectiveness.

1. Select an article which represents one theoretical viewpoint and build your lesson(s) around an opposing idea or position.

A teacher of Economics, for example, assigns a reading that represents the point of view of liberal economists, but designs his teaching sessions around the opposing views of leading conservatives or radicals.

In addition to assuring a balanced presentation, this approach adds variety and interest to the course and stimulates students to think critically. Because the taught material complements rather than repeats the article, it has the added benefit of increasing student attention.

2. Contact university departments or other experts in the field to get the latest information.

Before giving her lesson on the heart, a Biology teacher contacts their local university to get the most recent statistics and findings on heart transplants. Similarly, a Law teacher directly contacts lawyers involved in important cases pending or under adjudication, making her lessons as up-to-date as possible.

A Geography teacher routinely calls his contacts in a local university to get the latest information on environmental legislation which he incorporates into his lessons.

This kind of up-to-the-minute reporting on a few major developments relevant to course content can help you convey a sense of the excitement of research to your students.

3. Invite guest speakers to offer contrasting viewpoints.

One teacher makes a point of doing this in his courses so that his students are exposed to a variety of positions. "I want them to understand what the different points of view are," he says, "and one of the best ways I have found to do this is to invite a colleague or practitioner whom I know to be an adherent of each view to make a presentation to the class."

"I always take detailed notes during a guest lecture," says a Biology teacher. "In this way I am able to answer student questions about the material during later sessions and may learn something new myself!"

4. Assign multiple readings to represent a variety of viewpoints.

"Because the most controversial issues covered in a course are ones on which my students have strong opinions but little information, I try to expose them to diametrically opposite positions or theories," says one Politics teacher.

"Developing a set of readings takes time. However, I can usually use it for two or three years with only minor modifications.

5. Assign readings directed toward revealing the reasons behind differing points of view.

A Business teacher says, "I lead my students through an analysis and critique of each theorist's position." The focus is not on opinions but on the reasons behind them. "Sometimes my own view is apparent either explicitly or implicitly; other times it is not. Sometimes it is not possible to find a reading which gets at the basis for a particular point of view. However, any reading that presents a clear statement of the features of the theory is useful. Students can be directed to a lively discussion of reasons that are tenable. It gives them experience in learning the criteria of a good argument."

6. Use your students' opinions to create a microcosm of society's attitudes on social, political, and economic issues.

At the beginning of the term, a teacher of Economics gives his students a questionnaire in which they are asked to agree or disagree with a series of controversial statements on the functioning of the economy. "I use the survey results to introduce a variety of student viewpoints. Throughout the term I reveal selected results from the survey as these relate to new concepts or issues covered in homework and lessons."

"This technique gives my students a sense of personal involvement in the subject matter. Students learn that some of their peers may share their viewpoint. They also learn that some of their peers don't share their viewpoint and why. Use of student data allows me to introduce most of the views currently reflected in the society as a whole."

7. Draw upon the diverse backgrounds and experiences of your students to introduce different points of view.

At the beginning of the term, a teacher of Business asks his students to give written answers to questions about their backgrounds and reasons for taking the course. He asks students to focus particularly on experiences which might give them a particular viewpoint on social, political, and economic issues to be covered in the course.

Using a seating chart he calls on students whose prior experiences or interests may be relevant to a topic under discussion. In this way a full range of views is introduced in the course. "Often, with little or no effort, I am able to get students debating between themselves. In fact, I rarely give my own point of view until there has been a full discussion of the different points of view within the class itself."

This technique has additional advantages: introducing personal experiences and opinions makes the class livelier; and the teacher is given a method for learning about the students' backgrounds, experiences and personalities.

8. Present each of several competing theories as if you were an adherent of that position.

A Psychology teacher introduces three major approaches or schools of thought in the field. "I discuss each one historically and contrast the basic elements and implications of each," he says.

"I really don't have a point of view when I teach the course. There is so little known with impartial certainty; I don't think one is justified in taking a position. Therefore, I present the best case for each theory, then analyze each critically and comparatively."

Several other excellent teachers report that they also present the best case for each of several competing theories when teaching a course.

9 .Encourage students to take an approach different from the one you have adopted.

A teacher of English uses this strategy in all of his literature courses. "I always approach literature from an historical point of view: history is a particular passion with me," he says. "At the same time, I point out that there are many other perspectives and encourage students to use alternative approaches, e.g., the feminist approach."

10. Point out explicitly that there are alternative points of view.

"I indicate the polar principles which guide much of the research in the social sciences as well as much of our folk wisdom, e.g., 'opposites attract' versus 'birds of a feather flock together' or 'absence makes the heart grow fonder' versus 'familiarity breeds contempt'" says one teacher. In doing so I point out that they should be mindful that

there may be good reasons to believe the opposite of what a theory presents; that they should analyze all arguments in terms of their opposites."

11. Touch base repeatedly with the fundamentals or basics.

One Science teacher believes that too much of science is presented to students in a rote, plug-in-the numbers way.

"There are thousands of formulae," he points out, "but all of these are variations on a limited number of basic ideas of theories." "These basic ideas are 'ideal theories' from which are derived all the 'approximate' or 'technical theories' which scientists use."

"I try to teach my students how to judge when they can use an approximate theory with confidence and when they are obliged to go to a more rigorous level. In this way, I keep touching base with the fundamentals to reinforce students' understanding of them."

12. Focus your course on the classic issues and concepts in your discipline.

A History teacher explains that she has moved away from presenting the most esoteric and up-to- date concerns of professional historians to her students.

"The most interesting issues and themes for undergraduates," she explains, "generally turn out to be those which originally excited historians about a particular person, event, or epoch, not the controversies of present-day historians. The classic issue are the ones which attracted me to the field," she says, "and I find that they are still the most exciting for my students."

Following this approach does not mean that you cannot introduce newer research findings where they are relevant, of course. Nor does this suggest that ideas which have little or no current validity should be taught. It does mean that, in limiting your coverage, you select the major classic themes and concepts wherever possible.

13. Share your professional "junk mail" with your students.

One teacher makes a point of passing around announcements for local conferences, program proceedings, and advertisements for new books and journals in the field. "In this way I inform my students about professional activities and recent developments of which they might not otherwise be aware," he says.

"I also encourage my students to attend meetings and conferences and to read articles on topics of interest to them."

A staff member teaching French also shares copies of newsletters, newspaper clippings, and announcements of French movies, plays, or other cultural events in the area. "My students are often amazed and delighted to learn that there are so many opportunities to strengthen their language skills and to expand their understanding and enjoyment of French culture," he explains.

14. Require your students to read current journal articles.

"It's important for my students to be exposed to state-of-the-art ideas," says one Politics teacher. "I try to make sure that I direct students towards at least a few recent journal articles. [Sometimes] I translate the abstract of a journal article into layman terms. I present the basic findings in a narrative fashion, using little actual data."

"I want my students to become excited by the open-ended nature of our subject. I want them to understand that what they are learning is not the final word."

15. Require your students to read current newspapers or periodicals.

A teacher of Economics assigns the Tuesday editorials of the Wall Street Journal each week. She uses them as a basis for discussion, and she has her students compare them with textbook presentations on related topics.

A teacher of Biology also believes strongly in making use of articles in current periodicals. "I keep my eyes open for stories on recent developments which have become part of the 'current events' literature," he says. "For example, in a discussion of recombinant DNA, I was able to use photos from a recent issue of Life Magazine, as well as a story a broadsheet did on the Genetech Corporation."

16. Tell your students about local events which will expand their understanding of your subject.

"I distribute a calendar announcing course-related events in the area," one staff member explains. "The events include plays, lectures, demonstrations, poetry readings and so forth. In this way the content of my course is expanded far beyond what I can actually cover in class. I also encourage my students to use these local resources in their research and writing assignments."

17. Assign provocative or controversial topics for essays.

"I find that the quality of the essays I get often depends on the quality of the task / title I give," says a teacher of Business. He tries to give provocative tasks. For example, in a recent assignment he asked his students to respond to the question, "If you were working in a company that illegally pollutes the environment what would you do and why?" Giving provocative assignments not only challenges his students and makes for more interesting reading but also diminishes the chance that the essays will be plagiarised.

The importance of getting to know what gets the class "fired up" and what they relate to, is emphasised by several teachers in setting appropriately provocative assignments.

Teachers also noted that such assignments are an excellent way to encourage students to consider a broader range of issues in their responses.

18. Assign "thought problems" which are typical of the problems faced by professionals in the field.

One teacher assigns weekly "thought problems" which are of the same type of questions professional in the field are asked, such as, 'What caused the erosion in this picture?'; not 'Name six factors which can cause erosion."

Using real-life problems to encourage thoughtful reflection and/or discussion in this way, rather than requiring solutions, can be a particularly useful way to avoid overwhelming students with the complexities inherent in some tasks.

19. Give assignments which put your students in the role of another.

A History teacher reports that she used to give rather standard writing assignments, e.g., "compare author X and Y's views on A," where the two authors tended to be professional historians. "Most students, however, find the arguments of current historians somewhat arcane," she says.

"Therefore, most recently I have asked my students to read a collection of the 18th century speeches on why Louis XVI should be killed and assigned them the task of writing their own speech as if they had been living during the French Revolution.

"Students really are enthusiastic about this kind of assignment and do an incredibly good job. It helps them to identify with the issues of the time; in fact many of my students went to great lengths to research the authenticity of their own empathic interpretations. Next year, I intend to take this assignment a step further by dividing my

students into small groups and having them actually deliver their speeches to the group."

These strategies are adapted with permission from a package of materials available in: Marsh, H. W., and Roche, L. A. (1994). The Use of Students' Evaluations of University Teaching to Improve Teaching Effectiveness. Canberra: Department of Employment, Education and Training. Further information on the Improving Academic Teaching Project can be found in Marsh, H. W., and Roche, L. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. American Educational Research Journal, 30, 217-251.

Many of the ideas presented in the original package of materials were adapted with permission from Davis, B.G., Wood, L., & Wilson, R. (1983). ABC's of Teaching with Excellence. Teaching Innovation and Evaluation Services, University of California.

The authors would also like to express their appreciation to the many teachers who contributed strategies and other suggestions during the project.

Teaching strategy ideas for Exams & Grading

The instructional value of examinations and grading lies in the quality of the feedback and in the stimulus to study they provide. The items comprising this aspect of teaching apply specifically to feedback and less specifically to motivational issues. Students' perceptions of fairness and relevance of assessment procedures are probably associated with their motivation to learn. However, this aspect of teaching's main basis in principles of learning is reinforcement in the form of knowledge of results and affective consequences of that knowledge.

The following ideas are suggested and used by outstanding teachers across a range of institutions and disciplines. We suggest that you select no more than 3 or 4 strategies

which appear potentially the most profitable, then make a commitment to apply OR adapt them to improve your teaching effectiveness.

1. Give your students the tools to manage their own success.

A languages teacher says that she finds the most successful approach to providing feedback to students places emphasis on students being empowered to manage their own success. She does this by ensuring that each student has a clear idea of what success in a piece of work looks like (by using a mark scheme and exemplar work for every piece completed), a clear understanding of the process that should lead to this success (engaging students in structuring a plan of work), and mechanisms to ensure that errors are identified and addressed (demarcating peer-review sessions and self-review sessions).

2. Balance the difficulty of test question.

A teacher of Business distributes test questions as follows: about 25% are reasonably easy questions that nearly everyone gets correct. About 50% of the questions require a little more sophistication but can be answered by most students. About 25% of the items are quite challenging and generally are answered correctly only by the top 5-10% of my class.

"A balanced test with easy, moderate, and difficult items gives my students an opportunity to show whether they have mastered the fundamentals of the course or have gone beyond the minimum," explains this staff member. "I try to give my students a feeling of satisfaction by providing them with an opportunity to express what they have learned."

3. Prepare answers to exams and tests to hand out as soon as your students turn in their work.

One Chemistry teacher prepares a handout of correct answers which he gives to students as they hand in their answer sheets and leave the room. "There is no point in making students wait several days to find out how they did," this teacher explains. "They are most interested in the results at the time of the examination, and it is at the time of the examination that the greatest reinforcement of the learning can take place.

Note that this method gives students immediate feedback even though it may be a week or more before the assignments can be returned with comments or grades.

4. Give your students at least one assignment which consists of several options.

One teacher of English gives five or six options from which his students may choose the one which sounds interesting and most allows them to do their best.

Examples of the options which he offers include: a piece of creative writing; a dramatic representation to be performed in front of the class (which can be a small group or team project); an original videotape to be shown to the class (which can also be a team effort); or a third essay (a "safe" option generally selected by students who are happier with more conventional activities). In addition, with his permission, students can create an option of their own if they wish.

"More than five or six options tends to confuse some students; it becomes too difficult to decide," he believes. "Two few options, on the other hand, restricts unduly my more creative and daring students." He encourages his students to take an interdisciplinary approach and to link content and skills from other courses.

5. Discuss the answers to exams, tests, or homework during the next lesson.

One Science teacher says that even if he cannot return marked homework or exams, he always discusses the answers in the next lesson. "I want to correct any misunderstandings and reinforce their learning as soon as possible," he says. "Students are much more receptive to this right after completing an assignment."

6. Include peer-editing of student work (projects, essays, design work) in your course.

An History teacher has students exchange essays to take home and edit. "The final essay is submitted along with a copy of the first draft with its edited corrections in red," he explains. "Each paper then receives two grades, one for the author and one for the editor."

In this way, students receive prompt informal feedback from a peer, followed by a grade and a formal critique by the teacher. This technique helps students acquire good editing as well as good writing skills.

7. Return a 'model answer' to your students along with their own corrected exams or essays.

A teacher of Business likes to provide a great deal of feedback to his students after exams as a way of re-emphasizing the themes of the course.

"I generally spend about half the class walking my students through an exemplar essay that I distribute to them along with their own corrected exams. I try to explain the ways

in which most of their responses differ from what I consider to be an exemplary answer or solution."

8. Allow your students to bring in one page of notes to use during a mock exam.

Several teachers found it useful to allow their students to take an A4 sheet of notes into mock examinations during term time. This decreases students' anxiety about having to memorise formulae and facts and allows them to focus on exam technique. Preparing these crib sheets helps students focus their studying. Restricting students to one page of notes forces them to synthesise the most important aspects of the course or topic.

"My exams are quasi-open book," explains one teacher. "Letting them bring in condensed notes not only gives them a crutch against anxiety, it also provides an excellent form of review."

9. Include a "bonus" question on your exams or tests which asks your students to write an exam question rather than an exam answer.

Some teachers include an extra question, sometimes phrased in the following way: "Almost inevitably teachers fail to ask you in an exam all those things for which you so carefully prepared. As it happens, writing good questions is almost as difficult as writing adequate answers. So, if you have the time and inclination, write an original exam question. You will receive between 0 and 10 points depending upon the quality of your question. JUST THE QUESTION PLEASE, DON'T SUPPLY THE ANSWER."

This technique helps establish good rapport with your students, gives you additional information on their sense of what is important in the course, and becomes an excellent source of future exam, test, or discussion questions for the course.

10. Distribute anonymous copies of five different answers to one of your essay questions.

A Politics teacher who does this tells his students that one of the five answers received an "A", one a "B", response, etc. Finally, I explain what I am looking for in a response and why I gave each sample response the grade I did."

"I am much more interested in helping students learn how to do well in the course than I am in grading them," he explains. "As a result of this discussion, student improvement in subsequent tests is often remarkable. I am convinced that the value of spending a small amount of class-time this way far outweighs any loss of coverage of additional content."

11. Share success criteria and educate students in ways of interpreting and implementing these criteria.

One outstanding teacher ensures that all of her students have a clear set of success criteria, in the form of a mark scheme. She then encourages her students to ask the question 'How can I make this work better?' and seek the answer with as little teacher input as possible. Early on in the course, she helps students understand the criteria, but increasingly expects her students to act autonomously.

12. Peer-marking prior to hand-in.

One teacher of German carefully structures the lesson when students hand-in work, so that an opportunity is given for each student to comment on the work of another (or their own work) in light of the marking criteria (or 'criteria for success' as he calls it). In doing this, each student has a further opportunity to receive feedback and amend their work prior to final submission. He cautions that this only works if students have been instructed clearly in the mechanics of the mark scheme.

13. Ask your students what they need.

Several teachers engage in a dialogue with their students, focusing on what the students need in terms of feedback. Being adaptable to students' needs in this area creates a cooperative approach to learning, but one in which the onus is on the student to identify what works for them. Carefully structuring the discussion with students is important, as students may not know what is feasible for the teacher providing the feedback.

14. Plot student performance on a graph.

One outstanding teacher uses a simple graph to plot student performance. Sharing the graph with his students, he sets a line to demonstrate a target (for instance, 80%) and then another line 10% higher to show an area of 'plausible stretch'. For every piece of work submitted and assessed (tests, essays and any other assessed pieces), the student's mark is plotted on their graph and, if the point is below the target line, the student is asked to review the work, stating where they underperformed then instructed to resubmit the work, sometimes with annotations demonstrating what was done to make the improvements. By doing this, the teacher says his students appreciate that he will not allow an individual piece of work to remain as an underachievement, impressing on his students that improvement is always possible.

These strategies are adapted with permission from a package of materials available in: Marsh, H. W., and Roche, L. A. (1994). The Use of Students' Evaluations of University Teaching to Improve Teaching Effectiveness. Canberra: Department of Employment, Education and Training. Further information on the Improving Academic Teaching Project can be found in Marsh, H. W., and Roche, L. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. American Educational Research Journal, 30, 217-251.

Many of the ideas presented in the original package of materials were adapted with permission from Davis, B.G., Wood, L., & Wilson, R. (1983). ABC's of Teaching with Excellence. Teaching Innovation and Evaluation Services, University of California.

The authors would also like to express their appreciation to the many teachers who contributed strategies and other suggestions during the project.

Teaching strategy ideas for Breadth of Coverage

This aspect of teaching reflects students' responses to items concerning the contrasting of implications of various theories, the provision of the backgrounds of ideas and concepts, the presentation of different points of view and the discussion of current developments. These all have to do with substantive qualities of teaching. Each would seem to have the potential to increase student knowledge and understanding through facilitating generalisation beyond the confines of the specific situation, to clarify the material to be learned and its meaningfulness to the learner.

The following ideas are suggested and used by outstanding teachers across a range of institutions and disciplines. We suggest that you select no more than 3 or 4 strategies

which appear potentially the most profitable, then make a commitment to apply OR adapt them to improve your teaching effectiveness.

1. Select an article which represents one theoretical viewpoint and build your lesson(s) around an opposing idea or position.

A teacher of Economics, for example, assigns a reading that represents the point of view of liberal economists, but designs his teaching sessions around the opposing views of leading conservatives or radicals.

In addition to assuring a balanced presentation, this approach adds variety and interest to the course and stimulates students to think critically. Because the taught material complements rather than repeats the article, it has the added benefit of increasing student attention.

2. Contact university departments or other experts in the field to get the latest information.

Before giving her lesson on the heart, a Biology teacher contacts their local university to get the most recent statistics and findings on heart transplants. Similarly, a Law teacher directly contacts lawyers involved in important cases pending or under adjudication, making her lessons as up-to-date as possible.

A Geography teacher routinely calls his contacts in a local university to get the latest information on environmental legislation which he incorporates into his lessons.

This kind of up-to-the-minute reporting on a few major developments relevant to course content can help you convey a sense of the excitement of research to your students.

3. Invite guest speakers to offer contrasting viewpoints.

One teacher makes a point of doing this in his courses so that his students are exposed to a variety of positions. "I want them to understand what the different points of view are," he says, "and one of the best ways I have found to do this is to invite a colleague or practitioner whom I know to be an adherent of each view to make a presentation to the class."

"I always take detailed notes during a guest lecture," says a Biology teacher. "In this way I am able to answer student questions about the material during later sessions and may learn something new myself!"

4. Assign multiple readings to represent a variety of viewpoints.

"Because the most controversial issues covered in a course are ones on which my students have strong opinions but little information, I try to expose them to diametrically opposite positions or theories," says one Politics teacher.

"Developing a set of readings takes time. However, I can usually use it for two or three years with only minor modifications.

5. Assign readings directed toward revealing the reasons behind differing points of view.

A Business teacher says, "I lead my students through an analysis and critique of each theorist's position." The focus is not on opinions but on the reasons behind them. "Sometimes my own view is apparent either explicitly or implicitly; other times it is not. Sometimes it is not possible to find a reading which gets at the basis for a particular point of view. However, any reading that presents a clear statement of the features of the theory is useful. Students can be directed to a lively discussion of reasons that are tenable. It gives them experience in learning the criteria of a good argument."

6. Use your students' opinions to create a microcosm of society's attitudes on social, political, and economic issues.

At the beginning of the term, a teacher of Economics gives his students a questionnaire in which they are asked to agree or disagree with a series of controversial statements on the functioning of the economy. "I use the survey results to introduce a variety of student viewpoints. Throughout the term I reveal selected results from the survey as these relate to new concepts or issues covered in homework and lessons."

"This technique gives my students a sense of personal involvement in the subject matter. Students learn that some of their peers may share their viewpoint. They also learn that some of their peers don't share their viewpoint and why. Use of student data allows me to introduce most of the views currently reflected in the society as a whole."

7. Draw upon the diverse backgrounds and experiences of your students to introduce different points of view.

At the beginning of the term, a teacher of Business asks his students to give written answers to questions about their backgrounds and reasons for taking the course. He

asks students to focus particularly on experiences which might give them a particular viewpoint on social, political, and economic issues to be covered in the course.

Using a seating chart he calls on students whose prior experiences or interests may be relevant to a topic under discussion. In this way a full range of views is introduced in the course. "Often, with little or no effort, I am able to get students debating between themselves. In fact, I rarely give my own point of view until there has been a full discussion of the different points of view within the class itself."

This technique has additional advantages: introducing personal experiences and opinions makes the class livelier; and the teacher is given a method for learning about the students' backgrounds, experiences and personalities.

8. Present each of several competing theories as if you were an adherent of that position.

A Psychology teacher introduces three major approaches or schools of thought in the field. "I discuss each one historically and contrast the basic elements and implications of each," he says.

"I really don't have a point of view when I teach the course. There is so little known with impartial certainty; I don't think one is justified in taking a position. Therefore, I present the best case for each theory, then analyze each critically and comparatively."

Several other excellent teachers report that they also present the best case for each of several competing theories when teaching a course.

9 .Encourage students to take an approach different from the one you have adopted.

A teacher of English uses this strategy in all of his literature courses. "I always approach literature from an historical point of view: history is a particular passion with me," he says. "At the same time, I point out that there are many other perspectives and encourage students to use alternative approaches, e.g., the feminist approach."

10. Point out explicitly that there are alternative points of view.

"I indicate the polar principles which guide much of the research in the social sciences as well as much of our folk wisdom, e.g., 'opposites attract' versus 'birds of a feather flock together' or 'absence makes the heart grow fonder' versus 'familiarity breeds contempt'" says one teacher. In doing so I point out that they should be mindful that

there may be good reasons to believe the opposite of what a theory presents; that they should analyze all arguments in terms of their opposites."

11. Touch base repeatedly with the fundamentals or basics.

One Science teacher believes that too much of science is presented to students in a rote, plug-in-the numbers way.

"There are thousands of formulae," he points out, "but all of these are variations on a limited number of basic ideas of theories." "These basic ideas are 'ideal theories' from which are derived all the 'approximate' or 'technical theories' which scientists use."

"I try to teach my students how to judge when they can use an approximate theory with confidence and when they are obliged to go to a more rigorous level. In this way, I keep touching base with the fundamentals to reinforce students' understanding of them."

12. Focus your course on the classic issues and concepts in your discipline.

A History teacher explains that she has moved away from presenting the most esoteric and up-to- date concerns of professional historians to her students.

"The most interesting issues and themes for undergraduates," she explains, "generally turn out to be those which originally excited historians about a particular person, event, or epoch, not the controversies of present-day historians. The classic issue are the ones which attracted me to the field," she says, "and I find that they are still the most exciting for my students."

Following this approach does not mean that you cannot introduce newer research findings where they are relevant, of course. Nor does this suggest that ideas which have little or no current validity should be taught. It does mean that, in limiting your coverage, you select the major classic themes and concepts wherever possible.

13. Share your professional "junk mail" with your students.

One teacher makes a point of passing around announcements for local conferences, program proceedings, and advertisements for new books and journals in the field. "In this way I inform my students about professional activities and recent developments of which they might not otherwise be aware," he says.

"I also encourage my students to attend meetings and conferences and to read articles on topics of interest to them."

A staff member teaching French also shares copies of newsletters, newspaper clippings, and announcements of French movies, plays, or other cultural events in the area. "My students are often amazed and delighted to learn that there are so many opportunities to strengthen their language skills and to expand their understanding and enjoyment of French culture," he explains.

14. Require your students to read current journal articles.

"It's important for my students to be exposed to state-of-the-art ideas," says one Politics teacher. "I try to make sure that I direct students towards at least a few recent journal articles. [Sometimes] I translate the abstract of a journal article into layman terms. I present the basic findings in a narrative fashion, using little actual data."

"I want my students to become excited by the open-ended nature of our subject. I want them to understand that what they are learning is not the final word."

15. Require your students to read current newspapers or periodicals.

A teacher of Economics assigns the Tuesday editorials of the Wall Street Journal each week. She uses them as a basis for discussion, and she has her students compare them with textbook presentations on related topics.

A teacher of Biology also believes strongly in making use of articles in current periodicals. "I keep my eyes open for stories on recent developments which have become part of the 'current events' literature," he says. "For example, in a discussion of recombinant DNA, I was able to use photos from a recent issue of Life Magazine, as well as a story a broadsheet did on the Genetech Corporation."

16. Tell your students about local events which will expand their understanding of your subject.

"I distribute a calendar announcing course-related events in the area," one staff member explains. "The events include plays, lectures, demonstrations, poetry readings and so forth. In this way the content of my course is expanded far beyond what I can actually cover in class. I also encourage my students to use these local resources in their research and writing assignments."

17. Assign provocative or controversial topics for essays.

"I find that the quality of the essays I get often depends on the quality of the task / title I give," says a teacher of Business. He tries to give provocative tasks. For example, in a recent assignment he asked his students to respond to the question, "If you were working in a company that illegally pollutes the environment what would you do and why?" Giving provocative assignments not only challenges his students and makes for more interesting reading but also diminishes the chance that the essays will be plagiarised.

The importance of getting to know what gets the class "fired up" and what they relate to, is emphasised by several teachers in setting appropriately provocative assignments.

Teachers also noted that such assignments are an excellent way to encourage students to consider a broader range of issues in their responses.

18. Assign "thought problems" which are typical of the problems faced by professionals in the field.

One teacher assigns weekly "thought problems" which are of the same type of questions professional in the field are asked, such as, 'What caused the erosion in this picture?'; not 'Name six factors which can cause erosion."

Using real-life problems to encourage thoughtful reflection and/or discussion in this way, rather than requiring solutions, can be a particularly useful way to avoid overwhelming students with the complexities inherent in some tasks.

19. Give assignments which put your students in the role of another.

A History teacher reports that she used to give rather standard writing assignments, e.g., "compare author X and Y's views on A," where the two authors tended to be professional historians. "Most students, however, find the arguments of current historians somewhat arcane," she says.

"Therefore, most recently I have asked my students to read a collection of the 18th century speeches on why Louis XVI should be killed and assigned them the task of writing their own speech as if they had been living during the French Revolution.

"Students really are enthusiastic about this kind of assignment and do an incredibly good job. It helps them to identify with the issues of the time; in fact many of my students went to great lengths to research the authenticity of their own empathic interpretations. Next year, I intend to take this assignment a step further by dividing my

students into small groups and having them actually deliver their speeches to the group."

These strategies are adapted with permission from a package of materials available in: Marsh, H. W., and Roche, L. A. (1994). The Use of Students' Evaluations of University Teaching to Improve Teaching Effectiveness. Canberra: Department of Employment, Education and Training. Further information on the Improving Academic Teaching Project can be found in Marsh, H. W., and Roche, L. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. American Educational Research Journal, 30, 217-251.

Many of the ideas presented in the original package of materials were adapted with permission from Davis, B.G., Wood, L., & Wilson, R. (1983). ABC's of Teaching with Excellence. Teaching Innovation and Evaluation Services, University of California.

The authors would also like to express their appreciation to the many teachers who contributed strategies and other suggestions during the project.

Teaching strategy ideas for Homework

Positive ratings in the Homework section probably indicate that activity in learning outside of class was found to be valuable and that the learning experiences involved were meaningful. Homework provides students with opportunities to practise new knowledge and skills and, therefore, the Homework aspect of teaching seems consistent with sound principles of learning.

The following ideas are suggested and used by outstanding teachers across a range of institutions and disciplines. We suggest that you select no more than 3 or 4 strategies which appear potentially the most profitable, then make a commitment to apply OR adapt them to improve your teaching effectiveness.

1. Give your students at least one assignment which consists of several options.

One teacher of English requires students to write two essays on assigned topics. His third assignment, however, gives five or six options from which his students may choose the one which sounds interesting and most allows them to do their best.

Examples of the options which he offers include: a piece of creative writing; a dramatic representation to be performed in front of the class (which can be a small group or team project); an original videotape to be shown to the class (which can also be a team effort); or a third essay. In addition, with his permission, students can create an option of their own if they wish.

"More than five or six options tends to confuse some students; it becomes too difficult to decide," he believes. "Two few options, on the other hand, restricts unduly my more creative and daring students." Although optional assignments must be related to the subject matter of the course, he encourages his students to take an interdisciplinary approach and to link content and skills from other courses.

2. Plan to return homework in the next lesson.

"When I schedule homework assignments, I block out my own time to mark them immediately following the lesson the students hand them in," one Science teacher says. "This is important for two reasons. First, the quick turn-around time ensures that my students are still thinking about the assignment. Thus any feedback is likely to have a stronger impact than if it were delayed a week or more. Second, prompt feedback indicates to my students the importance of what they are doing and my concern for their learning the material."

An English teacher agrees. "The impact is enormous when you return assignments in the next lesson. Students are still anxious to know how they have done. That's a tremendous advantage in maximizing the impact of feedback on their learning."

3. Set up student panels.

One humanities teacher organises a series of student-led discussions each term. "I believe my students can teach themselves a great deal; therefore I do not play an active role in the student-led discussions. My role is to serve as organiser and facilitator.

In the first week his students select the topic and the date of their presentation. Generally, there are three to four students per topic. Outside class, his students meet

as groups him to discuss how to organise their topics for presentation and discussion. It is up to each student group to select whatever format they wish for their presentation.

"In the past, student groups have conducted a debate, performed a skit, or simply led a discussion about the topic," he says. "They learn a lot about the topic and they really get to know one another while preparing their presentations.

4. Prepare answers to exams and quizzes to hand out as soon as your students turn in their work.

One Chemistry teacher prepares a handout of correct answers which he gives to students as they turn in their answer sheets and leave the room. "There is no point in making students wait several days or weeks to find out how they did," this teacher explains. "They are most interested in the results at the time of the examination, and it is at the time of the examination that the greatest reinforcement of the learning can take place.

Note that this method gives students immediate feedback even though it may be a week or more before their work can be returned with comments or grades.

5. Give students frequent assignments and make extensive, constructive comments on them.

"Students need to know what they are doing well in addition to what they need to improve," says one teacher of History. "I am always careful to praise their strengths and to be as constructive and helpful as possible in pointing out their weaknesses."

"I make a point of writing extensively on my students' papers," says one teacher. "I make comments in the margins as I am reading and then append lists of strengths and suggestions for improvement."

6. Discuss the answers to exams, quizzes, or homework assignments at the next class meeting.

One Science teacher says that even if he cannot return graded assignments or exams, he always discusses the answers in the next lesson. "I want to correct any misunderstandings and reinforce their learning as soon as possible," he says. "Students are much more receptive to this right after completing an assignment."

7. Create opportunities for role playing.

A Law teacher makes use of role playing to encourage his students to develop skills they will need in their careers. "I give my students copies of a legal report, for example. Then one half of the class is asked to assume the role of the authors of that report and prepare an oral presentation for the client. The other half of the class is assigned to act as representatives of the client and to prepare questions to be asked of the lawyers.

"About a week later, during class time, I select certain students to actually enact these roles in front of the class. My students do not know ahead of time who will be called upon, so everyone has to be prepared. Those not called on join me in the role of the observer. When the students have enacted the meeting, the rest of us review each side's performance."

8. Assign provocative or controversial topics for papers.

"I find that the quality of the essays I get often depends on the quality of the assignment I give," says a Business teacher. He tries to give provocative titles for essays. For example, in a recent assignment he asked his students to respond to the question, "If you were working in a company that illegally pollutes the environment what would you do and why?" Giving provocative assignments challenges his students and makes for more interesting work.

One teacher who successfully engages students this way warns that it is important, even when deliberately trying to be provocative, or "realistic", to choose topics that the students are "ready" to deal with in the context of the material being covered.

The importance of getting to know what gets the class "fired up" and what they relate to, is emphasised by several lecturers in setting appropriately provocative assignments.

9. Ask students to analyse an essay or journal article and to write a critique of it.

One teacher of English assigns the work of a literary critic and then asks his students to write an essay taking an adversary position. "If my assignments are provocative," he says, "I get better results. I stress the importance of their presenting a personal point of view. They should enjoy doing the paper; it should provide them with a personal learning experience."

10. Give assignments which put your students in the role of another.

A History teacher reports that she used to give rather standard, sometimes dull, writing assignments.

"Most recently, however, I have asked my students to read a collection of the 18th century speeches on why Louis XVI should be killed and assigned them the task of writing their own speech as if they had been living during the French Revolution.

"Many students really are enthusiastic about this kind of assignment and do an incredibly good job. It helps them to identify with the issues of the time; in fact many of my students went to great lengths to research the authenticity of their own empathic interpretations."

11. Assign "thought problems" which are typical of the problems faced by professionals in the field.

A Biology teacher assigns weekly "thought problems" which are of the same type of questions professionals in the field are asked, such as, 'What is killing that tree?'; not 'Name six factors which can kill trees."

12. Have students solve problems at the board.

A staff member who teaches statistics calls on students to come up to the board to solve problems. Each student is permitted to bring a fellow student as a "coach" so that he or she is not put on the spot. At the beginning of the term the problems are based on homework assignments. Toward the end of the term, they are based on impromptu examples. This method increases student discussion and interaction and encourages your students to pay close attention in class.

13. Have your students carry out independent research projects.

A Chemistry teacher who uses this approach believes that too many courses follow a "recipe" approach and thus do not really introduce students to science as well as they might. "I want my students to get a feel for real scientific research," he explains. "Therefore, I require them to develop the questions, select the methods by which they are going to carry out their investigations, review the relevant research literature, and report their findings in both written and oral form." Some teachers frame this as a WRAP (Wonder, Research, Analysis, Publication) project.

These strategies are adapted with permission from a package of materials available in: Marsh, H. W., and Roche, L. A. (1994). The Use of Students' Evaluations of University Teaching to Improve Teaching Effectiveness. Canberra: Department of Employment, Education and Training. Further information on the Improving Academic Teaching Project can be found in Marsh, H. W., and Roche, L. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. American Educational Research Journal, 30, 217-251.

Many of the ideas presented in the original package of materials were adapted with permission from Davis, B.G., Wood, L., & Wilson, R. (1983). ABC's of Teaching with Excellence. Teaching Innovation and Evaluation Services, University of California.

The authors would also like to express their appreciation to the many teachers who contributed strategies and other suggestions during the project.

Teaching strategy ideas for Workload & Difficulty

Work that is seen by students to be too much or too difficult, is often by definition, given without consideration of learners' capacities and prior learning. Moreover, such work cannot be appropriately paced or presented in desirably learnable units. Overloaded students find it difficult to experience subjective feelings of success. They are likely to be forced into adopting learning strategies that minimise their ability to understand and generalise from the specific learning situation. On the other hand, students for whom success is too easily won lose motivation to succeed and are unlikely to value such learning highly. It would seem that students would value more highly achievement that involved them in overcoming substantial obstacles and that have necessitated relatively enduring commitments. This suggests the possibility of nonlinear relations between Workload/Difficulty and other indicators of teaching effectiveness such as student

learning. There can be little or no doubt that appropriate consideration of the Workload factor is consistent with accepted principles of teaching and learning.

The following ideas are suggested and used by outstanding teachers across a range of institutions and disciplines. We suggest that you select no more than 3 or 4 strategies which appear potentially the most profitable, then make a commitment to apply OR adapt them to improve your teaching effectiveness.

1. Empathise with the students' difficulties in learning the material for the first time.

"It is important to distinguish between appreciating the difficulty students have in understanding new material, and the rather simpler but less effective option of allowing the subject difficulty to act as an excuse for the lecturer's quality of teaching or the students' quality of learning," according to an outstanding teacher.
A staff member in the sciences says that he noticed that he had taught the course better the first time than he did in subsequent years. "When I asked myself why, I realised that in preparing the course for the first time, I really had to work hard to master certain parts of the material in order to explain it to my students. The next time, however, the mastery and explanation of these concepts no longer seemed difficult to me. Unfortunately, I forgot that they would still be difficult for the students. Now I colour-code all of my lesson plans, keying the parts that students are likely to find difficult and making a special effort to make points very clear."
A Physics teacher also tries to put himself in the students' shoes. "After I have finished writing up a set of lesson plans," he says, "I review them carefully, asking myself: 'What might my students find hard to follow in that line of reasoning?' 'What examples might make that more clear?' This has now become the most important part of my lesson preparation."
Several staff members report making notes to themselves of explanations that worked well and those that didn't. They also keep records of the kinds of errors students most commonly make in assignments and exams as a reminder of what students find most difficult to understand.

2. Get to know your students: Where they are "at"; and what they relate to.

Knowing your students is important for a number of reasons. Several outstanding teachers stressed that new learning must begin from what students are already familiar with. "Otherwise they quickly become confused, disinterested or anxious," one teacher explains. "Students will also open up more in class discussion if they feel a comfortable rapport with the teacher".
Getting to know what gets the class "fired up", or what they relate to is a strategy that several teachers have recommended as a means of generating a vibrant learning

atmosphere. This requires establishing good rapport and making time to chat with students in non-teaching situations.

3. Stress the most enduring values or truths in your discipline.

"I stress the permanent values in literature, the emotional responses that a particular novel or collection of novels elicits from us all," says one teacher of English. "I try to get my students to understand why they respond to a given novel in the way they do."
After a class has discussed how they feel about a novel - the common emotions it arouses - he tries to lead them to analyse, understand, and explain why nearly everyone feels the way they do. He poses questions such as: What must literature be like in order to get us to respond the way we do?
Why does a particular novel affect everyone in the same way? "Behind all my questions is the search for a way of analysing and discussing literature that will explain the most with the fewest assumptions."

4. Acknowledge the difficulty of concepts students are likely to find hard to understand.

"Acknowledging difficulty avoids the risk of belittling the students' efforts in mastering the concept, or the students themselves if they do not master the material easily," according to an exceptional Chemistry teacher. "It is important to admit to the difficulty of understanding material for the first time, but not to make that difficulty an excuse. A good way of achieving this aim is to offer a specific 'strategy' for mastering the material, such as '...so listen carefully...', or
'...so remember this simple example'".
One teacher says, "I consciously cue students to the most difficult ideas by saying such things as, "Almost everyone has difficulty with this one, so listen closely.' Because the level of students' attention varies throughout the hour, it is important to get everyone listening carefully before introducing a new concept or explaining a difficult point".

5. Touch base repeatedly with the fundamentals or basics.

"Students like to be challenged, but they need to feel confident and well-prepared to meet the new challenges, and that often means consciously retreating a little to reinforce the foundations on which new material is based," according to an outstanding teacher.

6. Focus your course on the classic issues and concepts in your discipline.

A History teacher explains that she has moved away from presenting the most esoteric and up-to-date concerns of professional historians in some of her teaching.
"The most interesting issues and themes for older students," she explains, "generally turn out to be those which originally excited historians about a particular person, event, or epoch, not the historiographical controversies of present-day historians. The classic issues are the ones which attracted me to the field," she says, "and I find that they are still the most exciting for my students."
Following this approach does not mean that you cannot introduce new research findings where they are relevant, of course. Nor does this suggest that ideas which have little or no current validity should be taught. It does mean that, in limiting your coverage, you select the major classic themes and concepts wherever possible.

7. Explicitly call attention to the most important ideas in each lecture.

"I began to emphasise the main points about ten years ago," says one Politics teacher, "when I discovered that you can't rely on students to intuitively know what the most important points are. You have to tell them."
Staff members in several disciplines stress the need to call students' attention to the most important ideas being presented. Some teachers announce the importance of an idea before presenting it, saying such things as "This is really important, so you have to be alert." Other teachers emphasise the most important ideas when summarizing, saying "The most important thing to remember here is..." or "This is so important that everyone of you should have it engraved on a gold plaque and hung over your bed!" as one ICT teacher puts it. ”There is no point in my students having to guess what is important if I can tell them," he says.

These strategies are adapted with permission from a package of materials available in: Marsh, H. W., and Roche, L. A. (1994). The Use of Students' Evaluations of University Teaching to Improve Teaching Effectiveness. Canberra: Department of Employment, Education and Training. Further information on the Improving Academic Teaching Project can be found in Marsh, H. W., and Roche, L. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. American Educational Research Journal, 30, 217-251.
Many of the ideas presented in the original package of materials were adapted with permission from Davis, B.G., Wood, L., & Wilson, R. (1983). ABC's of Teaching with Excellence. Teaching Innovation and Evaluation Services, University of California.
The authors would also like to express their appreciation to the many teachers who contributed strategies and other suggestions during the project.

Appendix 6b: email sent to all intervention arm teachers, November 2012

Good morning,

I hope this finds you well.

Thank you for volunteering to be part of the SecondarySEEQ student evaluation of teaching project, based here at CEM. Your enthusiasm is very much welcomed, as is your patience. This project is evolving every day and, as such, sometimes throws up issues that need resolving and may take a little time to do so. Irrespective of this, we will keep you informed so that, at the very least, you know what is going on.

By being involved in the project, you are very much at the forefront of research in an area that is generating a lot of interest. Your participation is at the heart of evidence-based practice in teaching.

With the first data collection point for the SecondarySEEQ Project coming up in two weeks, I'd like to take the opportunity to highlight some important pieces of information.

The student survey

1. The classes you have registered to take part in the project will need to take the online student survey during the week of 19th November. This should take place during lesson time; on average, students take around 6.5 minutes to complete the survey (the longest anyone has taken is 15 minutes). If you have difficulties with the survey, please seek help from your IT staff in the first instance. If your IT staff are unable to resolve the problem, please email the SecondarySEEQ Project Team on cemseeq@cem.dur.ac.uk.

2. I will send a link to the student survey page at the beginning of next week (12th November), by email.

3. If your students find that something on your student survey is incorrect (a class name is missing or a surname is not spelled correctly, for instance), please let me know as soon as possible.

4. It will be useful to have an activity planned for those who complete the survey quickly to move onto. This should, ideally, be something that doesn't involve noise and which students can begin easily; writing it on the board may be useful.

5. When you are ready for your students to take the survey, ask them to log on to the internet. Before they begin the survey, please read the following paragraph to them: 'You're going to take a survey that has been sent to us by a researcher at Durham University. The survey will ask you about my teaching and it shouldn't take very long to complete. I've volunteered to be part of the researcher's project and am interested to find out what you think of my teaching. As such, I want you to be honest when you give

your responses. Even though you will be asked to put your name on the survey, I will not see this when I get the results back; I will only see class averages for our class and not any individual responses you give. Please don't talk during the survey - allow each other quiet time to think carefully. When you finish, please follow the instructions I've put on the board. Once everyone has finished, we will continue with our lesson as normal. Please now type this address into the browser and then follow the instructions on the survey website.'

Your survey

6. I would be grateful if you would complete a short teacher survey; this can be found here: LINK TO CVENT TEACHER SURVEY. The survey shouldn't take very long to complete and will help us understand what effect the process of having your students evaluate your teaching has on your teaching.

Your secure website

7. You will be sent login details for the secure website hosted here at CEM on which you will be able to access your class-level student evaluation of teaching data. Usernames and passwords will be sent by post (for security reasons) to you at school, so please get in touch with us via cemseeq@cem.dur.ac.uk if you haven't received yours by 3rd December 2012.

Your peer-coached meeting

8. As your school is in the intervention group, you will need to ensure that you approach a colleague in school to take on the role of your peer-coach. The ideal person is a colleague who you trust and for whom you have respect as a teacher. They may teach the same subject as you, or you may approach someone who teaches a different subject. Your peer-coach might be involved in the SecondarySEEQ Project, or they may not. Your peer-coach is not there to be an 'expert', but someone who is happy to listen with genuine interest and who adopts a non-judgmental stance. Their role is very much to help you put the data into your class and school context, aiding you in looking forward to ways in which you might improve subsequent evaluation ratings.

9. You and your peer-coach will need to arrange a meeting of around an hour in length to take place at a convenient time between 3rd December (when we hope to have all data processed and returned to you via the secure website) and the end of your school term. You will need to access the secure website during this meeting so an internet connection will be needed. It will be important to make sure you are not disturbed during this meeting (I know this can be difficult in school!); allowing yourself and your peer-coach to focus and think critically and creatively will be important. Consider booking a room if this is possible. If unforeseen circumstances prevent you holding the meeting, please rearrange it.

10. Your peer coach will not need to prepare anything in advance of the meeting; it is advisable, however, that you log on to the website prior to meeting your peer-coach so

that you can view your own data and begin thinking about what it means and what uses you might make of it. Also, you will find a copy of the survey to which your students responded, along with a brief guide to start you thinking about interpretations of the data.

11. Because there are nine aspects of teaching the survey is trying to tap into, it is very likely that there will be peaks and troughs on the graph you see. Don't be surprised by this. Each point on the graph represents the average response for all students in your class (we haven't included the spread of results) and particularly where you have small numbers in your group, a single extreme response can shift the average for a specific aspect of teaching in one direction. So when you think about interpretations of the data, keep in mind this in mind.

12. On the website, you will find two types of resource to guide your peer-coached meeting: a video and a set of PDF documents with teaching strategies. We've created the video so that you can play and pause; it gives you questions to answer and things to discuss with your peer-coach. All of the resources are there to give you a direction, but not to give you all the answers. Our 'tight but loose' model invites you to use the resources as they are, adapt the strategies they suggest, or come up with your own, based on what you believe will have a positive effect on your class.

13. The outcome of your peer-coached meeting is twofold: to offer some interpretations of the data and to decide firmly upon what you will do in the following term to try and raise your class's subsequent student evaluations of your teaching. On the Feedback page of the secure website, you will find a box to type in the specific actions you will take in the following term with a specific class.

14. Reflections on being part of the SecondarySEEQ Project are really important to us. There is a tab on your secure website called 'Reflections'; it's here that we would like you to write about your experience of having your students evaluate your teaching. Ideally, this should be something you update periodically, whenever you feel there is something of interest to say. The more you can tell us about what it's like to be a part of the project, the more we can learn from you.

Thank you for reading through this. It might be useful to save it or print a copy for future reference. If you have problems accessing the website or have another query, please email cemseeq@cem.dur.ac.uk.

Best wishes,

Stuart Kime

PhD student and SecondarySEEQ Project Leader

Appendix 6c: SecondarySEEQ user terms and conditions of use

SecondarySEEQ Terms of Use

The services and content on the website https://css.cemcentre.org/seeq are owned, maintained and provided to you by the University of Durham ("us", "we", or "the University") whose registered address is University Office, Old Elvet, Durham, DH1 3HP, United Kingdom as represented by the Centre for Evaluation & Monitoring (CEM).

By proceeding with access to https://css.cemcentre.org/seeq (hereinafter the "Website"), including accessing or utilising any of the services or content therein you hereby agree to be legally bound by the following terms and conditions (the "Terms"). If you do not accept the Terms, please do not use or access the Website. Your use of the Website is at all times governed by the Terms as set out below, which should be read in conjunction with:

A. any details provided within the site in regards to its operation; and
B. any terms relating specifically to the services or content available on the Website as detailed therein.

1) General

a) We reserve the right to update or amend these Terms at any time, and therefore in continuing use of the Website and any content or service therein following any such change, you shall be deemed to be granting your acceptance of such change. It is therefore your responsibility at all times to check the Terms regularly to ensure you are aware of any such change. We reserve the right at all times to terminate your access to any or all of the services provided on the Website at any time, without notice, for any reason, including without limitation, breach of these Terms.

b) The Website and its contents (including without limitation the Website design, text, graphics and all software and source codes connected with the Website and its services) are protected by copyright, trade marks, patents and other intellectual property rights and laws. In accessing the Website you agree that you will access the contents solely for your personal, non-commercial use, or in accordance with any stated Commercial Commons Licence or other written

permission. Except for your downloading, copying and/or printing of pages of the Website for your personal use, in the absence of any applicable Commercial Commons Licence or other written permission granted in advance, none of the content may be downloaded, copied, reproduced, transmitted, stored, sold or distributed.

c) The content of the Website is intended for information purposes only and should not be considered to constitute any form of advice or recommendation upon which specific decisions should be made. The University has taken steps to ensure the materials contained on its website are current and accurate, but to the fullest extent permitted by law the University hereby excludes any warranty, whether express or implied, by statute at common law or otherwise, relating to the website and the material contained on the website. In particular (but without limitation of the above), the University gives no warranties as to quality, accuracy, timeliness, completeness or fitness for a particular purpose of the website or of the material contained in it. The University shall not be responsible for any use you may make of any information contained on the Website, how such information may be interpreted or any reliance that may be made on the same. The University provides no guarantee that the information contained herein is fit for any particular purpose.

d) To the fullest extent permitted by law, the University shall not be liable for any claims, costs, penalties, loss (whether direct, indirect or consequential and whether economic or special loss of any nature), damages or expenses arising from the use or from the inability to use the website. The University does not accept liability from reliance on information or advice contained on the website or from any unauthorised access or alteration to the website by a third party.

e) Any information you may submit to the Website, other than personal data that is covered by the University's Privacy Policy, will be treated as non-confidential and the University shall be entitled to reproduce, distribute and communicate that information to others without restriction and without obligation. By submitting information to the Website, you warrant that such information:
   i) is your own original work and that you have the right to make it available the University;
   ii) is not defamatory; and
   iii) does not infringe any law.

f) By submitting any information to the Website, you hereby agree to indemnify the University against all liabilities, costs and damages (including legal expenses) that may be incurred by the University as a consequence of any breach of the warranties expressed above and you agree to waive any moral rights to your contribution.

g) The University, its representatives, employees, officers and agents shall not be liable for any loss or damages or expenses of any kind including without limitation compensatory, direct, indirect or consequential damages, loss of data,

income or profit, loss of or damage to property or claims by third parties howsoever arising (save for death or injury caused by the University's negligence or that of its representatives, employees, officers and agents) in connection with the copying or use of any information or material contained in or referred to on the Website or otherwise from the use of the Website.

h) The University is not responsible for the content of external websites that link to the Website or which are linked from the Website.

i) The University reserves the right to withdraw the Website at any time and without prior warning. The University will endeavour to advise users of the Website in advance of any such withdrawal, whether such withdrawal, but is under no obligation to do so.

j) Parents who wish to allow their children access to and use of the Website and any services therein should supervise such access and use. It is your responsibility to determine which content or services are appropriate for your child.

k) We reserve the right at our discretion to prohibit any link from another Internet site or equivalent entity to materials or information on the Website. Without prejudice to that, any link to material or information on the Website must be neither misleading nor deceptive.

l) Durham University, University of Durham, the University of Durham logo and CEM are either trade marks of the University or Trade Marks of a third party to which the University holds a licence and therefore the same may not be reproduced without our prior written consent. Other organisation and company names referred to on this Website are the trade marks of their respective owners as indicated.

2) Data Protection and Privacy Policy

a) Where you provide us with any personal information from which we can identify you as a living individual (whether by taking that information by itself or when using it in conjunction with other information which we hold or are likely to hold) we shall process such information only in accordance with the provisions of the Data Protection Act 1998.

b) We will not collect any personal information about you, except where it is specifically and knowingly provided by you, for example when you register as a user of the Website or request information or provide us with feedback on the Website. We do not sell or share any information about you as an individual user of this website unless you agree to us doing so.

c) Any personal data you submit to the Website will be used according to these principles:
   i) Your data will be stored securely on the server hosting this website;
   ii) We will never sell or transmit your data to any third party, unless required to do so by law;
   iii) We won't collect personal information about you without your knowledge.

3) General

a) Nothing in the Terms is intended to confer or purports to confer on any third party any right to enforce any provision of these Terms.

b) Should any provision or provisions of the Terms be determined to be invalid, illegal or unenforceable for any reason by any court of competent jurisdiction then such term shall be severed and the remaining terms shall survive and continue in full force and effect and remain binding and enforceable.

c) These Terms shall be governed by and construed in accordance with the law of England and you hereby submit to the exclusive jurisdiction of the English courts. Where you use the Website outside of the United Kingdom you agree to comply with all applicable laws regarding the transmission of technical data exported from the United Kingdom or the country in which you reside (if different) and with all local laws and rules regarding acceptable use of and conduct on the Internet.

Appendix 6d Peer-coached meeting guide

CEM Centre for Evaluation & Monitoring

Durham University

SecondarySEEQ termly peer-coached meeting guide

SecondarySEEQ PhD Project 1

CEM Centre for Evaluation & Monitoring

Introduction

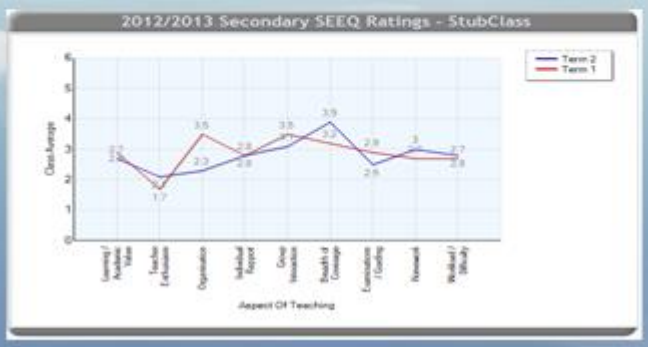
Durham University

- Set aside one hour for this meeting

CEM Centre for Evaluation & Monitoring

Class evaluation data (sample)

Durham University



In view

- Keep the relevant class's graph in view
- Open the relevant teaching strategy ideas PDF (linked on the Feedback page)

Why are we here?

Specific
Measurable
Attainable
Relevant
Time-bound

Peer-coach role

Attentive
Accepting
Warm
Genuine
Non-judgmental
Disciplined
Possessing a sense of humour!
Empathic
Curious

Discussion: step 1

- *Note that you'll need to repeat each of these steps for each class from which you've received evaluation data*
- What do you think the data presented on the graph mean?
- Why do you think your students have given these evaluations of your teaching?

Discussion: step 2

- *Consider two aspects of teaching for which you received the lowest evaluations*
- If, when we meet towards the end of next term, you see improvements in these aspects of teaching, what helpful things might you or others have done?
- What would have been the first thing that happened?

Discussion: step 3

- *Something amazing happens*
- *Positive changes happen without your knowledge*
- When you wake up, what would be the first, specific indication that your positive changes had been made and had worked?

Discussion: step 4

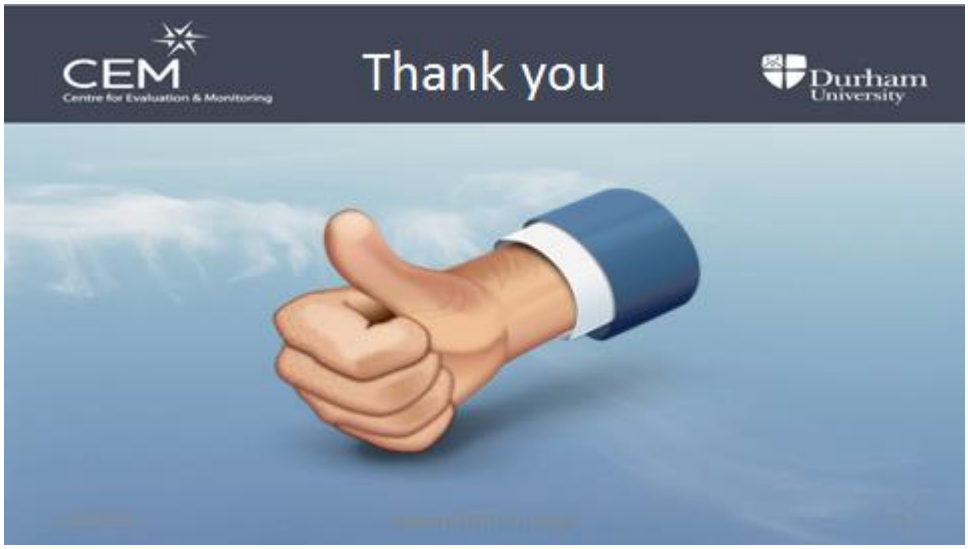
- If these amazing positive things had happened, how would the students in your class know?
- How do you think your students would react?
- What else would be happening that would indicate the amazing positive changes had taken place?

Goals and strategies

- Decide on and record the specific goals you want to achieve in your teaching with a specific class
- Select and record the specific strategies you will use to try and achieve improvements in next term's evaluations

Reflection

- Use the Reflection page to record your thoughts and observations on your experience in the SecondarySEEQ Project



Appendix 7a: teacher mission texts

- Aims: to reduce the difficulty for weaker students and increase challenge for stronger students. Strategies: Find out where students are 'at' through 1:1 tutorials. This will build more rapport with weaker students and encourage them to ask more questions during class time to resolve any misconceptions. Also arrange work groups for class work to give a variety of opinions during small group work & presentations. Require the reading of articles from Philip Allan Chemistry Review (made available through the learning platform) on a variety of topics to broaden the range of chemical contexts to some real-world situations. This will aim to stimulate broader questions.

- The area I am going to focus on improving over the course of next term are 'selling' my subject, so all students are aware of their learning and its academic value. Acknowledging that the lowest score (apart from workload and difficulty) comes from the 'breath of coverage' category, I am going to implement the following suggested teaching strategies from the secondary SEEQ notes: 1, 14 and 16. I believe these will be the most useful for the students following this course at the current time.

- Next term, one of the ways I am going to improve my teaching is through the more blatant promotion of the academic (and other values) of this subject. The strategies I have decided to use from the SEEQ resource pages are 1,14 and 16. I believe these are the most helpful and relevant to my students at this time.

- Workload and difficulty - I am going to give them longer term plans so they can see ahead and plan their time more effectively. At present I tell them of one deadline at a time and only paint the overall picture at the start. This term I will provide a constant update and 'big picture' to help them manage their workload. Learning & Academic Value - I will adapt topics to suit their needs and match their interests. Their next task on team-working will involve them looking at their own character and matching their skills and abilities to a team role linked to theory. Students will then make a choice of business to set up for their controlled coursework. This will allow them to see more value in their studies as they have control and choice of topic.

- I am going to make a list of appointment times to see students who have underachieved on their next assessment and to talk through the main areas in which they need to improve. In addition to this there will be parents evening where I will have a 15 min. slot to meet with them and talk through any issues with the course.

- 21 Students, 7 new to Trinity Sept 2012 Teaching strategy ideas for Breadth of Coverage 15. Require your students to read current newspapers or periodicals 18. Assign "thought problems" which are typical of the problems faced by professionals in the field. The students will have to do both of these this term as part of their UNIT 3 Report on a biological problem. Teaching strategy ideas for Group Interaction 11. Redirect your students' questions: basketball vs. Ping-Pong I think I do basketball questions but perhaps slip into Ping-Pong without noticing, so will make an effort to basketball the questions between the students. 14. Encourage students to make presentations in class
- 3 students, taught by me since Sept 2012 Teaching strategy ideas for Workload & Difficulty 5. Touch base repeatedly with the fundamentals or basics. 6. Focus your course on the classic issues and concepts in your discipline. Unit 5 is synoptic so doing the above is essential to ensure progress Teaching strategy ideas for Teacher Enthusiasm 2. Vary the pace and type of activities in lessons. It is easy to slip into a comfort zone and pace with only 3 students, so variation of activities and pace needs to be focused on.

- Unclear whether work load should be high or low. Group interaction, I will try and include more practicals in my teaching. Solving problems in a group

- explain the wider importance of what we are studying, real life applications. minute responses in lessons

- Workload and difficulty approaching that of other subject. More homework needs setting, although they compared it to maths homework where they said they had far too much.

- Breath of study - offering different views is very difficult in physics, as the physics they are studying is accepted by the scientific community, unlike a humanity. I will set more research opportunities for different concepts on the course, and practical applications. Group interaction - incorporate more practicals, where the group have to work together. Homework - lack of effort on homeworks

- Breath of study - offering different views is very difficult in physics, as the physics they are studying is accepted by the scientific community, unlike a humanity. I will set more research opportunities for different concepts on the course, and practical applications. Homework - the homeworks from this group are particularly poor, with minimal effort. If they cannot do it straight away they give up. I have explained that they need to use resources available (notes, textbook, each other, me) if they cannot do it. I have also emphasised that by not trying they will learn very little. I will also give more worked examples in class, for them to refer to if needed.

- A very able group who prefer to learn concepts using calculus. I will try and find more challenging problems for them to solve from a university textbook (which I used when I studied physics at Durham!) Try to improve the breath and difficulty by setting them research on applications of the concepts we are studying in class. I believe the workload is compared to maths, for which they are studying maths and further maths and therefore have a high workload for that, and therefore I think the workload I set them is suitable.

- Try to contextualise teaching with real life examples and applications. Introduce a greater variety of activities and teaching styles. Ensure students are aware of my genuine interest in the subject and their progress in it.

- Try to contextualise teaching with real life examples and applications. Introduce a greater variety of activities and teaching styles although this will be difficult given the fast pace at which we are moving through the course. Ensure students are aware of my genuine interest in the subject and their progress in it.
- Try to contextualise teaching with real life examples and applications, particularly in the Mechanics course. Introduce a greater variety of activities and teaching styles. Ensure students are aware of my genuine interest in the subject and their progress in it.

- Try to contextualise teaching with real life examples and applications. Introduce a greater variety of activities and teaching styles. Look to

encourage students to present to others. Ensure students are aware of my genuine interest in the subject and their progress in it.

- To find more resources on topics that represent contrasting perspectives on an issue. e.g. journals and books.

- Workload and difficulty has decreased t0 3.1, so more in line with other subjects. Is this not what we are aiming for? Homework - more emphasis on independent study and reading around the subject, and evidence of this brought to class.

- Homework: Return Hwks more promptly. Workload and difficulty: Give out extension hwks where student have to read articles beyond the A level textbooks and report back to the class.

- Workload and difficulty: Empathise with students as to the difficulty of work when looking at challenging concepts. (i feel i have been a little dismissive of their trials) Breadth of coverage: require students to read more articles and text books pages as hwks, and in class to consolidate their learning.

- Feedback was similar to the other class and it was also very positive. U6D also identified “Workload and difficulty” as an aspect that could be less high (although it was not too far from a score of 3). To get closer to 3 I will follow similar goals to those as planned for U6A. Goal 1 - Highlight in lesson plans aspects of unit of known difficulty, accumulate ideas that I see help them with their understanding, use eAQA to create a record of common mistakes, as well as using a series of PPQs to help draw this out. Goal 2 - acknowledge the difficulty of certain concepts and share this with students (so they listen extra carefully to certain aspects) but also share with them suggestions/strategies for coping with these sections. Goal 3 - Touch base repeatedly with the basics. I have found this to be successful when teaching a "step subject" because it helps to keep previous but fundamental secure in the minds of students; so I will continue with this, especially when teaching Unit 5 transition metal chemistry by repeatedly revisiting the hexa-aqua ion and linking all aspects of the chemistry to it. The second aspect to work on was “Homework”. This got a positive score of 4.1 but obviously the closer to 5 the better. To try to achieve this I will: Goal 1 - return homework ASAP so that it remains fresh in their minds and the feedback has greatest impact. Goal 2 - always hand out answers, to questions in student booklets as well as set tasks. Goal 3 - make sure homework is always targeted to reinforce learning, work on difficult aspects and build up from areas of weakness so students always feel that homework is of the greatest value and the greatest use to them.

- Overall I am pleased that the results are all in the 4s except for workload and difficulty which should be lower; and except for "breadth and

coverage". This remains a contentious issue because with limited time it is my job to focus on the specification and prepare students competently for the exam. I will however seek opportunities to link work to everyday aspects and direct students to articles in science magazines to widen their experience. With losing the January exams there may also be time to include a visit to a local university. I am also involving students in more competitions.

- Although I am only typing this now - plans were made a while ago. Workload - I will discuss deadlines with pupils and respond/adapt to their workload. I will give longer for each task. I will also spend longer discussing plans. Breadth of coverage - I discussed this target with the boys and agreed I should give more actual historian names. In light of the resources provided I also will do more washing line type activities and trial of MacDonald.

Appendix 7b: reflection texts

- This has been an extremely helpful exercise, primarily because of the difference in feedback from 13 AD and 13AC. In many ways, I am not surprised by this (a combination of group-type and make up as well as topic taught) has meant that the character of these two groups has been quite different. It was useful meeting with Stephen who has given me a few useful pointers for the classes in question. We would value more explanation of what might be meant be 'good' in the workload and difficulty section. Would it be possible to break this into two separate categories?

- Reflection Statistics Advent term: Initial feedback from the class is primarily positive with consistency across all areas except breadth of coverage and workload/difficulty. I am unsure if the students feel neutral about these categories or feel that they are unable to make a measurable judgement. The spread of ability in the class is large and I would have been interested to see more detailed feedback for individual results in the workload/difficulty category. In particular, I was surprised that the data was not skewed by the input of one student who is finding the course particularly difficult. With the first term of the AS course complete, I anticipate the fact that the increased level of difficulty inherent in the work as the course progresses to impact upon students' grading in the workload and difficulty category. This will need to be factored in to any changes identified. Reflection Mechanics Advent term: I did not teach this class last year. The unit I take them for (AS Mechanics) is new to them. Feedback substantially positive – dipping for breadth of coverage and workload/difficulty. There is a clear divide in the class between those that find

the work accessible (mainly A level Physicists) and those that do not. With this in mind, I suspect that the spread of results in several categories would have been significant. The weaker, quieter students in the class can be reluctant to ask for help.

- Feedback from students: Yr 13 students commented that 'Mechanics' was not an option in the questionnaire as a subject studied and they found it difficult to answer questions requiring them to comment on how successfully I, as a teacher, managed to present points of view other than my own. Inevitably in Maths there is limited scope for this. From a teaching point of view it has proved difficult to implement all my targets from last term (short term, school inspection, January external unit exams). Nevertheless, I did implement strategies with both classes. Students gave positive verbal feedback to the tasks set.

- Relatively easy so far. Some difficulties with the website which, when teachers have to book time to meet, doesn't go down well. One pair haven't done this yet as they found the site down when they went to complete it. Some of the students were a bit fed up of having to do the survey three times for different teachers as well. Other than that, all good and straightforward and the results, as all teachers have said to me, have been enlightening.

- I have found the results quite interesting. In some areas such as student teacher relationship and group interaction, the results are as I expected however an interesting finding is that in all classes the lowest scores are recorded on the Learning and academic value section. I have found this surprising and unexpected, my students clearly do not feel they are learning anything valuable in the lessons. This is something I feel I can address. Overall, the class I expected to get my best results in I did and the same goes for the class I expected to see lower scores. I feel some of the statements could have been worded in a more student friendly way, as a number of students said they were unsure what it was asking them and so ended up putting neutral.

- Looking at the two feedback forms has been really useful. Initially, I was really pleased by my feedback - I'm only in my second year of teaching A Level so it was really a boost. But I was somewhat concerned that my lowest score was on what they learnt/ academic value which seemed rather important in the grand scheme of things. So, I tried to ensure that the lessons were more focused and improved my score there. I think the improvement in other scores is just down to an improved relationship over the few months, but it is good to realise at any rate, and I think it gives me confidence to be a bit harsher with

them when I need to be. It's interesting that homework has slipped down because that is something that I feel I have let go a little bit too, so again it reflects the reality. Also good that my 'workload/ difficulty' is heading back towards average. Think that's probably good as I can have a tendency to pitch high. It would however be absolutely fascinating to see the breakdown in terms of the individual questions? Think that would be incredibly useful. But overall this has been a good experience for me!

- Was surprised about the homework about the outlier of ranks for the homework in my year 13. Partly this is based upon a sample size of 4 who do not like my harsh stance of late homeworks. Difficult to determine if there responses to the question are reflective on this or about the lack of connectedness with the assigned work.

- Autumn term 2012 - very active teaching with Y12 students. Spring term 2013 - as Exams Officer as well as Geog teacher, the Spring term presents problems for me in terms of very active teaching! - an amount of the work done with my y12 pupils during this term was covered by departmental colleagues and also was independent work - the drop in some of the "marks" the pupils have given me is no surprise. Summer term - once more Exam Officer duties mean I do not teach so actively - much of the Y12 work was revision work.

- Interesting results. This is a group new to geology and so to have such an apparently high rating so early in the academic year is very pleasing.

- Interesting to see the results and the similarities to both the Y12 and Y13 profiles. They both have similar numbers (11 and 10) in the survey yet the responses for the Y13 were all generally a little lower than for the Y12 - complacency and comfortableness in teaching style and confidence in their own abilities? We identified that I might not be helping some pupils who don't engage in the lessons, as much as I could be. I will be watching out for that.

- All Good over here still

- Took me a while to formalise the action points on the website although I had been discussing what to do for a while. My concern is that I have picked on too many changes to make at once but we shall see. Some action points have been easier to get going than others so progress using the techniques feels a bit patchy. Some interesting ideas though!

- It has been valuable to see how students feel, especially in regards to work load and the teacher/student relationship.

- It would be useful to see a bar chart showing the frequencies of each score. Just seeing the mean values doesn't let me see the spread of opinions.

- It is rather surprising to see the negative impact on my U6 class. This is in part due to the rather strange classroom dynamics, which I referred to in Term 1 reflections & goals. L6 has remained broadly similar which suggests limited impact of my slight modifications. Perhaps if I had been bolder in one area, we may have seen a more pronounced effect.

- I have had a number of issues with this so far, which has made participating in the project difficult. I could not log onto the website when initially we were told it would be available, although that did not overly bother me initially as there are always going to be teething problems, so it is something I can anticipate. However, when I then went to my designated 1 hour with my peer coach, the website would not load (Wednesday 5th Dec, pm); similarly on Thursday I tried to access it with no results. Today, I have been filling in my different feedback for different classes, and have still one to do (U6C HIstory). However, the site is now telling me that my feedback for both L6 classes is now the same as what I just input for my U6P class, despite the fact I have spent time carefully entering different information meeting the different critiques of the feedback. It also means I cannot make the correct comments for my U6 History class. I am very interested in helping with research projects and I am interested in building on and improving my teaching, but with serious time constraints, it has been very difficult to fit this in, and to have the system not work properly is frustrating. Also, regarding the Workload and difficulty section - your notes suggest that this should not be a 5 or a 1, but somewhere in the middle, and yet because it's a lower score than my others is it suggesting that this is an area where I am weaker, whereas actually I feel being between 3 and 4 is about where I would like to be, particularly if it is in comparison with other teachers. Therefore it would have been more useful for my two other weakest topics to have been selected.

- I'm not sure that the feedback has guided me to think about anything of which I am not already aware. This is a small class of seven students; there is one pupil in it who takes up a disproportionate amount of time because he is not really capable of A2 work but is carrying on because he just passed AS and we have allowed him to continue with A2 in order to be able to try for a 3 A level

portfolio. Because he is one out of seven, I wonder if the feedback will be 'skewed', both because of his responses and his needs which certainly affect the others' learning experience. For example, the group does not feel cohesive; whatever I try to do, there is always 6 and 1. (Out of the other six, three are Oxbridge candidates and all will be predicted A grades whilst I doubt he will pass.) I would like to know whether 4 is a very positive score or whether anything below 4.5 needs a good deal of improvement. I would also like to know if one person's response in a small group is going to have a significant statistical impact - I imagine that it will.

- Using the ideas for increasing breadth covered in lessons has also improved group interactions and made the syllabus more relevant for students. I have included articles from journals for discussion, homework and presentations which have been a valuable addition to teaching. This has been easier to achieve with the modules being taught this term as they are more applied, whereas last term's were more theoretical and this approach would have been more difficult.

- The raw data provided by the pupils has made interesting reading; it has covered areas of my teaching which I had never previously considered in real depth. I am particularly eager to harness the potential of breadth and better homework provision next term and look forward to deploying my strategies for this.

- As an idea I thought this project would be a useful way to measure the "Leaner Voice" and give some impartial information about the different 6th Form classes which I teach. However in practice this proved to be an extra burden to an already pressurised course and I felt that the questions asked and feedback given were not truely clear and some of the questions appeared difficult to understand. Feedback to teachers was presented in a useful graph but it was certainly not made clear of the intentions or how the process would work which can be frustrating when we are so pressurised for time.

- See above for my comments as I see this as an example of duplicate work

- It has been very useful to receive feedback from my students.

- Ok, although it would be useful to see the average response to individual questions rather than just one for the category.

- I have found the feedback very useful. I wonder if the pupils will be become less neutral as the year progresses - particularly as they are in Year 12 and have just started the A Level course. I also think the questions 30-32 may not be greatly relevant to Spanish at this level or need further explanation perhaps. Certainly I have questions about the feedback and it would be good to be given it from an individual who can answer these rather than getting them solely from a computer.

- Had my meeting with my peer coach on Tuesday after College during a week where we are all completely exhausted after a really punishing term. It was hard to get time together, especially as we have so many other after-College commitments - meetings, marking, UCAS references, lesson observation write-ups, you name it. I had been dreading viewing the data and wondered what on earth I had been thinking to ever volunteer to take part. In fact, it gave me a massive boost as the ratings were much more positive than I had anticipated - far more than I felt I deserved given that I felt my teaching has 'slipped' down the list of priorities this term because of my new management role. It actually made me evn more determined to reflect on my teaching in a meaningful way - I feel I owe those lovely kids that. I think we were both surprised at how quickly the hour went; the ideas just flowed and it was far easier to set SMART targets than I had anticipated. Since I originally trained to teach over 10 years ago, I have not actually sat down with anyone and thought in a clearly focussed and supportive way about how I can improve things. Particularly important to consult the students in the process I think. So glad I agreed to take part!

- The results are the graph are interesting but there is not much difference between them. The workload section is difficult to assess as the A Level is broken up into units so it is dependent on where they are in the course. We had just come back from a weeks fieldtrip so they perhaps felt under pressure. If the end result is to raise achievement I am not sure that focussing on these two areas will have a dramatic difference. Reflecting on the class and what would make the biggest difference to them is changing their behaviours. The class is boy heavy and very enthusiastic. Therefore getting them settled at the start is the lesson is the most important bit.

- Generally very positive feedback from students. Very useful discussion with peer-coach about T & L, not sure about some of the suggested strategies? Theory perhaps a bit dated in some cases. Reflection of teaching and reasons for underachievement useful though not sure it tied in with graph. Have focused on 2 lowest figures to see if there is change.

- On the whole I was very happy with the results. All areas appear to be very positive. My first thought is have the students answered in a way that they think will assist me rather than their true feelings? Teacher enthusiasm was lower in my Yr 13 class which has not come as a real suprise as the current work is a little more dry than usual and suggests that my own feelings are passed onto the students. Organisation and clarity was again scored lower in Year 13 class. I need to reflect on this to ascertain why this may be as I feel nothing is different from class to class. Breadth of coverage does not concern me at present as I have a very specific area of content to cover within the subject. Homework is lower in Yr 13. I try to give more independent challenges in Yr 13 where students set their own work. Perhaps this needs to be more structured to raise the grading. I believe relationships to be high within the groups and the scores appear to support that. I have made it a target to have more group interaction within classes and this would appear to have worked. Workload and difficulty - I believe should be in the middle, which it is and so I see that as a positive. Once again I feel the feedback is very positive.

- The year 13 scores are lower than the year 12. This is possibly due to the fact that I see them less each week and so the flow of thoughts through the topics are more disjointed.

- The information provided by the questionnaire feedback has been very useful to gauge the students' feelings on the topic covered in lessons. The questions are very in depth and require students to give a lot of thoughts to their answers. We have already got a similar student voice survey in the science department, however, there are aspects of the SEEQ project that would greatly improve both the experience of answering the questions for the students and the data generated for staff.

- I am pleased to be taking in part in this project as it has focused my aims and purposes in terms of my teaching methods and style for this year. I am aware that my sixth form students value very different aspects of their education experience and they have are willing to contribute more to the relationship. It is clear to me that without a positive and productive relationship students can soon become disengaged and disillusioned with the subject. this can lead to underachievement as well as the student being reluctant to take an active role in their own learning, which is key at this stage in their education. More to follow....

- At this time of year there is great emphasis on preparing for exams, "getting through" the syllabus and working on practice papers. Inevitably this makes for

less diversity in lessons, so I am pleased that my scores this second time round haven't dropped. The students welcome the exam focus; they have university offers to meet and want to do well. They know that our students generally achieve splendid results but understand that this is almost always the result of sheer hard work. I think that it may be worth bearing in mind, however, that Year 13 feedback in general, at this time of year, may be influenced by the change in emphasis of lessons.

- I was very interested in the outcome, even though it was only for a small group. This year we have been very focussed on getting through a large quantity of subject matter and preparing for exams in Jan/June. The students are focussed on results and their grades for university offers, so they welcome this. It seems that this leaves less time for enthusiasm for the subject, which is a shame.

- On the first feedback I was interested to see the difference in perception between my L6 French and German classes (one pupil is in both classes). The second survey was more in line with my perceptions for the AS French class. Interesting in both classes and both surveys that the breadth of coverage was rated as quite low (I think that has something to do with the course as it stands). Increased workload for the French set reflects reality (also for German set). Upper 6th French set have produced almost exactly the same stats both times (note sure how many complete the 2nd one); it tallies with my views of them. So the most interesting feedback was from 6FREA1A since it was not quite what I expected. I did increase their workload (this happens naturally anyway) but I didn't feel I could be any more enthusiastic than I had already been - if anything I had less energy as we progressed (!) but they still rated my enthusiasm more highly on the second survey. They have certainly got to know me and I them.
- As there are only 2 students in this group and the question was completed quite early in the term I would like to do it again later in the year as a contrast. This is my first year of teaching this level with this exam board so it has been interesting to see at the start they found the tasks easy. Thank you.

- What about doing this across Key Stage 3/4 where the class sizes are larger?

- Useful to discuss graph with a colleague. And to consider reasons for results. Videos were slightly drawn out and a bit unnecessary.

- Useful to meet and discuss the graph with another colleague although some of the coaching video was slightly patronising? Maybe instead of answering

questions about how it would feel etc. there could be a more structured format to facilitate writing down the targets and intended outcomes.

- The process has caused me to think whether having taught some of the courses several times means that I should aim to freshen up the approaches I am using to try to put the maths into real life context when I can. Also to vary my teaching style to include more group/discussion work where possible. The lower scores on teacher enthusiasm have made me reflect on how I come across to the students both in terms of class teaching and individual relationships.

- unclear whether workload and difficulty should be 3 , lower or higher. Am unclear whether 3.3 should be increased, or is a bit too high? n.b. would be useful to see last terms goals when setting this terms.

- When the first questionnaire event occurred the experience was slightly scary because I did not know how my students would score my teaching. I have always tried to be the best teacher I can and spend a lot of time trying to improve, especially since I took over as head of department and I am working on improving A level results. When the results of the surveys were published I was pleased by the positive responses from my students although there is always room for improvement. I have read the pdf files for suggestions to improve and many aspects already strike true because they cover some of the aspects that I have identified myself over the last 2 years. So this activity will make me renew my efforts over the next term. Whether this will be enough to get my scores nearer to 5 (or nearer to 3 for Workload) I am not sure because I feel that sometimes it is subjective and not objective whether a student selects "agree" or "strongly agree" - but this won't stop me trying. The aspect that most provoked my thinking was "breadth of coverage". Whilst I completely agree with the benefit of stimulating interest by making sure there are extra-curricular activities etc for students, some of the questions they were asked in this category did not seem really appropriate for teaching A level chemistry. Of course I taught them all about the contrasting ideas about the history behind the determination of the structure of benzene but there are not many other opportunities for discussing controversial theories or contradicting ideas, and students need to know the key content at the end. of the day. So I'm not sure if I actually will have opportunity to increase my score in this area over the next term. Overall, however, this exercise is of value because it makes me consciously reflect and I believe this is one of the most important aspects of trying to be a better teacher; so I will keep on having a go!

- I have to confess that, with such a busy day-to-day workload the project has not had any direct influence upon my teaching this term. Rather it has been an exercise for the students to complete their survey towards the end of the term. However, it is interesting to view the collated results of the survey and shows an interesting similarity of areas of strength & weakness across each group. This is certainly making me think about my own teaching and perhaps areas for personal development. The other point of note was that at least one of my pupils found the wording of some of the questions difficult ("I really like the way you teach Sir.....do I put totally agree or totally disagree"!!). Interesting that (a far as I am aware) you have not collect information about the background of the pupils (racial or national origin etc) to look at whether the intervention package has differing effects.

- This was a good evaluation and the questions were relevant. The only thing that that I would have liked was more data. I would be interested to know how many people answered one of the four choices for each question.

- Having viewed the feedback I was pleased to see that my students rated my lessons highly, this has given me confidence that I am meeting their current needs. The results I have seen relate to fairly broad categories rather than individual questions, and it would be interesting to see more detailed results to see where any perceived strengths and weaknesses lie, and then be able to consider strategies for improvement. The results for perceived workload and difficulty have made me consider changing my practice slightly in this area.

- It is beneficial to be part of a project where students can, anonymously, provide critical feedback on their experience of my subjkect at post-16 Understanding the 'highs and lows' of student evaluation can help inform my own planning and practice, so that I can improve my own standards and transmit this to the student's learning environment. I would also welcome a brief qualitative analysis of the data, in writing (by e-mail!) to discuss with my faculty colleagues so as to address issues that may be common across the school in the post-16 experience. This should help to drive up standards of learning and teaching at this level. Thank you again for providing me with the opportunity to take part.

References

Abrami, P. C., & d'Apollonia, S. (1991). Multidimensional Students' Evaluations of Teaching Effectiveness - Generalizability of "N=1" Research: Comment on Marsh (1991). *Journal of Educational Psychology, 83*(3), 411-415.

Abrami, P. C., d'Apollonia, S., & Cohen, P. A. (1990). Validity of student ratings of instruction: What we know and what we do not. *Journal of Educational Psychology, 82*(2), 219-231.

Abrami, P. C., d'Apollonia, S., & Rosenfield, S. (1997). The dimensionality of student ratings of instructions: What we know and what we do not. In J. C. Smart (Ed.), *Higher Education: Handbook of Theory and Research* (Vol. 11, pp. 213-264). New York: Agathon.

Ackland, R. (1991). A review of the peer coaching literature. *The Journal of Staff Development, 12*(1), 22-27.

Adey, P., Hewitt, G., Hewitt, J., & Landau, N. (2004). *The Professional Development of Teachers: Practice and Theory*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

AERA, APA, & NCME. (1999). Standards for Educational and Psychological testing. Washington, DC: American Educational Research Association.

Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes, 50*(2), 179-211.

Aleamoni, L. M. (1981). Student ratings of instruction. In J. Millman (Ed.), Handbook of teacher evaluation (Vol. 110, pp. 110-145). Beverley Hills, CA: SAGE.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association, 91*(434), 444-455.

Apodaca, P., & Grad, H. (2005). The dimensionality of student ratings of teaching: integration of uni- and multidimensional models. *Studies in Higher Education, 30*(6), 723-748. doi:10.1080/03075070500340101

Arain, M., Campbell, M. J., Cooper, C. L., & Lancaster, G. A. (2010). What is a pilot or feasibility study? A review of current practice and editorial policy. *BMC Med Res Methodol, 10*, 67. doi:10.1186/1471-2288-10-67

Arnold, D. M., Burns, K. E., Adhikari, N. K., Kho, M. E., Meade, M. O., Cook, D. J., & McMaster Critical Care Interest, G. (2009). The design and interpretation of pilot trials in clinical research in critical care. *Crit Care Med, 37*(1 Suppl), S69-74. doi:10.1097/CCM.0b013e3181920e33

Arthur, J., Waring, M., Coe, R., & Hedges, L. V. (2012). *Research Methods and Methodologies in Education*. London: SAGE Publications.

Askew, S. (Ed.) (2000). *Feedback for learning*. London :: RoutledgeFalmer.

Asparouhov, T., & Muthén, B. (2009). Exploratory Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(3), 397-438. doi:10.1080/10705510903008204

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences, 42*(5), 815-824. doi:10.1016/j.paid.2006.09.018

BBC. (2010). Pupils 'interviewing teachers for jobs'. Retrieved from http://news.bbc.co.uk/1/hi/education/8599485.stm

BBC. (2011). Exam boards: how examiners tip off teachers to help students pass. Retrieved from http://www.telegraph.co.uk/education/secondaryeducation/8940781/Exam-boards-how-examiners-tip-off-teachers-to-help-students-pass.html

Beauducel, A., & Wittmann, W. W. (2005). Simulation Study on Fit Indexes in CFA Based on Data With Slightly Distorted Simple Structure. *Structural Equation Modeling, 12*(1), 41-75.

Benton, S. L., & Cashin, W. E. (2012). *Student Ratings of Teaching: A Summary of Research Literature*. Retrieved from Manhattan, NY:
Berk, R., & Freedman, D. (2001). *Statistical assumptions as empirical commitments*. Retrieved from http://www.stat.berkeley.edu/~census/berk2.pdf
Biesta, G. (2007). Why "what works" won't work: Evidence-based practice and the democratic deficit in educational research. *Educational theory, 57*(1), 1-22.
BIT. (2012). *Applying Behavioural Insights to Reduce Fraud, Error and Debt*. Retrieved from London:
Borko, H. (2004). Professional Development and Teacher Learning: Mapping the Terrain. *Educational Researcher, 33*(8), 3-15. doi:10.2307/3699979
Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The Concept of Validity. *Psychological Review, 111*(4), 1061-1071.
Boruch, R. (1997). *Randomized Experiments for Planning and Evaluation: A Practical Guide*. Thousand Oaks, CA: Sage.
Brinko, K. T. (1990). Instructional consultation with feedback in higher education. *The Journal of Higher Education*, 65-83.
Burdsal, C., & Harrison, P. (2008). Further evidence supporting the validity of both a multidimensional profile and an overall evaluation of teaching effectiveness. *Assessment & Evaluation in Higher Education, 33*(5), 567-576. doi:10.1080/02602930701699049
Burnyeat, M. (1994). Enthymene: The Logic of Persuasion. In D. J. Furley & A. Nehamas (Eds.), *Aristotle's Rhetoric* (pp. 3-55). Princeton: Princeton University Press.
Campbell, D. (1975). *Assessing the Impact of Planned Social Change.* Paper presented at the Social Research and Public Policies: The Dartmouth / OECD Conference, Hanover, N.H.
Campbell, M. K., Elbourne, D. R., & Altman, D. G. (2004). CONSORT statement: extension to cluster randomised trials. *BMJ, 328*(7441), 702-708. doi:10.1136/bmj.328.7441.702
Cantrell, S., & Kane, T. (2013). Ensuring Fair and Reliable Measures: Practitioner Brief: The Bill and Melinda Gates Foundation.
Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: a practical guide to doing it better*. Oxford: Oxford University Press.
Carver, R. (1978). The case against statistical significance testing. *Harvard Education Review, 48*, 378-399.
Centra, J. A. (1993). *Reflective Faculty Evaluation: Enhancing Teaching and Determining Faculty Effectiveness. The Jossey-Bass Higher and Adult Education Series*. San Francisco: Jossey-Bass.
Centra, J. A. (2003). Will Teachers Receive Higher Student Evaluations by Giving Higher Grades and Less Course Work? *Research in Higher Education, 44*(5), 495-518. doi:10.2307/40197319
Chingos, M. M., & Peterson, P. E. (2011). It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness. *Economics of Education Review, 30*(3), 449-465. doi:10.1016/j.econedurev.2010.12.010
Chowdry, H., Greaves, E., & Sibieta, L. (2010). *The Pupil Premium: Assessing the Options* (C113). Retrieved from London:
Clipa, O. (2011). Teacher Perceptions on Teacher Evaluation: the Purpose and the Assessors within the Assessment Process. *Procedia - Social and Behavioral Sciences, 29*, 158-163. doi:10.1016/j.sbspro.2011.11.220
Cochran-Smith, M. (2003). Teaching Quality Matters. *Journal of Teacher Education, 54*(2), 95-98. doi:10.1177/0022487102250283
Coe, R. (2002). *It's the effect size, stupid: What effect size is and why it is important*. Paper presented at the Annual Conference of the British Educational Research Association, Exeter, UK.

Coe, R. (2012). The Nature of Educational Research. In J. Arthur, M. Waring, R. Coe, & L. V. Hedges (Eds.), *Education Research: Methods and Methodologies*. London: Sage.
Coe, R., Aloisi, C., Higgins, S. E., & Elliot Major, L. (2014). What makes great teaching? London: Sutton Trust.
Coe, R., Fitz-Gibbon, C. T., & Tymms, P. (2000). *Promoting Evidence-Based Education: The Role of Practitioners*. Paper presented at the British Educational Research Association, Cardiff University.
Coe, R., Kime, S., Nevill, C., & Coleman, R. (2013). The DIY Evaluation Guide: Education Endowment Foundation.
Cohen, J. (1992). A Power Primer. *Psychological Bulletin*(112), 155-159.
Cohen, P. A. (1980). Effectiveness of Student-Rating Feedback for Improving College Instruction: A Meta-Analysis of Findings. *Research in Higher Education, 13*(4), 321 - 341.
Collins, K. (2013). Innovating with impact. *Teaching Leaders Quarterly, Q2*, 4-5.
Corcoran, J., & Pillai, V. (2007). A Review of the Research on Solution-Focused Therapy. *British Journal of Social Work, 39*(2), 234-242. doi:10.1093/bjsw/bcm098
Cordingley, P., & Bell, M. (2012). *Understanding What Enables High Quality Professional Learning: A report on the research evidence*. Retrieved from Centre for the Use of Research Evidence in Education (CUREE):
Cousin, G. (2009). *Researching Learning in Higher Education*. London: Routledge.
Creemers, B. P. M. (1994). Effective Instruction: An Empirical Basis for a Theory of Educational Effectiveness. In B. C. D Reynolds, P Nesselrodt, E Schaffer, S Stringfield and C Teddlie (Ed.), *Advances in School Effectiveness Research and Practice* (pp. 189 - 205): Permagon Press.
Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.
d'Apollonia, S., & Abrami, P. C. (1997). Navigating student ratings of instruction. *American Psychologist, 52*, 1198-1208.
Darling-Hammond, L., & McLaughlin, M. W. (1995). Policies that Support Professional Development in an Era of Reform. *The Phi Delta Kappan, 76*(8), 597-604.
de Shazer, S. (1994). *Words Were Originally Music*. New York: Norton.
Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *Jama, 263*(10), 1385-1389.
Dickerson, S. S., Gruenewald, T. L., & Kemeny, M. E. (2004). When the social self is threatened: Shame, physiology, and health. *Journal of personality, 72*(6), 1191-1216.
Dominick, P. G., Reilly, R. R., & McGourty, J. W. (1997). The Effects of Peer Feedback on Team Member Behavior. *Group & Organization Management, 22*(4), 508-520. doi:10.1177/1059601197224006
Dommeyer, C. J., Baum, P., Chapman, K. S., & Hanna, R. W. (2002). Attitudes of business faculty towards two methods of collecting teaching evaluations: Paper vs. online. *Assessment & Evaluation in Higher Education, 27*(5), 455-462.
Drucker, A. J., & Remmers, H. H. (1950). Do alumni and students differ in their attitudes toward instructors. *Purdue University Studies in Higher Education, 70*, 62-64.
Eldridge, S., & Kerry, S. (2012). *A Practical Guide to Cluster Randomised Trials in Health Services Research*. Chichester: Wiley.
Falk, R., & Greenbaum, C. (1995). Significance tests die hard: the amazing persistence of a probabalistic misconception. *Theory and Psychology, 5*, 75-98.
Fan, X., & SIvo, S. A. (2007). Sensitivity of Fit Indices to Model Misspecification and Model Types. *Multivariate Behavioral Research, 42*(3), 509-529.
Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction, 29*, 1-9. doi:10.1016/j.learninstruc.2013.07.001

Feldman, K. A. (1976). The Superior College Teacher from the Students' View. *Research in Higher Education, 5*(3), 243-288.
Fenstermacher, G. D., & Richardson, V. (2007). On making Determinations of Quality in Teaching. *Teachers College Record, 107*, 186-213.
Ferguson, R. F. (2008). *The Tripod Project Framework*.
Flutter, J., & Rudduck, J. (2004). *Consulting pupils: what's in it for schools?* London: RoutledgeFalmer.
Foundation, E. E. (2013). Recruitment and Retention Pack. In E. E. Foundation (Ed.). London.
Frey, P. W., Leonard, D. W., & Beatty, W. W. (1975). Student ratings of instruction: Validation research. *American educational research journal, 12*, 327-336.
George, E., Iveson, C., & Ratner, H. (2000). *Solution focused brief therapy course notes*. Retrieved from London:
Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*: WW Norton.
Gibbs, A. (2012). Focus groups and group interviews. In J. Arthur, Waring, M., Coe, R., Hedges, L. V. (Ed.), *Research Methods and Methodologies in Education*. London: SAGE.
Gingerich, W. J., & Eisengart, S. (2000). Solution-Focused Brief Therapy: A Review of the Outcome Research. *Family Process, 39*(4).
Giroux, H. A. (1986). Radical pedagogy and the politics of student voice. *Interchange, 17*(1), 48-69.
Goldacre, B. (2013). *Building evidence into education*. Retrieved from London:
Goldhaber, D., & Anthony, E. (2004). Can Teacher Quality Be Effectively Assessed?
Gollwitzer, P. M. (1993). Goal Achievement: The Role of Intentions. In W. Stroebe & M. Hewstone (Eds.), *European Review of Social Psychology* (Vol. 4). London: John Wiley & Sons.
Gollwitzer, P. M. (1999). Implementation Intentions: Strong Effects of Simple Plans. *The American Psychologist, 54*(7), 493-503.
Gorard, S. (2013). *Research Design: Creating robust approaches for the social sciences*: Sage.
Gorard, S. (2014). Confidence intervals, missing data and imputation: a salutary illustration. *International journal of research in educational methodology., 5*(3), 693-698.
Gorard, S., & Cook, T. (2007). Where does good evidence come from? *International Journal of Research & Method in Education, 30*(3), 307-323. doi:10.1080/17437270701614790
Gregory, J. B., Beck, J. W., & Carr, A. E. (2011). Goals, feedback, and self-regulation: Control theory as a natural framework for executive coaching. *Consulting Psychology Journal: Practice and Research, 63*(1), 26-38. doi:10.1037/a0023398
Gruenewald, T. L., Kemeny, M. E., Aziz, N., & Fahey, J. L. (2004). Acute threat to the social self: Shame, social self-esteem, and cortisol activity. *Psychosomatic medicine, 66*(6), 915-924.
Guskey, T. R. (1988). Teacher efficacy, self-concept, and attitudes toward the implementation of instructional innovation. *Teaching and Teacher Education, 4*(1), 63-69.
Halcomb, E., Gholizadeh, L., DiGiacomo, M., Phillips, J., & Davidson, P. (2007). Literature review: considerations in undertaking focus group research with culturally and linguistically diverse groups. *Journal of Clinical Nursing, 16*(6), 1000-1011.
Hampton, S. E., & Reiser, R. A. (2004). Effects of a theory-based feedback and consultation process on instruction and learning in college classrooms. *Research in Higher Education, 45*(5), 497-527.
Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research, 77*(1), 81-112. doi:10.3102/003465430298487
Hayduk, L., Cummings, G., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. (2007). Testing! testing! one, two, three–Testing the theory in structural equation models! *Personality and Individual Differences, 42*(5), 841-850.
Haynes, L., Service, O., Goldacre, B., & Torgerson, D. J. (2012). Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials. In B. I. T. Cabinet Office (Ed.). London: Cabinet Office.

Hedges, L. V. (2007). Effect Sizes in Cluster-Randomized Designs. *Journal of Educational and Behavioral Statistics, 32*(4), 341-370.
Hedges, L. V. (2013). *Types of Effect Sizes and Their Relative Merits*. Paper presented at the Eighth Annual Conference on Randomsied Controlled Trials in the Social Sciences, Durham University, Durham.
Hedges, L. V., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. New York: Academic Press.
Hernán, M. A., & Robins, J. M. (2006). Instruments for causal inference: an epidemiologist's dream? *Epidemiology, 17*(4), 360-372.
Higgins, J., & Green, S. (Eds.). (2008). *Cochrane Handbook for Systematic Review of Interventions*. Chichester: John Wiley & Sons.
Higgins, S. E., Cordingley, P., Greany, T., & Coe, R. (2015). Developing Great Teaching: Teacher Development Trust.
Higgins, S. E., Katsipataki, M., Kokotsaki, D., Coleman, R., Major, L. E., & Coe, R. (2013). The Sutton Trust-Education Endowment Foundation Teaching and Learning Toolkit. In E. E. Foundation (Ed.). London.
Hildebrand, M., Wilson, R. C., & Dienst, E. R. (1971). *Evaluating university teaching*. Berkeley: Center for Research and Development in Higher Education, University of California, Berkeley.
Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical Benchmarks for Interpreting Effect Size in Research. *Child Development Perspectives, 2*(3), 172-177.
Hoell, G. H., Witt, J. C., Gilbertson, D. N., Ranier, D. D., & Freeland, J. T. (1997). Increasing Teacher Intervention Implementation in General Education Settings through Consultation and Performance Feedback. *School Psychology Quarterly, 12*(1), 77-88.
Hopkins, P. E. (2007). Thinking critically and creatively about focus groups. *Area, 39*(4), 528-535.
Hox, J. (2010). *Multilevel analysis: Techniques and applications*: Routledge.
Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.
Hutcheson, G. D., & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models*: Sage.
Jinga, I., & Negreţ-Dobridor, I. (2004). *Inspecţia şcolară şi design-ul instrucţional*: Aramis.
Kahneman, D. (2011). *Thinking, fast and slow*. London: Penguin.
Kane, T., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment*. Retrieved from
Kelley, T. L. (1927). *Interpretation of educational measurements*.
Kember, D., Leung, D. Y. P., & Kwan, K. P. (2002). Does the use of student feedback questionnaires improve the overall quality of teaching? *Assessment & Evaluation in Higher Education, 27*(5), 411-425.
Kim, J. S. (2007). Examining the Effectiveness of Solution-Focused Brief Therapy: A Meta-Analysis. *Research on Social Work Practice, 18*(2), 107-116. doi:10.1177/1049731507307807
Kluger, A. N., & DeNisi, A. (1996). The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis and a Preliminary Feedback Intervention Theory. *Psychological Bulletin, 119*(2), 254-284.
Kolenikov, S., & Bollen, K. A. (2012). Testing Negative Error Variances Is a Heywood Case a Symptom of Misspecification? *Sociological Methods & Research, 41*(1), 124-167.
Kraemer, H. C., Mintz, J., Noda, A., Tinklenberg, J., & Yesavge, J. A. (2006). Caution regarding the use of pilot studies to guide power calculations for study proposals. *Archives of General Psychiatry, 63*, 484-489.
Krueger, A. B. (2000). *Economic Considerations and Class Size*. Retrieved from Princeton, NJ:

L'Hommedieu, R., Menges, R. J., & Brinko, K. T. (1990). Methodological Explanations for the Modest Effects of Feedback From Student Ratings. *Journal of Educational Psychology, 82*(2), 232 - 241.

Layne, B. H., DeCristoforo, J. R., & McGinty, D. (1999). Electronic versus traditional student ratings of instruction. *Research in Higher Education, 40*(2), 221-232.

Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The Role and Interpretation of Pilot Studies in Clinical Research. *Journal of Psychiatric Research, 45*(5), 626-629.

Levanthal, H., Singer, R., & Jones, S. (1965). Effects of Fear and Specificity of Recommendation Upon Attitudes and Behavior. *Journal of Personality and Social Psychology, 2*(1), 20-29.

Lindquist, E. F. (1940). *Statistical Analysis in Educational Research*. Boston, USA: Houghton Mifflin.

Lipsey, M. W., Puzio, K., Yun, C., Herbert, M., Steinka-Fry, K., Cole, M., . . . Busick, M. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*. Retrieved from Washington DC:

Little, J. W. (1993). Teachers' professional development in a climate of educational reform. *Educational evaluation and policy analysis, 15*(2), 129-151.

Locke, E. A., Chah, D., Harrison, S., & Lustgarten, N. (1989). Separating the effects of goal specificity from goal level. *Organizational Behavior and Human Performance, 43*, 270-287.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Eaglewood Cliffs, NJ: Prentice Hall.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57*(9), 705-717. doi:10.1037//0003-066x.57.9.705

Lodi-Smith, J., & Roberts, B. W. (2007). Social investment and personality: a meta-analysis of the relationship of personality traits to investment in work, family, religion, and volunteerism. *Pers Soc Psychol Rev, 11*(1), 68-86. doi:10.1177/1088868306294590

Louis, K. S., & Marks, H. M. (1998). Does Professional Community Affect the Classroom? Teachers' Work and Student Experiences in Restructuring Schools. *American Journal of Education, 106*(4), 532-575.

Marincovich, M. (1999). Using student feedback to improve teaching. In P. Seldin & Associates (Eds.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 45-69).

Marsh, H. W. (1982). SEEQ: A reliable, valid, and useful instrument for collecting students' evaluations of university teaching. *British Journal of Educational Psychology, 52*, 77-95.

Marsh, H. W. (1987). Students Evaluations of University Teaching: Research Findings, Methodological Issues, and Directions for Future Research. *International Journal of Education Research, 11*(3), 253-388.

Marsh, H. W. (1991). Multidimensional Students' Evaluations of Teaching Effectiveness: A Test of Alternative Higher-Order Structures. *Journal of Educational Psychology, 83*(2), 285-296.

Marsh, H. W. (2007a). Do university teachers become more effective with experience? A multilevel growth model of students' evaluations of teaching over 13 years. *Journal of Educational Psychology, 99*(4), 775-790. doi:10.1037/0022-0663.99.4.775

Marsh, H. W. (2007b). Students' evaluations of university teaching: A multidimensional perspective. In R. P. Perry & J. C. Smart (Eds.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective* (pp. 319-384). New York: Springer.

Marsh, H. W. (2007c). Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases and Usefulness. In R. C. Perry & J. C. Smart (Eds.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*: Springer Netherlands.

Marsh, H. W. (2011). *Students Evaluations of University Teaching: Ratings of Teachers, Universities and Departments*. Paper presented at the Higher Education Academy Conference.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103*, 391-410.
Marsh, H. W., & Dunkin, M. J. (1992). Students' evaluations of university teaching: a multidimensional perspective. In J. C. Smart (Ed.), *Higher education:handbook of theory and research* (Vol. 8). New York: Agathon Press.
Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralising Hu & Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320-341.
Marsh, H. W., & Hocevar, D. (1991a). Multidimensional students' evaluations of teaching effectiveness: Factor structure stability across academic discipline, instructor level, and course level. *Teaching and Teacher Education, 7*, 9-18.
Marsh, H. W., & Hocevar, D. (1991b). Students' evaluation of teaching effectiveness: the stability of mean ratings of the same teachers over a 13-year period. *Teaching and Teacher Education, 7*(4), 303-314.
Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis. *Annu Rev Clin Psychol, 10*, 85-110. doi:10.1146/annurev-clinpsy-032813-153700
Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory Structural Equation Modeling, Integrating CFA and EFA: Application to Students' Evaluations of University Teaching. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(3), 439-476. doi:10.1080/10705510903008220
Marsh, H. W., Nagengast, B., Fletcher, J., & Televantou, I. (2011). Assessing Educational Effectiveness: Policy Implications from Diverse Areas of Research. *Fiscal Studies, 32*(2), 279-295.
Marsh, H. W., & Overall, J. U. (1980). Validity of Students' Evaluations of Teaching Effectiveness: Cognitive and Affective Criteria. *Journal of Educational Psychology, 72*(4), 468 - 475.
Marsh, H. W., & Roche, L. A. (1993). The Use of Students' Evaluations and an Individually Structured Intervention to Enhance University Teaching Effectiveness. *American educational research journal, 30*(1), 217 - 251.
Marsh, H. W., & Roche, L. A. (1997). Making Students' Evaluations of Teaching Effectiveness Effective. *American Psychologist, 52*(11), 1187 - 1197.
Marsh, H. W., & Roche, L. A. (1999). Rely upon SET Research. *American Psychologist, 54*(7), 517-518.
Mayo, E. (2004). *The human problems of an industrial civilization*: Routledge.
McCarney, R., Warner, J., Iliffe, S., van Haselen, R., Griffin, M., & Fisher, P. (2007). The Hawthorne Effect: a randomised, controlled trial. *BMC Med Res Methodol, 7*(1), 30.
McGowan, W. R., & Graham, C. R. (2009). Factors contributing to improved teaching performance. *Innovative Higher Education, 34*(3), 161-171.
Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist, 35*(11), 1012.
Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.). Washington, DC: American Council on Education / Macmillan.
Messick, S. (1990). Validity of Test Interpretation and Use. Princeton, NJ: ETS.
Messick, S. (1994). The Interplay of Evidence and Consequences in the Validation of Performance Assessments. *Educational Researcher, 23*(2), 13-23.
Muthen, B. (1997). Latent variable modeling of longitudinal and multilevel data. *Sociological methodology, 27*(1), 453-480.
Muthén, L. K., & Muthén, B. O. (1998 - 2014). Mplus User's Guide (Vol. Version 7.2). Los Angelese, CA: Muthén & Muthén

Myhill, D., Jones, S., & Bailey, T. (2011). *Grammar for Writing? The impact of contextualised grammar teaching on pupils' writing and pupils' metalinguistic understanding.* Retrieved from Swindon:

Newmann, F., & Wehlage, G. G. (1996). *Authentic Achievement: Restructuring Schools for Intellectual Quality*. San Francisco: Jossey-Bass Publishers.

Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice, 14*(2), 149-170. doi:10.1080/09695940701478321

Newton, P. E. (2013). *The importance of ideas: the semantics of validity*. Hughes Hall, Cambridge University. Presentation retrieved from

Newton, P. E., & Shaw, S. D. (2014). *Validity in Educational and Psychological Assessment*. London: SAGE.

Noyes, A. (2005). Pupil voice: purpose, power and the possibilities for democratic schooling. *British Educational Research Journal, 31*(4), 533-540. doi:10.1080/01411920500153614

Oxman, A. D., & Guyatt, G. H. (1993). The Science of Reviewing Research. *Annals of the New York Academy of Sciences, 703*, 125-134.

Pallett, W. (2006). Uses and abuses of student ratings. *Evaluating faculty performance: A practical guide to assessing teaching, research, and service*, 50-65.

Penny, A. R. (2004). *Effects of Student Ratings Feedback and a Group Intervention on the Quality of University Teaching: a randomised controlled trial.* (PhD), University of Durham, Durham.

Penny, A. R., & Coe, R. (2004). Effectiveness of Consultation on Student Ratings Feedback: A Meta-Analysis. *Review of Educational Research, 74*(2), 215-253.

Puffer, S., Torgerson, D. J., & Watson, J. (2003). Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. *BMJ, 327*, 785-789.

Raudenbush, S. W., & Bloom, H. S. (2015). Learning about and from variation in program impacts using multisite trials.

Raykov, T. (1997). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research, 32*(4), 329-353.

Raykov, T., & Marcoulides, G. A. (2012). Evaluation of validity and reliability for hierarchical scales using latent variable modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 19*(3), 495-508.

Remmers, H. H., & Brandenburg, G. C. (1927). Experimental data on the Purdue rating scale for instructors. *Educational Administration and Supervision, 13*(6), 399-406.

Remmers, H. H., Martin, F. D., & Elliott, D. N. (1949). Are students' ratings of instructors related to their grades? *Purdue University Studies in Higher Education, 44*, 17-26.

Richardson, J. T. E. (2005). Instruments for obtaining student feedback: a review of the literature. *Assessment & Evaluation in Higher Education, 30*(4), 387-415. doi:10.1080/02602930500099193

Rindermann, H., Kohler, J., & Meisenberg, G. (2007). Quality of Instruction Improved by Evaluation and Consultation of Instructors. *International Journal for Academic Development, 12*(2), 73-85. doi:10.1080/13601440701604849

Robinson, C., & Taylor, C. (2013). Student voice as a contested practice: Power and participation in two student voice projects. *Improving Schools, 16*(1), 32-46. doi:10.1177/1365480212469713

Roche, L. A., & Marsh, H. W. (2002). Teaching self-concept in higher education: reflecting on multiple dimensions of teaching effectiveness. In N. Hativa & P. Goodyear (Eds.), *Teacher thinking, beliefs and knowledge in higher education*. Dordrecht: Kluwer.

Rothstein, J., & Mathis, W. J. (2013). *Review of Two Culminating Reports from the MET Project*. Retrieved from Boulder, Colorado:

Ryan, T. A. (1970). *Intentional Behaviour*. New York: Ronald Press.
Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of psychological research online, 8*(2), 23-74.
Schochet, P. Z., & Chiang, H. (2009). Technical Methods Report: Estimation and Identification of the Complier Average Causal Effect Parameter in Education RCTs. NCEE 2009-4040. *National Center for Education Evaluation and Regional Assistance*.
Schulz, K. F. (1995). Subverting Randomization in Controlled Trials. *The Journal of the American Medical Association, 274*(18), 1456-1458.
Schulz, K. F., Altman, D. G., & Moher, D. f. t. C. G. (2010). CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *Annals of Internal Medicine, 152*.
See, B. H., Gorard, S., & Siddiqui, N. (2015). Best practice in conducting RCTs: Lessons learnt from an independent evaluation of the Response-to-Intervention programme. *Studies in Educational Evaluation, 47*, 83-92.
Seijts, G. H., & Latham, G. P. (2000). The construct of goal commitment: Measurement and relationships with task performance *Problems and solutions in human assessment* (pp. 315-332): Springer.
Service, O., Hallsworth, M., Halpern, D., Algate, F., Gallagher, R., Nguyen, S., . . . Kirkman, E. (2014). *EAST: Four simple ways to apply behavioural insights*. Retrieved from
Showers, J., & Joyce, B. (1996). The evolution of peer coaching. *Educational leadership, 53*(6), 12-16.
Siddiqui, N., Gorard, S., & See, B. H. (2015). Accelerated Reader as a literacy catch-up intervention during primary to secondary school transition phase. *Educational Review*, 1-16.
Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107-120.
Sorenson, D. L., & Reiner, C. (2003). Charting the uncharted seas of online student ratings of instruction. *New directions for teaching and learning, 2003*(96), 1-24.
Spaulding, L. S., Mostert, M. P., & Beam, A. P. (2010). Is Brain Gym® an effective educational intervention? *Exceptionality, 18*(1), 18-30.
Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the Validity of Student Evaluation of Teaching: The State of the Art. *Review of Educational Research, 83*(4), 598-642. doi:10.3102/0034654313496870
Stuart, E. A., & Jo, B. (2015). Assessing the sensitivity of methods for estimating principal causal effects. *Statistical Methods in Medical Research, 24*(6), 657-674.
Tashakkori, A., & Teddlie, C. e. (2003). *Sage Handbook of Mixed Methods in Social and Behavioural Research*. Thousand Oaks, CA: Sage.
Thompson, M., & Wiliam, D. (2008). Tight by loose: A Conceptual Framework for Scaling Up School Reforms. In E. C. Wylie (Ed.). Princeton NJ: ETS.
Timperley, H. (2008) Teacher professional learning and development. *Educational Practices Series - 18*. Geneva: International Academy of Education / International Bureau of Education.
Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). *Teacher professional learning and development: Best Evidence Synthesis Iteration (BES)*. Retrieved from Wellington, New Zealand: http://www.oecd.org/edu/school/48727127.pdf
Toner, J. (2009). Small is not too small: reflections concerning the validity of very small focus groups (VSFGs). *Qualitative Social Work, 8*(2), 179-192.
Torgerson, D. J., & Torgerson, C. J. (2001). The Need for Randomised Controlled Trials in Educational Research. *British Journal of Educational Studies, 49*(3), 316-328.
Torgerson, D. J., & Torgerson, C. J. (2008). *Designing randomised trials in health, education and the social sciences: an introduction*. Basingstoke, England, and New York: Palgrave Macmillan.

Torgerson, D. J., & Torgerson, C. J. (2013). Randomised trials in education: an introductory handbook. London: Education Endowment Foundation.
Convention on the Rights of the Child, Resolution 44/25 Stat. 14 (1989).
Vaughn, S., Schumm, J., & Sinagub, J. (1996). *Focus Group Interviews in Education and Psychology*. London: Sage.
Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education, 24*(1), 80-91. doi:10.1016/j.tate.2007.01.004
Wanzare, Z., & da Costa, J. L. (2000). Supervision and staff development: Overview of the literature. *National Association of Secondary School Principals (NASSP) Bulletin, 84*(618), 47-54.
Warrington, W. G. (1973). *Student evaluation of instruction at Michigan State University.* Paper presented at the The first invitational conference on faculty effectiveness as evaluated by students, Philadelphia, Measurement and Research Center, Temple University.
Watts, D. G. (1991). Why Is Introductory Statistics Difficult to Learn? And What Can We Do to Make It Easier? *The American Statistician, 45*(4), 290-291. doi:10.2307/2684456
Wei, R. C., Darling-Hammond, L., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the United States and abroad*. Retrieved from Dallas, TX:
Weimer, M. (2009). Teachers who improved. *The teaching professor, 23*(2).
Wiggins, G. (1989). A True Test: Toward More Authentic and Equitable Assessment. *The Phi Delta Kappan, 70*(9), 703-713.
Wiliam, D. (2007). Content Then Process: Teacher Learning Communities in the Service of Formative Assessment. In D. Reeves (Ed.), *Ahead of the Curve: The Power of Assessment to Transform Teaching and Learning* (pp. 183-204). Bloomington, IN: Solution Tree Press.
Wilkinson, R., & Pickett, K. (2009). *The Spirit Level: Why equality is better for everyone*. London: Penguin Books.
Willingham, D. T. (2012). When Can You Trust the Experts? How to Tell Good Science from Bad in Education. San Francisco: Jossey-Bass.
Wilson, R. C. (1984). *Using Consultation to Improve Teaching*. Retrieved from California:
Wilson, R. C. (1986). Improving Faculty Teaching: Effective Use of Student Evaluations and Consultants. *The Journal of Higher Education, 57*(2), 196 - 211.
Wilson, S., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. *Review of research in education, 24*, 173-209.
Wong, K., & Nicotera, A. (2003). *Enhancing Teacher Quality: Peer Coaching as a Professional Development Strategy. A Preliminary Synthesis of the Literature*. Retrieved from
Wong, K., & Rutledge, S. A. (2006). *System-wide efforts to improve student achievement*. USA: Information Age Publishing.
Wright, S. L., & Jenkins-Guarnieri, M. A. (2012). Student evaluations of teaching: combining the meta-analyses and demonstrating further evidence for effective use. *Assessment & Evaluation in Higher Education, 37*(6), 683-699. doi:10.1080/02602938.2011.563279
Wylie, E. C. (2008). *Tight but Loose: Scaling Up Teacher Professional Development in Diverse Contexts*. Retrieved from Princeton, NJ:
Xiao, Z., Kasim, A., & Higgins, S. (2016). Same difference? Understanding variation in the estimation of effect sizes from educational trials. *International Journal of Educational Research, 77*, 1-14.
Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement*. Retrieved from Washington, DC:

[i] Hopewell, S., Clarke, M., Moher, D., Wager, E., Middleton, P., Altman, D.G., et al. (2008). CONSORT for reporting randomised trials in journal and conference abstracts. *Lancet*, 371:281-283

[ii] Hopewell, S., Clarke, M., Moher, D., Wager, E., Middleton, P., Altman, D.G. at al (2008). CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med* 5(1): e20

[iii] Ioannidis, J.P., Evans, S.J., Gotzsche, P.C., O'Neill, R.T., Altman, D.G., Schulz, K., Moher, D. (2004). Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 141(10):781-78