Developing, Testing and Interpreting a Cross Age Peer Tutoring Intervention for Mathematics: Social Interdependence,Systematic Reviews and an Empirical Study.

ZENELI, MIRJAN

Developing, Testing and Interpreting a Cross Age Peer Tutoring Intervention for Mathematics: Social Interdependence, Systematic Reviews and an Empirical Study

Mirjan Zeneli

Submitted in partial fulfilment of the requirement of

Doctorate of Philosophy in Education

School of Education, Durham University

September 2015

To:

Antje Hornburg and my children

Acknowledgement

I would like to start by first thanking my partner for all the patience and discussions on peer tutoring and my daughter for simply being around, reminding me when to take a break. My acknowledgement also goes to Professor Peter Tymms for all his support as a supervisor to make this work possible; you know very well that without your help this work would have not even started. I have learned a great deal from you and I am lucky to have had you as my mentor, thank you for going that extra mile. I would also like to thank Dr. David Bolden for his endless contributions and support; Dr. Maria Cockerill for her help with approaching participants and providing me with encouragement during difficult times. Special thanks also go to Professor Allen Thurston for his advice on peer tutoring theories. Acknowledgments also go to all administrative staff at Durham University, thank you for your efficiency, support and quick personalized responses. I would also like to thank all the academic authorities who helped one way or another with the schools, all the head-teachers, teachers and students who took part in this research, your help will always be remembered, too. I have learned so much from you and I look forward to carrying on working with you. Special thanks also go to ESRC and the North East Doctorial Training Centre staff for their training and financial support. Finally I would like to thank my family for motivating me, and my previous college teacher Alison Halsey, who during my time at Canterbury College served as a parent, friend and teacher, providing me with all the support that was needed to come this far.

Thank you to everyone.

Abstract

Cross-age peer tutoring is a peer learning strategy which has been shown to improve both socio and academic process of learning factors as well improve attainment in various subjects. There is, however, still room for the intervention to be developed: which was the aim of this work. This was done by applying important socio interdependent aspects such as resource, interpersonal and goal interdependence to a cross-age-peer tutoring intervention in mathematics. Prior to developing the method, the researcher engaged with the theoretical literature as well as provides two forms of systematic reviews. The newly informed cross-age peer tutoring method was then tested on three schools, two of which adopted a pre-post-test quasi-experimental design and one took a single group pre-post-test design. All the schools applied an Interdependent Cross-Age Tutoring (ICAT) format for a period of 6 weeks, on the basis of a 30 minute session once a week. Mathematics head-teachers, facilitators, teachers and students were all trained in various aspects of ICAT. To capture and interpret the impact of the intervention, performance instruments were innovated for each school, together with various previously established attitude sub-scales. In order to measure implementation fidelity ICAT lesson materials were collected for most of the topics and each school received general as well as structured pair observations from the researcher. Also, in order to explore how different groups learned under ICAT the lesson materials of the higher performing tutees were compared to those of the lower performing tutees on various aspects. The findings were mixed, with one of the quasi-experimental design schools showing a highest effect size of 0.81 favoring the ICAT group. The impact of ICAT on important and broader processes of learning attitude variables, social as well as academic, are also discussed. Comparisons of lesson materials between higher performing tutees and lower performing tutees revealed that the highest performing tutees showed better implementation of an essential socio-interdependent aspect: setting a shared academic goal.

Table of content:

List of Figures:

List of Equations:

List of Tables:

List of Abbreviations:

ANOVA	Analysis of Variance
ANCOVA	Analysis of Covariance
ASDQ-1	Academic Self-Description Questionnaire
CFA	Confirmatory Factor Analysis
CAT	Cross-age Peer Tutoring
CEM	Centre for Evaluation and Monitoring
CFI	Comparative Fit Index
CONSORT	Consolidated Standards of Reporting Trials
CWPT	Class Wide Peer Tutoring
EFA	Exploratory Factor Analysis

ERIC	Education Resource Information Centre
ES	Effect Size
ESL	English as Second Language
ICAT	Interdependent Cross-age Peer Tutoring
IMI	Intrinsic Motivation Instrument
IMPROVE	Introducing the new concepts, Metacognitive questioning, Practicing, Reviewing and reducing difficulties, Obtaining mastery, Verification, and Enrichment
PAL	Peer Assisted Learning
PALS	Peer Assisted Learning Strategies
RCT	Randomised Controlled Trial
RMSEA	Root Mean Square Error of Approximation
RPT	Reciprocal Peer Tutoring
SEN	Special Education Needs
Std.	Standard Deviation
TLI	Tucker-Lewis Index
ZPD	Zone of Proximal Development

This is an ESRC funded thesis, DTC training grant number: ES/I007296/1

1 Introduction: Context, Definitions and Thesis Structure

1.1 The English Context

There is a need in England to further improve students’ mathematics attainment, especially in comparison to other subjects. The Department of Education in England provides national data for Key Stage 2, ages 7-11, Key Stage 4, 14-16, and Key Stage 5, ages 16-18. The national data set shows that in 2014 the percentage of students who gained their Key Stage 2 exam results at a level 5 or above was 50% for reading, 52% for grammar, and only 42% for mathematics. The UK as a whole has often performed below the international average in mathematics (Pisa, 2009, 2012), although it must be acknowledged that within the UK there are contextual and performance variations.

The English educational system has been under constant radical reforms since the 1988 Education Reform Act, which introduced the National Curriculum and the National Testing Programme (Tymms, 2011). In the (2011) paper, Tymms outlines the major reforms that have taken place in England, concluding that most of those reforms have not improved the basic performance of English elementary and secondary students. This is regardless of the £500 million spent on the National Numeracy Strategy, the £500 million spent on the National Literacy Strategy and the £40 per-subject/per-student for the national test (taking into account 600,000 students with 3 subjects each year, this is £72 million per year).

In other words, regardless the billions of Pounds worth of reform in England the standards have not improved by a great deal. This was especially the case for literacy, while for numeracy there had been only modest albeit questionable progress. One of the biggest issues that the Tymms (2011) paper identifies is that for all these reforms the government failed to carry out research to actually monitor if the reforms were working, especially when considering the vast amount of taxpayers’ money being spent.

Similar conclusions on the impact of reforms in England are not recent; Tymms, (2004) and Tymms, Coe and Merrell (2005) made this point a decade ago. They did so by analysing data from various data sets such as those from the Key Stage 2 results for mathematics and reading, the Centre for Evaluation and Monitoring (CEM), and the findings by Massey, Green, Dexter, Hammet (2003), etc.

The Tymms (2011) paper concludes with three important recommendations for policy makers: a) evidence based education, rather than un-tested opinions, should lead reforms, b) moving towards a learning society, i.e. even if a reform is influenced by academic evidence, the government still needs to constantly monitor the effectiveness of their reforms since the educational context is constantly changing and c) Campbell's (1969) idea of 'Reforms as Experiment' should be a guide for politicians involved in education, hence employing more scientific ways of conducting reforms.

Evidence based interventions are therefore needed to improve students' performance and teaching standards in England and the UK as a whole. One evidence based intervention which has shown to be effective in the USA and has potential benefits for England is cooperative learning, a form of peer learning. There is a need to try and test more cooperative learning interventions in England; research suggests that although the interventions have been proven to be very effective, in the UK cooperative learning is still lacking behind (Jolliffe, 2007, 2015). This is important especially when considering the English context driving towards the idea of 'Every Child Matters: Change for Children, 2004' and how most of the advantages steaming from the cooperative learning interventions align extensively with such national direction (Jolliffe, 2007).

To conclude research suggests that it is necessary to improve mathematics performance in schools in England; one intervention type which can help improve performance as well as aligns well with the national educational directions is cooperative learning.

1.2 Definitions: Peer tutoring (a peer learning technique)

Peer tutoring is an old teaching method. In Ancient Greece the method was endorsed by Socrates and was widely in use. In the UK the method emerged during the 19th century and went by the name of the Bell-Lancaster system (Allen, 1976). Andrew Bell, not an educator then, wanted to find a way to help students in an orphan school in India and introduced the idea of peer tutoring. The system was then adopted by Joseph Lancaster in England and later in Wales (Allen, 1976). The method subsequently started to fall into disuse at the national scale, as students were not trained to tutor (Duress, 1971).

Peer tutoring is one among many forms of peer learning techniques, with peer learning being defined as the broader umbrella referring to students learning together. Peer learning varies in number, usually involving two or more students. Topping and Ehly (1998) and Topping (2007) provide a helpful typology of various forms of peer assisted learning methods. They classify most peer learning strategies under the following three umbrellas:

1) Peer facilitation and education, i.e. *peer modelling*, *peer education for health* and *peer counselling*;

2) Peer feedback, referring to *peers monitoring peers* and *peers assessing peers*, and finally;

3) Peer tutoring, i.e., one-to-one[1] interactions such as same-age *class wide peer tutoring (CWPT), same-age reciprocal peer tutoring (RPT), same-age peer assisted learning strategies (PALS), paired learning (same or cross-age tutoring), etc.* [2].

[1] Although one student might also be tutoring or tutored by two or three peers.

[2] A broader definition and explanation of different forms of peer assisted learning strategies is presented in chapter 3, Empirical Literature Review, under the section ‘Review of peer assisted learning techniques’.

The term ‘cross-age tutoring’ in this paper refers to an age gap of two-three years. ‘Same-age’ will refer to when tutoring is conducted within a classroom, either similar-ability students or mixed-ability. Reciprocal peer tutoring will be defined broadly as a classroom situation where every student has the ability to tutor and be tutored by their classroom friends in different ways. As figure 1 illustrates, at the heart of the definition of peer tutoring is the idea that the tutoring is done for the improvement of an academic subject.

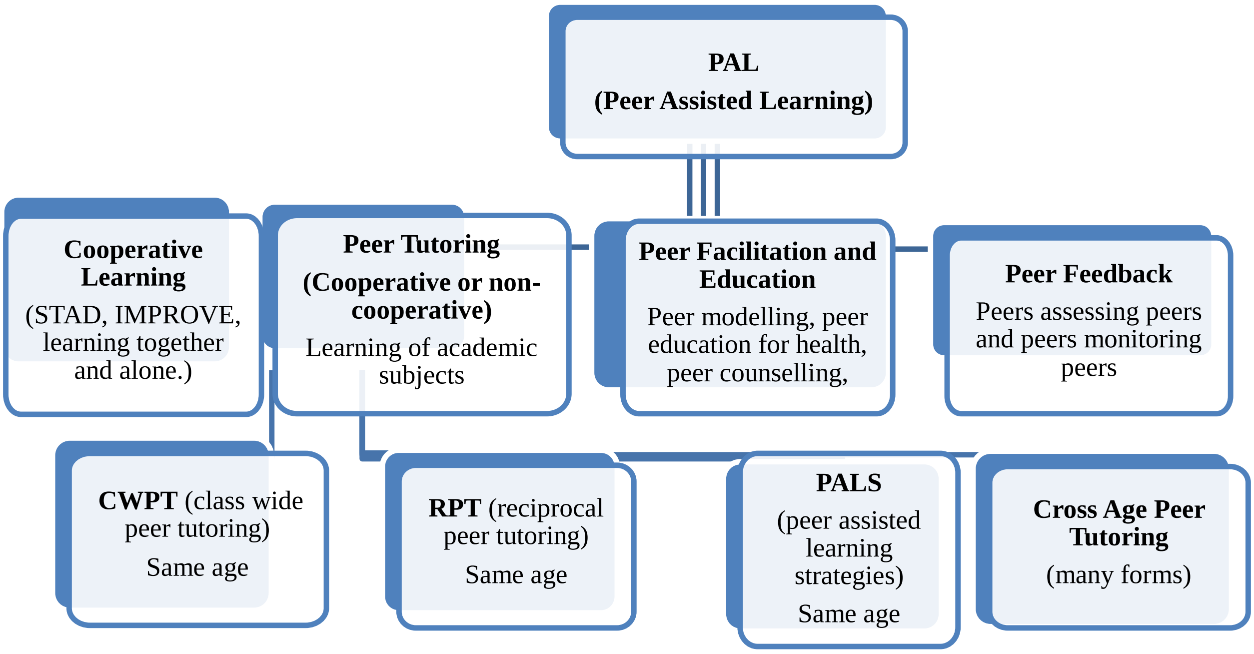


Figure 1. *Diagram categorising paired learning*

1.3 The Structure of this thesis

Following this short introduction, the work is divided into three main parts, each with three chapters: The first part covers various reviews and the aim of the research, the second part covers the methodology and the final part covers the reflections and discussion of the findings:

The review consists of a chapter entitled ‘theoretical review’, a chapter entitled ‘empirical review’ and a chapter on ‘aims, objectives research question and significance’.

The second part on the methodology looks at: ‘overall research method’, ‘instrument development and coding’ and ‘reporting the analysis and findings'.

The final part, reflection, concentrates on three aspects: ‘discussing the findings’, ‘discussing the limitations’ and ‘conclusion and recommendations’.

1.3.1 Part one

Chapter two is a review of theory and it looks at both traditional and new or dominant theories in peer tutoring, while simultaneously linking theories to one another in various ways. The chapter concludes that Social Interdependence Theory provides more answers than other theories in terms of how to make peer tutoring work, as the theory places great importance on making students’ interactions cooperative.

Chapter three is one of the lengthiest and provides an empirical review of five topics:

a) A review of the ‘what works’ literature;

b) A short review and explanation of the four most common types of peer tutoring interventions;

c) A detailed review of meta-analyses on the impact of peer learning and peer tutoring, while investigating the process of learning variables in cooperative learning environments;

d) A structured and methodical review of peer tutoring interventions in the core subjects of literacy, mathematics and science. This is undertaken in order to investigate the extent to which peer tutoring interventions have incorporated broad social interdependent elements and identified whether many social interdependent aspects have been applied to 'cross-age' peer tutoring; and

e) A systematic review of peer tutoring in mathematics in order to establish the average effect size of performance, academic attitude, and social attitude. This is conducted for same-age traditional, same-age reciprocal and cross-age peer tutoring.

The last two reviews are exclusive to this research paper and help establish: a) the rationale for this research, b) more insight into how different peer tutoring interventions relate to different theories, bringing order to any current misunderstanding and, c) the magnitude of the effect sizes for peer tutoring research in mathematics, in both achievement and social and academic attitudes.

Chapter four focuses on the *aims, objectives, research questions* and *significance*.

1.3.2 Part two

Chapter five sets out the research methods and covers issues such as: participants, research designs, data collection procedures and the details of the intervention.

Chapter six covers instrument development and looks at four topics: a) it analyses the reliability of the performance tests via Cronbach alpha for each school, b) investigates the questionnaires' reliability and validity via exploratory and confirmatory factor analysis, c) it also defines the items and coding procedures used to analyse students' lesson materials and, d) explains the observation items and the coding procedures.

Chapter seven reports the analysis methods and the findings for performance, attitude variables, students' lesson materials and the observations.

1.3.3 Part three

The discussion chapter, chapter eight, covers: discussions of implementation fidelity, students’ performance findings, the process of learning variables related to Interdependent Cross-age Peer Tutoring (ICAT), the new framework developed here and the general strengths of such framework.

Chapter nine discusses the limitations. Limitations are explored in various areas: limitations regarding the extent to which the intervention was successful in applying social interdependent elements, methodological limitations and other general limitations.

Chapter ten entitled ‘conclusion and recommendation’ provides: a) an overall conclusion, c) recommendations for how to implement cooperative learning successfully, d) how to implement cross-age peer tutoring, e) how to implement ICAT and, f) proposals on further development and research into ICAT.

The next chapter provides a theoretical literature review:

PART I

2 Theoretical Literature Review

2.1 Introduction to Peer Tutoring Theory

Peer tutoring has been explained by various theories. Some of the theories apply to the learning processes of the tutor, tutee, or both parties. The following are some of the main fields from which explanations have derived: a) educators – concentrating on the effect of *individualised instructional tasks* established within peer tutoring, b) developmental psychologists and Gestalt Theory – tackling the effect of tutoring on *socialisation, cognition* and *meta-cognition,* c) social psychologists – interested in *social interaction* and *psychological processes* and, d) educational psychologists – interested in *general improvements* of the method's procedures (Allen, 1976; Goodland & Hirst, 1989).

Prior to preceding any further it is necessary to highlight two points: firstly, the above categorisation is not mutually exclusive (Allen, 1976). For example, many educators who concentrate on individualised instructions can be either behaviourist, social or educational psychologists, developmental psychologists or all of the above. The trends could also be interlinked with each other in various ways, such as an educational psychologist could very easily subscribe to a social theory or psychological theory, for example Vygotsky can be thought to belong to various social sciences trends. Therefore, even though categorisation of theories and names will be used in this chapter, such categorisation is purely to aid the reader by organising the text.

Secondly, prior to looking at each field of theories it is necessary to make explicitly clear that most perspectives in peer tutoring do not contradict each other (Goodland & Hirst, 1989). Rather, each theory reports particular characteristics which it considers to be crucial in a peer tutoring programme; what is perceived as crucial by one theory is not perceived as such by another. It will be made evident, therefore, that the main critique towards most theories is that they do not take a broader perspective, detailing the broader socio, psychological and educational elements that are necessary for an effective peer tutoring intervention.

This chapter provides a critical evaluation of differing theoretical peer tutoring perspectives. In particular the aim of this chapter is to analyse what each theory has to say regarding the process of learning through which peer tutoring enhances performance. To achieve its aim, the paper is divided into two main parts:

The first part analyses most of the older theories and models which attempt to explain *why* and *how* peer tutoring enhances performance. Such analyses are necessary in order to understand the current dominant theories in peer tutoring, which have links with traditional views. Specifically, this section broadly analyses the similarities and differences of the following theories: Role Theory, Social Skills Theory, Self-Determination Theory, Behaviour Theory based on the S-R notion, and Piaget's Constructivist Theory.

The second part looks at current and dominant peer tutoring theories. It particularly considers Vygotsky's Social Constructivist Theory and Social Interdependence Theory. The essentials of each peer tutoring theory will be outlined.

Overall, this chapter makes three conclusions: Firstly, there are clear links between some of the peer tutoring theories. Secondly, those links can be easily recognised between the two currently dominant theories and the older theories. Finally, although there are strong links between Social Constructivism and Social Interdependence, due to efforts made by researchers such as Johnson and Johnson (1987) and Slavin (2010), this integration exists mostly at the group level of three students and over, rather than at the one-to-one peer tutoring level.

2.2 Traditional Peer Tutoring Theories

The term 'traditional peer tutoring' theory is used to encompass all those approaches that are not currently frequently used to understand or explain why or how peer tutoring works or should work.

2.2.1 Role Theory

Role Theory explains the effectiveness of peer tutoring by emphasising the benefits associated with roles. Specifically, one of the ideas why peer tutoring is superior to traditional classroom settings is that the method contains the appropriate roles in the tutoring process (Sarbin, 1976).

In order to explain why role is important Sarbin finds it necessary to first define roles in two different ways: role/status is either *earned* or *ascribed*, the former being an achievement by virtue such as being a teacher, a musician, a sportsman, etc., while the latter is assigned to a person by nature, in the sense that you may become a parent, a brother, older or younger (Sarbin, 1976). According to Sarbin (1976), when a person who has *earned* the role does not perform to the standard expected, that person is more likely to receive a neutral response from the community compared to a mother who does not perform in her role as a caring person.

The argument goes even further. It has been suggested that since the role of the classroom teacher is *earned* and the response he/she gets is neutral within the classroom setting, the teacher fails to give full dedication because there is less pressure from the classroom (Sarbin, 1976). The lack of pressure/motivation results in the teacher finding it hard to: a) understand non-verbal cues of communication, b) evaluate the true ability of the student and, c) utilise correct esteem mechanisms such as smiles, hugs and other signs of affection. Hence, as a way of evaluation and esteem improvement the teacher or adult uses stars and grades. The conclusion is that these non-verbal communication mechanisms are not effective in conveying the necessary emotional support and increase role misunderstandings between the teacher and the student (Sarbin, 1976).

The idea that adults find it hard to understand students' non-verbal cues is illustrated by an early study which recorded a short tutoring episode without verbal discussion and presented the video to two groups, students and adults: While the adults failed to understand whether the tutee in the video had comprehended his exercise, the child watchers, via reading non-verbal cues, predicted correctly (Allen & Feldman, 1975).

According to Role Theory, the beginning of peer tutoring is associated with unclear roles, as the tutor thinks he/she has to take the role of the teacher; until the tutor comes to the realisation that the role is not *earned* on the basis of merit, a condition also enabled by the fact that the tutor is unable to use esteem valuations such as stars and grades used by teachers (Sarbin, 1976). Consequently, the new identity forces the tutor to become caring, determined, and as a way of feedback and evaluation, to use smiles, affection and other non-verbal affective cues understood by the tutee. The tutor thereby gains respect from the tutee as well as recognition from the teacher (Sarbin, 1976).

Therefore, the peer tutoring setting is different from the teacher-student setting. Certain interrelated characteristics apply to both tutor and tutee: a) firstly, *ascribed* identity or role takes over, b) secondly, this in return increases involvement and a sense of obligation on both sides and, c) finally affective valuation reinforcement, or feedback, in the form of non-verbal or affective communication is used instead of esteem reinforcement based on non-affective systems such as stars or grades.

As a result of this understanding regarding the importance of the assigned role, proponents of Role Theory develop recommendations on how to improve peer tutoring interactions. Specifically, one of the strategies to improve peer tutoring effectiveness is to assign symbols to students, hence further increasing their *ascribed* roles (Sarbin & Allen, 1968). Also, in cross-age peer tutoring for instance, the age difference, among other things, can serve as a symbol in itself, reinforcing each other's roles.

Role Theory also seeks to explain psychological affective processes of learning in peer tutoring. Especially Role Theory has tried to explain why tutors in cross-age conditions have higher academic self-worth than tutors in same-age conditions. Higher academic self-worth for tutors has been established by Miller, Topping and Thurston (2010), Topping, Campbell, Douglas and Smith (2003), Tymms, Merrell, Thurston, Andor, Topping & Miller, (2011). The reason why Role Theory is strong at explaining this phenomenon is that it argues that being older is associated with more responsibilities towards the younger, and responsibility is associated with feeling important. As a result, on the tutor's side the following affective psychological factors are enhanced through peer tutoring: a) academic motivation, b) the school and subject specific attitude and, c) sense of responsibility (Allen, 1976). These conclusions have also been reported earlier by names such as Cloward (1967), Gartner, Kohler and Reisman (1971).

Moreover, for the tutor the above affective factors derive also from the class teacher role; specifically, as a teacher you: a) are respected by the old and the young, which boosts your self-esteem, b) gain recognition and a sense of responsibility to fulfil the role and the expectations, c) understand the tutee's position by empathising with the other's situation, a skill required by a good teacher, finally, d) there is the *help* element towards others, hence further enhancing the feeling of being useful and improving self-acceptance (Allen 1976, p21).

To conclude, the overall idea is that a person with an *ascribed* role is more pressured or motivated to be successful towards a particular activity than the one with an *earned* role. Evidence for Sarbin's claims is provided in the empirical evidence chapter, this is usually in the context of comparing fixed-role or cross-age peer tutoring to normal classroom teacher conditions.

2.2.2 Social Skills Theory

To recap, one of the weaknesses of Role Theory is that it over-emphasises the importance of roles in improving interactions and hence affective factors. Sarbin's view could be questioned if we were to look at this idea in the context of siblings; for example even though a brother has an assigned role it does not necessarily mean that he is pressured/motivated by society to help his siblings.

Also, simply placing an older student with a younger one does not necessarily guarantee that all their interactions will be pleasant. Some of the negative aspects in peer tutoring are that bullying may appear from the tutor side or the tutee may intimidate the tutor (Allen, 1976). Therefore, when such circumstances are allowed in the classroom, peer tutoring cannot be claimed to have its full positive effect.

Social Skills Theory, which was later fully integrated into Social Interdependence Theory reports that social skills are an important element in peer tutoring (Allen & Feldman, 1976; Argyle, 1976). Allen, Feldman & Devin-Shehan, (1976) have shown that when a tutee was programmed to provide wrong responses to the tutor, without the tutor's knowledge, it produced a negative effect on the tutor's interpersonal communication skills towards the tutee.

Social Skills Theory, however, does not emphasise the need to train students in social skills. Instead the theory suggests that the tutors in a tutoring situation already have social skills to interact and communicate with their tutee; this is the sole explanation for the effectiveness of peer tutoring. The argument goes that most of the students already have an older or younger sibling, or are part of social groups outside the classroom (Argyle, 1976).

The relationship proposed by Argyle is that social skills, such as synchronising body language, speech patterns, utterances and gazes are automatically gained by extensive peer tutoring experiences.

For Argyle (1976) social skills are the kernel of learning, since these are the skills through which feedback and assessment is gained and given within a learning context. For example, the act of looking is not just to obtain information but is also to send it. Therefore, since in dyadic interactions the parties look at each other 55% of the time and have a mutual gaze of 25% of the time, it becomes evident that intense non-verbal exchange of information takes place (Argyle, 1976; Argyle & Ingham, 1972). According to Argyle (1976), the relationship between effective peer tutoring and social skills is to some extent reciprocal as time passes, in other words, the more students interact with each other and know each other for a longer time, the more they synchronise their communication skills and enhance their communication.

Social Skills Theory also reports that peer tutoring has an effect on meta-perception. Such perceptions emerge when two students of different ages interact in order to see each other's inner mental state; hence requiring intense cognitive processes, although this can only take place with students over the age of 8 or 9 (Argyle, 1976). Why? The logic implies that meta-perception is enabled through cognitive structural similarities, since students have higher abilities than adults in putting themselves in their partners' shoes (Argyle, 1976). Such idea is based on Bonarious' (1965) conclusion that in *similar cognitive constructs* communication is easier. As already seen, this specific argument was also made by Role Theory when stating how teachers fail to communicate effectively with students, and will be made by Social Constructivism later on in the chapter.

2.2.3 Self-Determination Theory

To recap again, until this point we have seen how micro-elements of peer tutoring, such as the nature of the role and communication play an important aspect in the effectiveness of a peer tutoring programme, specifically in improving diverse emotional factors and consequently performance.

The idea that peer tutoring enhances emotional factors is also supported by Self-Determination Theory. For Self-Determination Theory, however, emotional enhancement derives not only from the micro elements of the interaction; a crucial element is also the classroom system or the activity context, which improves peers' academic attitude and consequently performance (Deci, Eghrari, Patrick & Leone 1994; Lepper & Greene 1975; Lepper & Henderson, 2000; Royan & Deci, 2000).

Hence, Self-Determination Theory, a social psychological theory, goes one step further than Role Theory and Social Skills Theory. Unlike Sarbin or Argyle, who place particular emphasis on role, identity, and interpersonal skills, Self-Determination Theory concludes that a crucial element in understanding why tutoring by a peer works is to recognise the lack of a control mechanism in the students' behaviour; fostering free will and tranquillity.

The theory goes as far as to suggest that having a teacher at the centre of the class is perceived as authoritative, and one of the reasons why many students do not find learning to be a fun activity (Royan & Deci, 2000). The argument is that although intentional monitoring and rewarding increases extrinsic motivation and hence performance in the short term, it is wiser to promote affective factors such as *intrinsic motivation;* specifically, the teacher should work towards making the students enjoy their learning by letting them determine their learning process as much as possible (Royan & Deci, 2000).

To conclude, the reason why peer tutoring works according to Self-Determination Theory, is that by its nature peer tutoring is non-competitive and reduces authoritativeness. As the tutee or tutor likes his/her partner, the activity itself transforms into a fun activity regardless of its instructional structure. Recent research has shown that when it comes to peer learning activities, the activities lacking *negative competition* or *controlling reward structures*, were overall more effective (Roseth, Johnson & Johnson, 2008). As will be shown later in this chapter, this view is also adopted by Constructivism, Social Constructivism and Social Interdependence Theory. The main contribution by Self-Determination Theory to peer tutoring is that students should be given more autonomy regarding their learning processes.

2.2.4 Skinner's S-R Model and Educationalists

Although rewards may be perceived as controlling mechanisms, one has to make a further distinction between the type of the reward and whom the reward comes from (Royan & Deci, 2000). This self-determination idea links well with Behaviour Theory explanations of peer tutoring. Studies have shown that specific and spontaneous praising, a form of reward, as opposed to constant praising, improves enjoyment and task involvement (Chalk & Lewis, 2004).

According to educationalists and Behaviour Theory, some of the most important elements in dyad interactions in tutoring sessions are the following (Harrison & Cohen, 1971):

- Praise appropriately;
- Do not punish;
- Avoid using delicate cues to prompt;
- Deal with the responses appropriately.

The above points have their origin in Skinners' behaviourist psychology of the S-R model, i.e. the stimulation/response idea (Goodland & Hirst, 1989). The idea is that in order for students to make further steps in learning by increasing both emotional and cognitive factors, they need to be constantly and appropriately *rewarded,* a strategy which can only be achieved via individualised structured interactions (Goodland & Hirst, 1989).

Hence, apart from increasing affective/emotional factors via reward and esteem mechanisms such as praising, behavioural psychologists point out how to also increase *cognitive elements* via structure and task involvement, an idea highly applied by many educational psychologists (Goodland & Hirst, 1989).

Individualised instruction within peer tutoring is seen as one of the crucial elements which enhance students' performance (Harrison 1976; Topping & Thurston 2012; Yarrow & Topping 2001). Unlike texts or workbooks, which often confuse students, personalized peer tutoring has the potential to provide students with support in reading complicated materials or in understanding tasks (Harrison, 1976). In a comparison made between tutors who were trained and tutors who were untrained in how to provide individualized instructions, Harrison and Cohen (1971) found that tutors trained in personal instructional skills helped their tutee perform four times better than those tutees who worked with un-trained tutors. Similar findings have been produced by John, Fantuzzo, King, & Heller (1992).

In a comparison between instructional types, individualized instruction via peer tutoring was found to provide the highest scores (Klosterman, 1970). Specifically, individualized instruction in peer tutoring in mathematics has been reported to provide higher performance outcomes than any other instructional technique in education in one of the earliest systematic reviews regarding mathematics instruction (Hartley, 1977).

All forms of personal instructions require certain skills, such as avoiding punishment, increasing verbal praising and task clarification in order to be effective in improving students' academic attainment. On the other hand, not all forms of individual instruction are effective if applied universally. According to educationalists the logic goes that personal instructions may differ across the subjects. For example, while it is necessary to correct a tutee with the right pronunciation fairly quickly when reading incorrectly, in mathematics the tutee needs to be shown how to answer the exercises independently via the use of physical or conceptual aids (Harrison & Cohen, 1971). Some of the first names to apply and test the effect of academically structured tutoring methods have been Ellson, Barber, Engle & Kampwerth (1965), Ellison, Harris and Barber (1968), Harrison (1971, 1972), Harrison and Brimley (1971), Niedermayer (1970), Niedermayer and Ellis (1971).

2.2.5 Constructivism

Up to this point, most of the above theories, with the exception of educationalists touching upon cognitive elements, have concentrated on the affective/emotional psychological factors enhanced by peer tutoring, as opposed to wider psychological elements such as cognition, meta-cognition or organizational skills. For example, we have looked at how role, social skills, the class system, and the praising as reward mechanism influence mainly affective psychological factors. As noted, educationalists on the other hand were the exception. By emphasizing individualized instructions they have made claims as to how wider cognitive psychological factors are enhanced by peer tutoring.

The names accredited for looking at the affective, cognitive and meta-cognitive impact of peer tutoring are those belonging to the Constructivist school of thought. For Piaget (1950) peer interactions are superior to student-adult interactions precisely due to the influence they have on enhancing even wider psychological factors, especially cognitive factors. Piaget (1950) concluded that discourse is more fruitful among peers. Discussion, the main element that results in criticism, is better fostered among equals. Thus, for Piaget it was essential that the teacher and the learner had the same authority, yet differing levels of knowledge (Piaget, 1950). The idea of reduced authority is similar to that of Self-Determination Theory, discussed earlier on.

It is suggested that cognitive change is only possible when there is a balance between *accommodations*, i.e. the level of authority is equal, and *assimilation*, referring to the introduction of new ideas and knowledge via the acceptance by the tutee of new phenomena to be cognitively internalised (De Lisi & Golbeck, 1999). It is for these reasons that peer tutoring is considered superior to adult teaching. Simply put, constructivism argues that peer tutoring provides prime conditions for a change in cognitive structures (De Lisi & Golbeck, 1999).

Regarding cognitive processes, Piaget's tutoring methods are built on theories of *equilibration*, i.e. new ideas have to be reconciled with older ones, which can be achieved when the new ideas do not extremely deviate from the existing ideas. However, in order for cognitive change to take place there has to be some form of cognitive conflict in which the tutee is lead and instructed by the tutor in a structured manner (De Lisi & Golbeck, 1999).

De Lisi and Golbeck (1999) provide a comprehensive model of the cognitive process through which peer-tutoring in general enables cognitive change and therefore enhances learning. To begin with, the existing cognitive structure is tested by a hypothesis by the peer-learner which then first leads to *assimilation* and later to *accommodation*. *Accommodation,* however, does not necessarily imply a permanent change to the existing structure (De Lisi's & Golbeck, 1999). In order for permanent change to emerge to the existing cognitive structure there has to be what is called the *perturbation-regulation-compensation sequence*. This rather extensive term relates to the idea that first the new knowledge needs to be perceived as different from the existing knowledge, hence entering the *regulation stage of perturbation*, then a cognitive decision is made, either the existing cognitive system is changed permanently or the old system is retained (De Lisi & Golbeck, 1999).

To conclude, the crucial element in understanding how Piaget's peer tutoring method results in cognitive reshaping, is to realise that among peers *cognitive challenge* is more acceptable and congruent, and reinforced by the *reflections* (meta-cognition) which emerge only once the interactions have ended, or when the tutor needs to prepare teaching materials in advance (De Lisi & Golbeck, 1999). This is important to bear in mind when it comes to comparing Piaget's view of how the social interaction enhances meta-cognition in peer tutoring to those explanations given by the current theory of Social Constructivism, the latter arguing that cognitive and meta-cognitive changes emerge during the actual interactions. The next section explores current theories:

2.3 Current Leading Peer Tutoring Theories

There are currently at least two extensively used peer tutoring theories, Social Constructivism and Social Interdependence Theory:

2.3.1 Social Constructivism

For social constructivists, appropriate academic social interaction dimensions are necessary to achieve higher psychological performance. Hence, in a more recent theoretical review of peer tutoring and Social Constructivism, Thurston, et al, (2007) suggest that a distinction and a link is usually made between *inter-psychological* (social) and *intra-psychological* (within the mind) functions when reviewing Social Constructivism. *Inter-psychological* functions are those which emerge via inter-subjectivity; in the peer tutoring context this is the degree to which peers can engage in the type of dialogue which revolves around their own worlds or views, a task easily performed by peers, however with difficulty between students and adults (Donaldson, 1987).

The claim made by Vygotsky is that *inter-psychological functions* lead to *intra-psychological function*, which makes learning possible among peers. Hence what is important is the context provided by inter-subjectivity in the 'Zone of Proximal Development' (ZPD); that is "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance, or in collaboration with more capable peers" (Vygotsky, 1986, p86).

Therefore, the link between *inter-psychological* and *intra-psychological functions* explains how a child internalizes and processes new knowledge, as well as promoting lengthy mental developments via questions and debates which extend their peers' thinking (Thurston, et al, 2007).

Gestalt theories, Social Constructivism being one of them, have emphasised that learning will emerge when a learner can place his/her ideas in an intellectual context (Goodland & Hirst, 1989). When students teach one another they first have to put effort into making the material understandable to the learner; by doing so they constantly reflect on their own learning process (Goodland & Hirst, 1989). By organising and preparing the materials for the learners, one has to reshape and reformulate ideas in new ways and consequently the tutor will understand the material better (Gartner, Kohler & Reisman, 1971).

A current dominant social constructivist figure in peer tutoring has been Keith Topping, Head of the Centre for Peer Learning at Dundee University. Topping and Ehly (1998), provide as many as 56 positive psychological processing elements involved in a peer tutoring intervention; elements, which were then used by Topping and Ehly (2001) to create a Social Constructivist model with differences to the De Lisi and Goldbeck (1999) constructivist model. The Topping and Ehly (2001) model is broader and more detailed.

The model emphasizes the effect of inter-subjectivity and communication as a means of minimizing psychological damage, under which the tutor is also a co-learner. Therefore the tutor has to learn how to find the balance of information a peer can take in, while simultaneously monitoring their performance, managing and detecting errors and giving diagnoses (Thurston, et al., 2007).

According to LeBlanc and Bearison (2004) and Topping and Ehly (2001) the following are some of the characteristics found within peer tutoring:

- Shared understanding. Understanding of a topic is perceived to be one of the crucial elements for long term memory, since understanding a material involves extensive cognitive processes (Craick & Lockhart, 1972);
- Shared positive experience**.** Having a shared positive experience is essential for the tutor since it creates a positive perception of the tutee, as well as the other way around;
- Instructional clarity**.** Another important aspect of the successful teaching session is instructional clarity, especially understanding what the roles of each party are.

The Social Constructivist links with other peer tutoring theories. There are strong links between Social Constructivism and Constructivism, Educationalist and Behaviour Theory, as well as Social Skills Theory.

The connection to Constructivist Theory. The main difference between Piaget's and Vygotsky's peer tutoring methods and theories revolve around the timing of when the organisational, affective, cognitive and meta-cognitive elements are restructured (Thurston, et al., 2007). Since Vygotsky placed more emphasis on the intra-psychological functioning and inter-subjectivity, most psychological elements, for both tutee and tutor, are reshaped during the interactions. For Piaget on the other hand the process of learning essentials on the tutor's side takes place at the pre-interaction stage, i.e. the stage of preparation, while for the tutee in the post-interaction stage (Thurston, et al., 2007).

It can be argued that if work is given to students in advance, psychological elements will also be reshaped during the interaction as well as prior to or after. If no material is given to the students prior to the interactions, however, it is most likely that most of the psychological changes will take place during and after the interactions. This explanation is consistent with the notion of learning as a circle proposed by Topping and Ehly (2001).

Also, both developmental psychologists, Piaget and Vygotsky, emphasized the importance of similar authority between the tutee and the tutor. The only major second difference between these two thinkers is that Vygotsky claimed that in order to successfully affect the ZDP cross-ability interactions are required (Foot & Howe, 1998). According to Socio Cognitive Theory (social constructivism), there is a strong connection between social settings and learning, with more competent learners acting as mediators for those less competent, providing the latter with scaffolds where knowledge can be tested (Vygotsky, 1968).

Whilst Piaget's peer learning methods may be characterized by exemplifying, questioning, disagreement and evaluating, Vygotsky's methods are characterized by exemplifying, providing explanations, bringing ideas together (co-construction), leading and hinting (Thurston, et al., 2007). Hence, in general Piaget's theory of peer learning models is usually applied to situations in which students have the same ability, while Vygotsky's models usually emphasise cross-ability and complementary interactions in which the tutor takes some form of control, an action which serves well both tutor and tutee (Foot & Howe, 1998).

Furthermore, under Topping and Ehly's (2001) social constructivist model *questioning* and *cognitive-conflict*, characteristics belonging to Piaget's constructivism, are also included. The reason for doing so is that in reality the student interaction mode does not entirely resemble that of a *cognitive-conflict* learning process of questioning and disagreement or entirely that of a cooperative learning process of providing explanations and bringing ideas together (Foot & Howe, 1998). The argument goes that the distinction is not clear since the students usually choose the strategy which suits them at a particular point in time, depending to a great extent on the knowledge the particular students hold of the task under consideration (Foot & Howe, 1998). On the other hand one can make the argument that the structure of the task does make a significant difference to shaping the nature of the interaction (Roseth, et al., 2008).

*Connection to Social Skills Theory***.** Topping and Ehly's (2001) model and Yoarrow and Topping (2011) make it clear that in order to understand the peer tutoring learning process thoroughly, it is also important to see the effect that social skills have on emotional elements, i.e. academic self-concept or motivation, as well as on cognitive and meta-cognitive states. The reason why this is important is that in return academic affective variables, such as mathematics self-concept, have a major impact on performance (Marsh, & Craven, 2006; Marsh & McDonald-Holmes, 1990; Marsh & O'Mara, 2008;; Royan & Deci, 2000), as do cognitive and meta-cognitive variables (Gabbert et al, 1986; Lovet 2008; Piaget, 1978; Valentine, DuBois, & Cooper, 2004; Vygotsky, 1968, 1978;).

For example, Yarrow and Topping (2001) in analysing the process through which paired writing enhances performance, also recommend that verbal and nonverbal cues are extremely useful for both: a) stimulating esteem and, b) increasing cooperation; which consequently shape the cognitive and meta-cognitive strategies of a particular tutoring method.

Connection to Educationalists and Behaviour Theory. The importance of individualized instructional strategy, as emphasised by educationalists, is also emphasized as a crucial part of the 2001 model, since it too improves affective, cognitive and meta-cognitive elements, and consequently performance (Yarrow & Topping, 2001). In mathematics for example, there is a difference between traditional instructions, which aims mainly to provide mechanical mathematical activities, and new individualized instructions, which incorporate strategic questioning, the latter method concentrating mostly on mathematical process and higher mathematical thinking (Burns, 1985).

Recently, educationalists have attempted to design diverse individualized instructions with the specific aim of increasing diverse socio-psychological processing factors. Yarrow and Topping (2001) presented a form of paired reading, which if followed faithfully by students would promote positive discourse among pairs. Also, if the interactions incorporate verbal mechanisms such as praising, the kernel contribution of Behaviour Theory to peer tutoring (Goodland & Hirst, 1989), this would then lead to more stimulation, task-concentration time and, reduced anxiety (Topping 1995, 2001a).

Process of Learning Outcomes. By analysing the processes and outcomes of peer tutoring methods, Topping and Thurston (2011) have tried to link specific socio interactions and cognitive talk among peers to performance. For social constructivists peer tutoring is thought to have psychological, behavioural and linguistic benefits:

Psychological Enhancements. According to meta-analyses by Cohen, et al., (1982), Leung and Marsh (2005), and reviews by Sharply, A.M., and Sharply, C.F., (1981), Topping and Ehly (1998), Roscoe and Chi (2007), the following have been identified as some of the psychological elements enhanced by peer tutoring:

Enhanced affect:

- Motivation to learn and subject specific self-conception.

Improved cognition:

- Higher level of cognitive reconstruction, via asking questions and providing explanations.
- Higher information transfer, due to better feedback.
- Intense engagement with the task and instructions.

Enhanced meta-cognition:

- Greater chance to self-correct/reflect.
- More chances to reshape and re-organise ideas, as tutees and tutors both need to provide comparative questions and elaborative explanations.
- They are also given more opportunities to assess knowledge boundaries via monitoring.

Improved Behaviour:

- Heightened academic interaction.

Enhanced cognitive talk:

- Explanations, feedback, questioning, concentrating on instructions, academic disagreements, suggesting ideas, adjusting ideas, linking ideas to real life.

Enhanced meta-cognitive talk:

- Constantly monitoring and assessing own and other's development.
- Engaging in reflective talk. Linking ideas to current knowledge.
- Elaborated/long explanations.

Linguistic:

- Students can improve their educational communication skills, specifically explaining, asking questions, clarifying, summarising or providing feedback correctly.

Considering the vast amount of advantages just outlined regarding peer tutoring, one can question why peer tutoring does not always work. Usually the response has been that researchers and teachers have not implemented peer tutoring correctly, or, as Lemons, Fuchs, D., Gilbert, Fuchs, L., (2014) have realised, the method's effectiveness has been underestimated due to school context changes, with peer tutoring subsequently being compared with another intervention rather than a true control group.

Conclusion on Social Constructivism. Social Constructivism's explanations of why and how peer tutoring works have their origin in names such as Vygotsky (1987, translated) and Brunner (1978). Their main assumption is that constructive *educational* peer interaction is the key to understanding why and how peer tutoring is effective.

According to social constructivism, peer tutoring enhances performance via a range of educational elements in the social, psychological, behavioural and linguistic domains (Cohen, et al., 1982; Leung & Marsh 2005; Sharply, A.M., & Sharply, .F., 1981; Topping & Ehly, 1998; Roscoe & Chi, 2007). A social constructivist model through which peer tutoring enhances performance is provided by Topping and Ehly (2001). The next section explores Social Interdependence Theory:

2.3.2 Social Interdependence Theory

In order to provide a full picture of the strengths of Social Interdependence Theory, it is necessary to go back to the distinction between “peer learning”, “group learning” and “peer tutoring”. As discussed in the introduction chapter, peer learning is an umbrella for all forms of peer interactions. Group learning refers to students in groups in general, peer tutoring refers only to were students take the role of the tutor; hence, peer tutoring is also by definition a form of group learning.

There are many similarities between group learning of 3 students and over and peer tutoring (usually one-to-one, but not necessary), there is however at least one main difference at the theoretical level; specifically, the mechanisms for improving cooperative group learning of 3 students and over have been extensively explained by social interdependence mechanisms. For peer tutoring such mechanisms have been limited.

Social Interdependence Theory has also brought most other peer learning theories together; however this inclusiveness of the theory has not been fully achieved by the researchers when implementing peer tutoring. A social interdependence approach has been applied to same-age peer tutoring, however not to cross-age peer tutoring[3]. Figure 2 below provides more information on the territory touched by Social Interdependence Theory.

Figure 2. *Social interdependence applicability*

[3] This argument is further developed in the latter chapters.

In order to provide a full picture of social interdependence in peer tutoring it is necessary to first look at social interdependence at the group learning level of 3, 4, and more students. Consequently, most of this section looks at social interdependence and the elements it shares with other peer tutoring theories.

Social Interdependence: An effective cooperative learning theory. Most of the above perspectives concentrate on one or two socio-psychological elements as a means of exploring the process through which peer tutoring enhances performance. Social Interdependence Theory acknowledges that the way teachers, educationalists and behaviour theorists construct their instructions has a bearing on peer-interaction and consequently diverse psychological elements (Johnson, D.W., & Johnson, R., T, 1987).

According to Social Interdependence Theory, cooperative instructional settings provide positive interdependence among peers, since the achievements of a particular student are positively interdependent and intertwined with that of the peer he/she is placed to work with (Deutsch, 1962). A recent meta-analysis has pointed out that cooperative settings have on average a higher effect size compared to competitive ones, weighted-mean effect size 0.46, and to individualistic settings, effect size 0.55 (Roseth, Johnson, D,W., & Johnson, R.T., 2008). Similarly, a best-evidence synthesis by Slavin, Lake, and Groff (2009) investigating effective interventions in primary and secondary mathematics classes revealed that the cooperative interventions had the highest weighted-mean effect size, 0.42.

The kernels of Social Interdependence Theory. The roots of social interdependence are traced to names such as Kafka (1935), Lewin (1948) and Deutsch (1949). *Positive and goal interdependence*, referring to the mastery of goals, such as becoming competent at a subject, or *promovotive interaction*, referring to the idea of encouragement, assessment and reflection, are the kernel to understanding why and how cooperative peer learning improves performance (Johnson, D.W., & Johnson, R.T., 2005; Jolliffe, 2011; Slavin 2010). According to social interdependence thinkers peer learning influences a variety of social processes of learning elements in the psychological, linguistic and behavioural domains (Ginsburg-Block, et al., 2006; Johnson, D.W., et al., 1981; Jollifee, 2007; Rohrbeck, et al., 2003; Roseth, et al., 2008).

Social Interdependence and its links to other theories. In order to get a full picture of the theory it is necessary to first identify its similarities to past peer learning theories.

*Connection to Social Skills and Role Theories***.** Johnson, D. W., (1990) reports that for cooperation to be effective, so that everyone benefits, a certain form of verbal and non-verbal dialogue needs to take place; for example, facial expressions showing smiles and interest, the tone of voice is soft, students are seated in close proximity and increasing eye contact.

Training for the tutor in how to employ academic skills is important, however training on interpersonal skills such as how to relate to the tutee, how to be friendly, how to minimise misbehaviour and praise appropriately, is equally important (Lippit, 1976); as are non-verbal cues such as smiling and tone of voice, which indicate approval and cooperation (Johnson, D. W., 1990). The same also applies for the tutee (Johnson, D.W., & Johnson, R.T., 1987).

Therefore, Johnsons' views are congruent with most theories reported earlier on. The difference, however, is that while Role Theory and Social Skills Theory suggest that peers already develop and possess social skills when interacting with one another (Sarbin, 1976), Social Interdependence Theory suggests that without further appropriate social skills training cooperation cannot function smoothly (Johnson, D.W., & Johnson R.T., 1987; Jolliffe, 2011). Simply combining cross-abilities and different ages is not sufficient to ensure that a peer learning project will actually be cooperative and hence provide positive results (Johnson, Johnson, Holubec & Roy, 1984)

The argument goes that in a peer learning problem solving setting students realise their state of understanding by talk, because all parties involved in the cooperation activity are required to justify and explain their strategies to the members of the group (Yackel, Cobb & Wood, 1991). Subsequently, communication is the first step in cooperation (Johnson, D. W.,

& Johnson, R.T., 1987; Johnson, D.W., 1990). For social interdependence these skills are usually referred to as *forming* and *functioning.*

Forming relates to teachers' organising skills such as moving the students into cooperative groups without noise and disturbance, as well as organizing the classrooms, making sure the students remain in their groups, encouraging students to participate.

Functioning relates to the skills necessary to maintain a successful working rapport among peers, such as offering procedures on how to complete the task, express verbal and non-verbal support for peers while maintaining eye contact, enthusiasm and praise (Johnson, D.W., et al., 1984).

Functioning skills, specifically non-verbal skills, which communicate feelings, are some of the hardest to train and maintain (Johnson, D.W., et al., 1984).What makes functioning skills more difficult is that peers find it hard to implement them, especially when either the tutee or the tutor do not meet the expectations of their peer. Allen and Feldman (1974) found that tutors whose tutee was successful used more positive verbal interaction, i.e. more affective phrases such as "that's right", "you are catching up".

When verbal communication is inconsistent with the non-verbal response it becomes clear that confusion will dominate the tutoring process (Johnson, D.W., 1990). Consequently, a verbal reinforcement mechanism such as praise will no longer serve their purpose for either the tutors or the tutees. The argument goes that inconsistency reminds the students of the same negative esteem mechanism as those employed by the teachers to the entire class, such as stars and grades, which are misunderstood by the students as mentioned by Role Theory. Therefore there are higher chances that underdeveloped discourse would emerge in tutoring situations where appropriate training has not been applied (Person & Graesser, 1999).

The lack of communicative social skill, could perhaps explain the lack of correlation between the frequencies of praising, as a process implementation mechanism, and outcomes in a study conducted by Topping and Thurston (2012), since the students may have either:

a) praised in the wrong sequence, b) praised simply due to being watched and, d) lacked the skills to synchronize their verbal talk with their non-verbal expressions.

It is consequently essential, in the opinion of Johnson, D.W., & Johnson, R.T., (1987), that students are aware and are trained on how they communicate verbally and non-verbally in a synchronised manner and avoid praising when they should not, an extremely difficult yet achievable task.

Connection to Behaviour and Social Constructivist Theories. In order to see how social interdependence relates to Behaviour Theory and social constructivism it is helpful to look at what Johnson, D.W., et al., (1984) call *formulating* and *fermenting* skills. These two concepts are different from the previous two just mentioned above, *forming,* referring to general classroom organisation skills, or *functioning,* referring to group social interaction skills.

By contrast, *formulating* refers to academic skills which provide all parties with a deeper understanding of the material, such as summarizing the findings out loud, aiming for accuracy by correcting the partner, finding appropriate ways to remember the findings, pushing for vocalization so that other members are able to correct, and planning out loud, – *fermenting* refers to using techniques which enhance reconceptualization of the materials being investigated. Some of the skills required to achieve *fermenting* are to criticize ideas not people, integrating different ideas within a single point, extending explanations, generating further answers or testing the ideas with reality (Johnson, D.W., et al., 1984).

However, the theory differs from social constructivism in particular in terms of the importance social interdependence places on *forming* and *functioning skills,* concluding that they are a pre-requisite for the *formulating* and *fermenting* skills to operationalize fully, hence training in *forming* and *functioning* skills is a requirement in order to move to higher interaction stages (Johnson, D.W., et al., 1984.)

Also like social constructivism, the theory points to the importance of cross-ability and shared-responsibility (Jonson, D.W., et al., 1984; Johnson D.W., & Johnson, R.T., 1987; Roseth, et al., 2008).

Social constructivist advocators (Topping & Ehly, 1998) too acknowledge that simply placing peers in groups and allowing them to talk to one another does not mean that cooperative learning will emerge. As Yarrow and Topping (2001) conclude, without appropriate verbal and non-verbal interaction in maintaining rapport (interpersonal interdependence), cognitive and meta-cognitive talk are extremely hard to achieve; hence, to an extent agreeing with social interdependence as mentioned above. However, just like Role and Social Skills Theory, Topping and Ehly do not place as much of an emphasis on training the teachers and the students on important social skills as do social interdependence advocators. This is one of the crucial points which separate social constructivism from social interdependence.

*Connection to Educational and Self-determination Theories***.** Social interdependence considers instruction (an educationalist concept) to be very important. However, there is a difference between social interdependence instruction and instruction as emphasised so far; specifically, social interdependence stresses the *style* of social interactions, rather than concentrating only on the academic contents.

For social interdependence it is essential that the social interactions are positively interdependent through mechanisms such as shared goals and the application of interpersonal communication skills. Such positive social interactions would enhance individual *intrinsic motivation* in group learning. Although intrinsic motivation by definition takes place within an individual, as mentioned according to Self-Determination Theory the context under which the individual operates can also have an influence. As we have seen so far, intrinsic motivation or enjoyment of the task is the kernel of Self-Determination Theory.

According to social interdependence there can be two forms of interactions; the following list is a comparison between peer interaction patterns under two different group conditions influenced by different instructional types such as cooperative and competitive (Johnson, D.W., & Johnson, R.T., 1987):

Positive Goal Interdependence

- Interaction is promoted
- Trust, support, liking and acceptance evolve as time passes
- High exploitation of resources, exchange of information and oral rehearsal
- Intrinsic motivation, such as high knowledge expectation, high commitment and mutual benefits
- Intensive emotional involvement while learning

Negative Interdependence

- Interaction is oppositional
- Misleading communication
- The process is characterized by dislike; rejection and distrust are the norms
- Extrinsic motivation, i.e. low success expectations, low persistence and low knowledge curiosity
- Low emotional involvement in learning

However, competitions are not necessarily all bad, contrary to self-determination general belief, some students, especially those who are the top performers, might find the process to be intrinsically motivating, especially if they are confident that they are going to win within their given groups.

Process of Learning Outcomes. According to the following meta-analyses in social interdependence literature - Johnson, D.W., Maruyama, Johnson, R.T., Nelson and Skon (1981), Ginsburg-Block, Rohrbeck, Fantuzzo and Lavigne (2006), Rohrbeck, Ginsburg-

Block, Fantuzzo & Miller (2003), - the following are the effects of a social interdependent peer learning setting on psychological, behaviour and linguistic areas:

Psychological.

- Enhancing cognitive and affective elements.
- Improved cognition and meta-cognition.
- Enhanced intrinsic motivation toward the subject, i.e. enjoying the subject.
- Improving general self-esteem.
- Students have better attitudes towards others.
- Students improve locus of control.
- Students improve subject specific self-concept.

Linguistic. Peer learning influenced by Social Interdependence Theory improves social communication skills.

Behaviour. Peer learning enhances positive interaction skills, particularly peer interactions and interactions with school personnel.

How can positive interdependence be achieved? For social interdependence thinkers, positive interdependence can be achieved at the following social context levels (Johnson & Johnson, 2005; 2008):

1) Interdependent rewards/goals: These can be divided into two: academic or non-academic interdependent rewards/goals (Johnson, D.W., & Johnson, R.T., 2005; 2008; Jolliffe, 2007). This refers to the idea that the success of one member in a group is related to the success of the other members, hence influencing students to help each other. A peer tutoring method with such characteristics is provided by Fantuzzo, et al., (1992) with same-age students; in this format the students choose their own rewards and goals together and take turns tutoring.

2) Role interdependence: This refers to the idea that students should be given clear roles such as tutor/tutee, which students then internalise and are dependent on. This is usually achieved best through cross-age tutoring (Sarbin & Allen, 1968; Sarbin, 1976; Fitz-Gibbon, 1985; 1992; 2000a; 2000b; Chambers, 2011).

3) Structural interdependence: Referring to the idea that students' interactions should be structurally guided by a peer tutoring script/framework in an informative way; in other words, guiding the students' interaction at every peer tutoring stage (Fantuzzo, et al., 1992).

4). Task interdependence: This is achieved by dividing tasks so that the students have to work individually in order to meet the common goal. (Chambers, 2011; Jolliffe, 2011)

5) Interpersonal interdependence: This refers to the idea that students need to be friendly to one another, and constantly be monitored and trained on how to master social skills such as praising, listening, asking questions, being polite, smiling at each other, etc. (Allen & Feldman, 1975). Such understanding derives from Deutsch's (1949) hypothesis that emotions are contagious and that in groups they create emotional interdependence if managed appropriately.

6) Relevance of the SPD: For example there should be cross-ability within a group so that students can give and take ideas. Students must be given materials or be paired in such a way that they influence each other's ZPD. The most effective way to ensure mixed-ability pairing is cross-age tutoring (Fitz-Gibbon, 2000a; 1992).

7) Scaffolding: This understanding derives from the idea that students test each other's knowledge while interacting and being there for each other (Brunner, 1978). Specifically, this can be achieved by asking open ended questions or giving elaborative explanations (Roscoe & Chi, 2007).

8) Employing pedagogical skills: students can be trained to improve their cognitive and meta-cognitive discussions and behaviours (Kramarski & Mevarech, 2003, 2004; Roscoe & Chi 2007; Topping & Ehly, 1998), improve pedagogical skills such as 'when' and 'how' to

give feedback (Fuchs, Fuchs, Bentz, Philiphs & Hamlett, 1994) or praise (Fitz-Gibbon 1992; 2000a).

9) Also, teachers should set, train and monitor students on most of the appropriate application of the elements outlined above (Chambers 2011; Fitz-Gibbon 1992; 2000a; Johnson, D.W., Johnson, R.T., & Helubec, 1990; Jolliffe 2011; Roseth, et al., 2008).

***Conclusion regarding Social Interdependence*.** The following distinctions have been made between a true cooperative group and a traditional learning group (Jonson, D.W., et al., 1984; Johnson, D.W., & Johnson, R.T., 1987; Roseth, et al., 2008):

Cooperative Groups

Positive Interdependence, i.e. everyone becomes concerned with the performance of the group.

Accountability is individual, i.e. everyone is held accountable by the group members.

Heterogeneous in ability.

Leadership is shared.

Responsibility is shared.

Concentration on a task is promoted by the teacher.

Social skills are taught.

Teachers intervene when observing something wrong.

Traditional Groups

No interdependence.

No individual accountability.

Homogeneous in ability.

Appointed leaders.

Responsibility is with individuals.

The teacher mainly describes the task.

Assumed social skills.

No feedback on group functioning is provided by teachers.

No group procedures to evaluate how the group is processing their interactions.

Groups are given procedures and instructions on how to interact.

In peer tutoring, social interdependence has not been fully achieved. There have, however, emerged names which have applied social interdependent elements to peer tutoring; these have mainly been in the same-age peer tutoring context rather than cross-age, in other words only in half of the peer tutoring spectrum. For example, names such as Fantuzzo, Davis and Ginsburg-Block (1995), Ginsburg-Block, and Fantuzzo (1997), Fantuzzo and Ginsburg-Block (1998), Ginsburg-Block, (2005), Ginsburg-Block, et al., (2006), have managed to bring the idea of enhancing 'cognitive skills' and 'academic structural engagement', in a 'goal/reward interdependent context' via RPT, in which it is the student who is empowered by the choice of rewards and goals.

For a detailed representational diagram of social interdependence please see Abrami, Chambers, Poulsen, DeSimone, d'Apollonia, & Howden (1995), Johnson, D.W., and Johnson, R.T., (1995). Or Slavin's (2010) diagram which also identifies the mechanisms and order through which cooperative group learning enhances performance.

2.4 Chapter Conclusion

To summarise, most peer tutoring theories overlap with each other to a great extent. Table 1 on the next page provides a summary of the various peer tutoring theories in two domains: classroom/pair context and process of learning through socio-psychological elements.[4]

Furthermore, the process of learning functions that have emerged in peer learning in general, have in turn been illustrated to enhance both social and academic performance. For example the link between the psychological process of learning elements and performance, has been referred to by different names: On attitude factors Marsh and O'Mara (2008), Valentine et al., (2004) illustrate the reciprocal effect of *academic self-concept* on performance, while Self-Determination Theory does the same with the effect of *enjoyment* on

[4] However, there is a limitation to table 1, that is, it is not detailed enough to explain which parts of each theory relate to tutees, tutors or both; Goodland and Hirst (1989) provide more information on this topic.

performance (Royan & Deci 2000). On the other hand, the relationship between cognition and performance has been reported by more traditional theorists such as Piaget and Vygotsky, while the effect of meta-cognition on performance is reported by Garofalo and Lester (1985), Weinstein, Husman, & Dierking (2000), and Lovet (2008).

Table 1. *Essential classroom context and socio-psychological elements by theory*

Perspectives	*Essentials Classroom Context*	*Socio-psychological elements predicted to be influenced.*
Role Theory (Sarbin & Allen, 1968)	Identity. Communication.	Affective factors, such as dedication, motivation, academic self-concept, reliability, affection towards others.
Social Skills Theory (Argyle, 1976)	Interpersonal skills, verbal and non-verbal communication skills.	Affective factors, specifically positive attitude towards peers.
Self-determination Theory (Royan & Deci, 2000)	The class system. Free will.	Intrinsic motivation, such as higher interest towards the subject, more choice how to conduct learning.
Behaviour Theory and Educationalist Theory (Harrison & Cohen, 1971)	Stimulation and Response. Praise. Evaluation. Individualized instructions.	Affective factor, specifically self-concept. Cognitive elements such as memory via the repetition element
Constructivism (De Lisi & Golbeck, 1999; Piaget, 1950)	Same level of authority. Cross-ability. Cognitive construction via cognitive conflict.	Organizational skills Affective Cognitive Meta-cognitive factors
Social Constructivism (Topping & Ehly 2001; Vygotsky, 1968)	Social interaction. Cognitive co-construction. Same level of authority & cross ability. Instruction. Pedagogical skills.	Organizational skills Affective (Educational) Cognitive and Meta-cognitive factors
Social Interdependence Theory. (Johnson, D.W., & Johnson, R.T., 1987; Jolliffe 2007; Slavin 2010)	All of the above, and emphasis on positive interdependence at various levels.	All of the above

Another point to conclude is that while social constructivism has been very successful in bringing within its orbits constructivism and behaviour theories (Skinner’s S-R, educationalists), social interdependence as an approach contains links with most peer tutoring theories.

In a cross age peer tutoring context, or in any peer tutoring context in general, it is clear that without the additional elements recommended by Social Interdependence theory cooperation could break down and domination or conflict could emerge due to the perceived role and authority of the tutor. A peer tutoring intervention without structure, training of students or requiring the teacher to monitor broader aspects of students' interactions, could therefore lead either to domination, conflict, cooperation, or a mixture.

In a cross age peer tutoring intervention social interdependence ideas are even more necessary since the tutor gains authority from the role as well as the age element. Specifically, without efforts at multi-levels - without a positive structure (goal, reward, autonomous, informative, resource, and task interdependence), without training the teachers to monitor both academic and interpersonal communication skills, without student training in both academic as well as cooperative skills, and without prior preparation of the classroom context such as seating quietly or ensuring there is enough space for the students to interact, - the chances for cooperation to fail are higher and the interactions could result in domination and conflict, and therefore hindering learning.

The next chapter provides an evaluation of the empirical research in peer tutoring. Amongst other findings, the empirical literature review illustrates that at the peer tutoring level in general, most peer tutoring interventions have taken a limited theoretical applicability of social interdependence in the main academic domains, literacy, mathematics and science.

3 Empirical Review

3.1 Introduction

The empirical review is the lengthiest chapter in this study, the review is divided into five parts: a) The general ‘what works’ literature, b) an overall review of different PAL methods, c) an overall review of different meta-analyses and reviews of peer tutoring effectiveness, d) a methodical evaluation of peer tutoring empirical studies and, e) a systematic review of peer tutoring interventions in mathematics.

The first section provides an overview of the ‘what works’ literature. Specifically it looks at the position of peer learning, in terms of both cooperative learning and more specifically peer tutoring.

The second section, the review of peer tutoring methods, seeks to achieve two goals: firstly to further clarify the various definitions of peer learning interventions and secondly, to identify some of the main peer tutoring techniques used to date.

The third section, i.e. the review of meta-analyses and other reviews, is also aimed at achieving two goals: The first goal is to illustrate the effectiveness of peer tutoring in terms of academic achievement outcomes. The second goal is to touch upon the process of learning mechanisms claimed to be impacted by peer tutoring.

The fourth section is a methodical empirical literature evaluation in peer tutoring; this part did not aim to estimate effect sizes or place methodological inclusion restrictions on the source of papers. Rather, quantitative, qualitative and mixed research design studies in peer tutoring were all reviewed and the aim was to evaluate the extent to which peer tutoring interventions in general had implicitly or explicitly taken a broadened theory standpoint when implemented. The evaluation was conducted by filtering through 11 social interdependent benchmarks, it then assessed how each of the peer tutoring studies in each of the main subject areas, literacy, mathematics and science, measured up to the benchmarks. In other words, the explicit aim of the empirical theoretical evaluation was to explain the significance and need for this study, as well as clarify where this study is positioned within the current peer tutoring literature overall.

The final section is a systematic review of peer tutoring interventions in mathematics. The systematic reviews included stricter inclusion criteria than the empirical theoretical evaluation. The systematic review of peer tutoring involving mathematics serves two aims: a) It identifies the overall mean effect size in mathematics achievement, academic self-concept and social self-concept; and b) it explores how all these elements differ in terms of the type of peer tutoring interventions, such as same-age, same-age reciprocal (mostly influenced by Social Interdependence Theory) and cross-age. The section provides more specific information regarding the nature and effectiveness of different peer tutoring interventions in mathematics. This is necessary since previous reviews or meta-analyses have not done so. As well as identifying what had previously worked in peer tutoring mathematics, the systematic review also aided the power analysis conducted in order to identify the sample size needed in peer tutoring mathematics for the purpose of this study, as well as identifying where this particular study sits within the peer tutoring in mathematics literature.

It is concluded that peer tutoring strategies have been found to be effective in improving learning processes by enhancing social and academic outcomes – linguistically, psychologically and behaviourally; and that peer tutoring enhances academic performance. It is also concluded that there are many forms of peer learning techniques. Within the main subject areas in peer tutoring the following have been the most common methods: Paired Tutoring (same-age or cross-age), CWPT, RPT and PALS.

It will be argued that most peer tutoring research has not taken a broader theoretical applicability when implementing peer tutoring; and that most peer tutoring methods have been conducted in the field of literacy, followed by mathematics and then science. In terms of theoretical elements *"training in peer tutoring academic behaviour"* or *"using peer tutoring* *scripts"* are the most frequent elements implemented in peer tutoring, whereas *"training in interpersonal communication skills"* is the least. Finally, the systematic review reveals that in mathematics cross-age peer tutoring and same-age reciprocal peer tutoring show higher mean effect sizes on performance and social and academic attitude variables than same-age peer tutoring in mathematics. The next section looks at the 'what works' literature:

3.2 The 'What Works' Literature

Recently Higgins, Katsipataki, Coleman, Henderson, Major & Coe (2014) have shown that cooperative learning interventions are effective, economic and well evidenced. In their premium teaching and learning toolkit, Higgins, et al., (2014) perform a thorough analysis looking at various systematic reviews and meta-analyses in education in order to determine which interventions have been shown to provide the highest educational achievements, as well as looking at the strength of the evidence. Additionally, the toolkit also looks at the costs of the interventions. The toolkit reviews 35 types of interventions, concluding that peer tutoring and meta-cognition were the only two interventions which were classified as "high impact for low cost with strong evidence". Feedback had high impact for low cost; however moderate evidence, peer group collaboration/cooperation was classified as having moderate impact for very low cost and very strong evidence.

Therefore peer tutoring is at the top of the 'what works literature' not only in terms of improving students' performance, but also due to its cost-effectives and well-established research base.

Prior to exploring the 'what works' literature it is necessary to define what is meant by an 'effect size':

There are many ways of interpreting the effect size (Coe, 2002), one of which would be to say that an effect size of over 0.9, suggests that the average person in the peer tutoring group has scored higher than 82% of the students in the comparison group when controlling for the pre-test data. Or an effect size of 0.2, which is usually the common finding in peer tutoring projects with strong designs as that by Tymms, et al, (2011), would suggest that the average student in the post-test has scored higher than 58% of the students in the comparison group.

Table 2 below is taken from Coe (2002) illustrating how other effect size coefficients can be illustrated in a similar way:

Table 2. *Interpreting the effect size*

Effect size	*Percentage in the control group*	*Effect size*	*Percentage in the control croup*
0.0	50%	0.9	82%
0.1	54%	1.0	84%
0.2	58%	1.2	88%
0.3	62%	1.4	92%
0.4	66%	1.6	95%
0.5	69%	1.8	96%
0.6	73%	2.0	98%
0.7	76%	2.5	99%
0.8	79%		

Two more recent systematic reviews have compared different educational interventions in mathematics in order to investigate what works:

First, Slavin and Lake (2008) conducted a systematic review of effective mathematics interventions in elementary schools. They used strict inclusion criteria such as matched or randomised controlled studies, studies which were 12 weeks and longer and those which had achievement measurements not inherent to experimental groups. Covering 87 studies, they found that the instructional interventions such as: cooperative learning, classroom management and motivational approaches, showed higher mean effect sizes, 0.33, than computer assisted instruction interventions, 0.19, or using mathematics curricula influenced texts, 0.10. The review divided the findings into three levels: strong evidence interventions, moderate evidence and low evidence interventions. They report that most of the interventions in the strong category were of the cooperative learning nature. It must be made clear at this point that ‘peer tutoring’ does not, by default, fall into the category of cooperative learning, especially when considering that the word ‘tutoring’ implies an asymmetry of relations between the peers. Whereas less structural cross-age tutoring is usually less cooperative, in the sense that the tutor takes control, same-age reciprocal tutoring tends to be more cooperative. Consequently, an argument can be made that what determines the nature of peer tutoring interactions, i.e. cooperative or dominating, is not only whether there is tutoring taking place, but rather the broader structure within which student pairs are placed to work.

Another similar systematic review was conducted by Slavin, Lake and Groff (2009), and investigated the most effective mathematics interventions, however this time for middle and high school students. Again, using strict inclusion criteria they compared 100 studies on instructional interventions, computer-assisted instruction and mathematics curricula. Similar to the Slavin and Lake (2008) study, they found that for mathematics curricula interventions the mean effect size was the smallest 0.03, for computer-assisted instruction the mean effect size was the second smallest 0.10, and for instructional interventions the mean effect size was 0.18. Some of the highest mean effect sizes of the instructional interventions were those of cooperative learning strategies, i.e. Student Teams Achievement-Division and IMPROVE (Introducing new concepts, Meta-cognition, Practicing, Reviewing, Obtaining, Verification and Enrichment), mean effect sizes 0.42 and 0.52 respectively.

Specific findings on the impact of peer tutoring in mathematics are also provided by Othman (1996) who conducted a meta-analysis on peer learning in general. When comparing different peer learning strategies in mathematics to one another he reports that peer tutoring showed the highest mean effect size based on 18 studies, 0.30.

3.3 Review of peer assisted learning techniques

Topping and Ehly (1998) categorise peer assisted learning as ‘peer facilitation and education’, ‘peer feedback’ and ‘peer tutoring’. This section investigates each of these categories:

3.3.1 Peer facilitation and education

Peer facilitation and education involves peer modelling, peer education for health and peer counselling (Topping & Ehly, 1998):

Peer modelling. According to Schunk (1998) peer modelling has its roots in Bandura’s Social Cognitive Theory, which emphasises the process of learning through elements such as modelling and observation of individuals; the elements in return enhance aspects such as production, attention, motivation or retention. There are at least three forms of modelling (Schunk, 1998):

*Cognitive modelling***.** Taking place when a peer verbally demonstrates and explains while the tutee mainly observes.

*Mastery and coping modelling***.** Referring to techniques which concentrate on skills, by a master model who conducts loud positive self-talk, - or developing learning self-efficacy, shifting from negative to persistent positive self-talk.

Self-modelling. Taking place when peers video/audio record themselves in pairs and then go back to reflecting on their own learning process.

Schunk (1998) concludes that *peer modelling* has produced positive results in wide areas such as eating habits, social/learning skills, self-efficacy and academic performance.

Peer counselling. This method is similar to peer education for health, with the difference of concentrating on providing help for the peer in life issues and problems. This is done by giving feedback, listening, or emphasising positive behaviour. The method has been shown to provide good results in elementary, middle, high schools, as well as university and special populations (Ehly & Vazques, 1998).

3.3.2 Peer feedback

Peer feedback can be divided into two parts, peer monitoring and peer assessment (Topping & Ehly, 1999):

Peer monitoring. This refers to the idea of peers recording the various social and academic behaviours of their peers through various systems of monitoring (Henington & Skinner, 1998). The method contains several steps: 1) identifying the behaviour or goal that needs to be monitored, 2) designing the recording system, 3) selecting the monitor, and 4) training the monitor (Henington & Skinner, 1998). However, it has been argued that the effectiveness of such method in social and academic behaviour and performance is not convincing (Hennington & Skinner, 1998).

Peer assessment. This method differs from peer monitoring in one major aspect, specifically in peer monitoring the emphasis is on improving learning behaviour rather than enhancing academic processes and outcomes, whereas in peer assessment improving academic outcome is the main aim. The method has gained credibility due to its ability of providing students with individualised feedback as well as altering cognitive structures (O’Donnell & Topping, 1998). According to O’Donnell and Topping (1998) several elements should be kept in mind in terms of peer assessment:

1) The issue of reliability and validity is a great concern in peer assessment;

2) The stated purpose of peer assessing could produce different results, i.e. when the purpose is to grade, rather than to provide developmental feedback, the peers are more lenient;

3) In summative assessment intrinsic motivation may be harmed;

4) Formative assessment provides more qualitative feedback in terms of linking the learning process to performance outcomes;

5) Although there are issues with validity and reliability in peer assessment, the vast amount and speed of feedback provided by a peer may compensate for any drawbacks;

6) The effectiveness of feedback and assessment depends on the characteristics of giver, receiver, the materials at hand, especially the extent to which the giver and receiver understand each other and the material, and finally;

7) Social dimensions, friends or competing peers, can negatively influence the reliability and validity of assessment.

3.3.3 Peer tutoring

Peer tutoring has been applied in the following conditions and contains the following characteristics: same-age or cross-age, cross-ability or same-ability pairings, same-sex or cross-sex pairs, structured or unstructured, with training or without training, intensive or short, fixed role or reciprocal, with or without goal interdependence, with or without contingencies, on male or female students, on normal populations, special populations or students with different socio economic background, and on a variety of academic subjects (Shapley & Sharply, 1981). Peer tutoring can also be applied with or without praise (Burns, 2006), with or without elaborative cognitive or meta-cognitive strategies (Fuchs, et al., 1997; King & Rosenshire, 1993; Topping, Campbell, Douglas & Smith, 2003) and with or without scripts to guide students' interactions (O'Donnell, 1999).

The peer tutoring models below are some of the most commonly used methods; some of them are very similar to one another, others are different. What the methods below all have in common is that they are all structured, they all emphasise academic skill training, usage of scripts to guide students' interactions and that they are used in diverse populations as well as in various academic subjects:

Cross-age peer tutoring: Cross-age peer tutoring takes many forms. However, as the title suggests the main basis is that the tutor and the tutee differ from one another in terms of age and consequently in ability level, with the older peers possessing a more advanced knowledge of the materials. Also, due to its nature cross-age tutoring is always fixed role, with the older student acquiring the role of the tutor (Sharply & Sharply, 1981). Beyond these three characteristics, age, ability and role, the method can have any combination of the remaining characteristics described in the first paragraph of this section.

The following is a description of common cross-age peer tutoring interventions: a) the older and the younger students are first tested, unless data on ability already exists, b) the higher performing older student is paired with the higher performing younger student, and so on down the line, to ensure that the gap in knowledge is neither too high nor too small, c) the students are trained on the peer tutoring structure, d) half of the older age moves into the younger age classroom and half of the younger age moves into the older age classroom, unless the classrooms are big enough to accommodate all students (Fitz-Gibbon, 2000).

One of the latest, lengthiest and methodologically strong cross-age peer tutoring programmes to have taken place in the UK, which also contains many of the characteristics of a peer tutoring programme outlined above, is that of Tymms, et al., (2011).

CWPT. There are at least 8 main elements in a CWPT tutoring method; 1) same-age, 2) random assignment of tutor and tutee, 3) reciprocity between peers, 4) structure, 5) method training, 6) group goal, 7) group contingency, and 8) scripts with instructions and answers (Mayer, Terry & Greenwood, 1999). Beyond these 8 elements, CWPT has tried to incorporate other strategies such as: praise, different cognitive or meta-cognitive strategies, and same or opposite sex pairings.

The following are the sequential processes applied to a traditional CWPT intervention: a) same-age students are trained in the peer tutoring strategy, b) pairs are randomly assigned to one another so that they have similar-ability, c) the class is divided into two main groups which compete for reward acquisition, d) the students switch roles during the day of tutoring each other, e) the pairs' goal is to score as many points as possible for their group, f) the pairs' interactions are guided by scripts, g) the winning team receives recognitions/a reward, this could be students clapping their hands (Delequari, Greenwood, Streeton & Hall 1983).

One of the lengthiest CWPT programmes was conducted by Greenwood, Delquary and Hall (1989); a two year study concentrating on literacy and mathematics.

The remaining two methods, RPT and PALS, derive from CWPT:

RPT. Reciprocal Peer Tutoring is similar to CWPT, in the sense that it contains reciprocity, same-age/same-ability students, training, structure, goal and reward interdependence, scripts with instructions and answer cards. Its main difference lies in the form of goal, reward structure and how students are paired. Whereas in CWPT the pairs are assigned to one of two groups and compete against one another for a reward set by the teacher, in RPT it is the pair who chooses the type of reward from a list, if they manage to meet a particular goal the pairs are then awarded their chosen reward (Fantuzzo, King & Heller, 1992). Other differences are that the students under RPT receive team work training and are not randomly assigned to one another, instead the pairs are always similar in ability while still aided by the questions and answers in the flash cards (Fantuzzo & Ginsburg-Block, 1999).

The randomised control study by Fantuzzo, et al., (1992) is one of the first to test RPT. Some RPT have gone one step further in also incorporating elaborative cognitive and meta-cognitive strategies (Ginsburg-Block & Fantuzzo, 1998).

PALS. Paired-assisted learning strategies contain all the above CWPT characteristics, however they are different to CWPT in at least four characteristics: 1) Even though same-age, the pairs are cross-ability and are not randomly allocated, 2) PALS incorporate a modelling characteristic in which the higher ability student leads the way, 3) emphasis is placed on praising, 4) emphasis is placed on cognitive and meta-cognitive strategies of learning (Fuchs, Fuchs, Karnes, Hamlett, Katzarokt & Dudka, 1998).

Extensive research has been conducted on PALS in literacy by Fuchs, Fuchs, Kazdan and Allen (1999) and Fuchs, Fuchs, Mathes and Simon (1997); and in mathematics by Fuchs, et al., (1998) and Fuchs, et al., (2000).

3.4 A review of attainment and process of learning effectiveness around peer tutoring

This section provides a review of the systematic and meta-analyses literature conducted around peer tutoring, as well as an important meta-analysis in peer learning in general by Roseth, et al., (2008).

Peer tutoring effectiveness can be evaluated in many ways, for example, it can be evaluated in terms of its impact on academic performance improvement or its impact on socio-psychological process of learning improvements: such as academic or social self-concept, motivation or cognitive and meta-cognitive behaviours. In both aspects, academic performance and process of learning elements, various meta-analyses and reviews have been conducted.

These meta-analyses and reviews can be divided into two categories: papers that have explicitly taken a social interdependent position in their hypothesis investigation (Ginsburg-Block, Rohrbeck, and Fantuzzo, 2006; Leung 2014; Rohrbeck, et al., 2003; Roseth, et al., 2008) and papers that have taken other perspectives, some of which also investigate certain social interdependent variables (Cohen, et al., 1982; Hartley, 1977; Leung, Marsh, Craven, Yeung, & Abduljabbar, 2013; Roscoe & Chi, 2007; Webb, 1991). The next two sub-sections look at these meta-analyses and other reviews.

3.4.1 Meta-Analyses review

One of the earliest meta-analyses to have been conducted in peer tutoring is that by Hartley (1977). Hartley's study looked at the impact of different forms of individualised structured learning, concluding that one to one interventions showed the highest mean effect sizes on performance outcome, 0.6. However, Hartley's study is criticised for using a limited number of studies (29 studies), including studies without control groups, or for concentrating only on performance outcome and only in mathematics (Cohen, et al., 1982).

The most commonly cited meta-analysis of peer tutoring was conducted by Cohen, et al., (1982). This study looked at the effect of peer tutoring in different academic subjects from 65 studies, students grade 1-9, for both tutor and tutee; as well as at the process of learning mechanisms, such as: the effect sizes of same-age or cross-age tutoring, structured or unstructured tutoring, trained or not trained, fixed or different instructors, short or long interventions, ability background and randomisation. The study further considered the impact of peer tutoring on attitude towards the academic subject and self-concept. The following were the findings from Cohen, et al., (1982):

On tutee's academic performance: The reported mean effect size was 0.40.

- Trained in academic skills mean effect size 0.41, untrained 0.36.
- Cross-age mean effect size 0.49, same-age 0.29.
- Structured mean effect size 0.51, unstructured 0.26.
- Random assignment to treatment/control groups mean effect size 0.46, non-random assignment 0.32.

- Fixed instructor mean effect size 0.41, different instructors 0.36.
- Time, 0-4 weeks mean effect size 0.95, 5-18 weeks 0.42, 19-36 weeks 0.16, group difference ($p<.001$).
- In mathematics effect size 0.60, reading effect size 0.29, other effect size 0.30, with a significant difference between mathematics and reading/other effect sizes, ($p<.05$).
- Low ability tutees mean effect size 0.42, middle ability tutees 0.33.
- For the tutees' attitude towards the subject mean effect size 0.29, towards general self-concept 0.09, however student general self-concept was non-significant.

On tutors' academic performance: Overall for the tutors the mean effect size was 0.33.

- Trained mean effect size 0.34, untrained 0.32, cross-age 0.35, same-age 0.28.
- Structured mean effect size 0.34, unstructured 0.32.
- Time: 0-4 weeks mean effect size 0.56, for 5-18 weeks 0.38, 19-36 weeks 0.10.
- In mathematics mean effect size 0.62, reading 0.21.
- Very low ability tutors mean effect size 0.42, low ability 0.23, and medium ability 0.25.
- For the tutors' attitude towards the subject mean effect size 0.42, and towards general self-concept 0.18.

A meta-analysis by Rohrbeck, et al., (2003) investigated the impact of PAL. This meta-analysis also included several triads. The main theoretical position of the meta-analysis is that of social interdependence; the experiments were coded for characteristics such as goal or reward interdependence, evaluation, student autonomy from the teacher, structure, as well as investigating demographic issues such as the socio economic background, age and gender pairing. The following were the findings with students of ages 5.65 to 11.50 years old, 90 studies were included, no distinctions were made between tutor and tutee effect size:

The overall mean effect size of all PAL studies tutors and tutees combined was 0.33.

- For mathematics mean effect size 0.22, social studies 0.49, science 0.62, spelling 0.21, writing 0.33, language 0.21 and literacy in total 0.27.
- Studies with fewer than 50% of their participants belonging to a low socio economic group mean effect size 0.32, studies with 50% or over 0.45.
- Urban students mean effect size 0.44, suburban-rural 0.23, significant at ($p<.001$).
- Cross-age tutoring mean effect size 0.80, same-age tutoring 0.47.
- Same-sex pairs mean effect size 0.63, mixed pairs 0.30.
- Below 23 hours mean effect size 0.38, over 23 hours 0.32.

In terms of some important socio interdependent variables:

- Interventions with goal set by teacher, mean effect size 0.30, interventions goals set by students 0.99, group difference significant at ($p<.001$).
- With reward mean effect size 0.34, without 0.26.
- Reward selected by teacher mean effect size 0.30, reward selected by students 0.89, significant at ($p<.001$).
- With less student autonomy mean effect size 0.30, with more student autonomy 0.94, group difference significant at ($p<.001$).
- Structured interactions mean effect size 0.30, unstructured interactions 0.33.

Ginsburg-Block, et al., (2006), present another meta-analysis with explicitly social interdependent variables, concentrating on paired tutoring and small group tutoring in 36 studies with elementary students. The investigation of the analysis focuses on: a) The peer tutoring impact on non-academic outcomes such as social skills, self-concept and positive behavioural outcome, as well as academic outcomes. b) The relationship of the non-academic outcomes to performance outcomes. c) The relationship of different interactional types, such as interdependent or non-interdependent to the non-academic outcomes. d) The effect of demographic characteristics on non-academic outcomes, and, e) the impact of gender pairing and PAL exposure on non-academic outcomes.

The results were:

a) Impact of peer tutoring in non-academic outcomes: on social skills outcomes mean effect size 0.28, group difference significant at ($p<.0001$). For general self-concept, mean effect size 0.18, significant at ($p<.0001$). For behaviour outcomes mean effect size 0.45, significant at ($p<.0001$). On academic outcomes mean effect size 0.35, also significant at ($p<.0001$).

b) The non-academic outcome correlation with performance outcomes was, $r=0.59$, significant at ($p<.01$).

c) Impact of different interaction structures on non-academic outcomes: for the impact on social skills: Structured interaction mean effect size 0.44, without structure 0.21, group difference significant at ($p<.01$). Student autonomy mean effect size 0.90, less autonomy 0.29, group difference significant at ($p<.0001$). With rewards mean effect size 0.35, without rewards 0.16, group difference significant at ($p<.01$). For the impact on self-concept outcomes: Structured interaction mean effect size 0.33, without structure 0.09, group difference significant at ($p<.01$). With rewards mean effect size 0.27, without rewards -0.03, group difference significant at ($p<.001$). On positive behaviour outcomes: Structured interaction mean effect size 0.41, without structure 0.77, non-significant. For student autonomy mean effect size 0.61, less autonomy 0.37. With rewards mean effect size 0.46, without rewards 0.42.

d) Effect of demographic characteristics on non-academic outcomes:

The impact on social skills outcomes: Studies with 50% or under of their participants belonging to a low socio economic group had a measure of homogeneity within the group that was Qw=12.12, non-significant, and mean effect size 0.38, while studies with over 50% of studies belonging to a low socio economic group, QW=32.16, significant at ($p<.05$) (two tailed), and mean effect size 0.51; also there was no significant difference between these two groups. Urban students mean effect size 0.41, sub-urban rural 0.27, group difference ($p<.05$). Tutee grade 1-3 effect size 0.35, grade 4-6 mean effect size 0.28.

For the impact of demography on self-concept: Studies with 50% or under of their participants belonging to a low socio economic group mean effect size .03, for studies with over 50% belonging to a low socio economic group the mean effect size was 0.23, group difference significant at ($p<.05$). Urban students mean effect size 0.55, sub-urban rural 0.07, group difference significant at ($p<.001$). Tutee grade 1-3, mean effect size 1.20, for grade 4-6 it was 0.17, significant at ($p<.01$).

The impact of demographic characteristics on positive behaviour: Studies with 50% or under of their participants belonging to a low socio economic group mean effect size 0.49, studies with over 50%, mean effect size 0.29. Urban students mean effect size 0.74, sub-urban rural 0.38, significant at ($p<.05$). Tutee grade 1-3, mean effect size 0.49, and for grade 4-6 mean effect size 0.43.

e) The impact of gender pairing and duration in hours, median=15 hours: For social skills, same gender mean effect size 0.65, mixed 0.26, significant at ($p<.01$); in terms of duration below median the mean effect size 0.38, over median effect size 0.26. For self-concept, same gender mean effect size 1.29, mixed gender 0.15, significant at ($p<.001$); in terms of time below median mean effect size 0.35, over median effect size

0.27. In terms of positive behaviour: same gender mean effect size 0.64, mixed gender 0.42; in terms of time over median mean effect size 0.99, over median .68.

One of the latest meta-analyses explicitly testing Social Interdependence Theory at the broader peer learning level, is that of Roseth, Johnson and Johnson (2008). The study represents 8 decades of research from 11 countries, including 148 studies in group learning and peer tutoring on 12-15 years of age. The aim of the research was to investigate the impact of cooperative/interdependent vs. competitive peer learning vs. individual learning settings on academic performance outcomes, as well as investigate the impact of such setting on social relationships, and the link between social relationship and academic performance outcomes. The prediction was that cooperative groups would perform better in academic and social gains, as well the creation of a positive correlation between positive peer relationships and performance outcome. The following were the results:

In terms of academic achievement:

- Cooperative vs. competitive peer learning mean effect size 0.57 in favour of cooperative peer learning, significant at ($p<.05$); cooperative vs. individual learning 0.65, significance at ($p<.01$) ; competitive vs. individual learning 0.20.

Regarding positive peer relations:

- Cooperative vs. competitive peer learning, mean effect size 0.48 in favour of cooperative peer learning, significant at ($p<.01$); cooperative vs. individual learning mean effect size 0.56, in favour of cooperative learning significant at ($p=.01$); competitive vs. individual learning mean effect size 0.03, in favour of competitive learning.

In terms of the link between positive peer relation effect size and performance outcome effect size:

- The study found a strong correlation 0.63, ($p<.05$), with a variation in achievement accounted for by positive peer relation by $R^2=0.40$.

Another meta-analysis which analyses the impact of peer tutoring on performance and academic self-concept is that by Leung, et al., (2013).This study concentrated on students from all ages and identified contextual variables which provide the highest effect sizes in peer tutoring, such as school level, ability, and form of peer tutoring. The following were the findings from 68 experiments of this meta-analysis:

In terms of self-concept: overall for self-concept mean effect size 0.88, significant at ($p<.05$). Studies targeting self-concept mean effect size 1.09, non-target self-concept 0.18, both significant at ($p<.05$).

Regarding academic performance:

For tutees: Average academic performance mean effect size 0.65, significant at ($p<.05$); elementary school 0.62, middle school 1.01, upper school 0.97. Ages 5-12 mean effect size 0.63, ages 13-18 it was 0.95, group difference significant at ($p<.001$). In terms of ability: low ability mean effect size 0.96, special needs 0.50, mixed 0.57, group difference significant at ($p<.001$).

For tutors: Average academic performance mean effect size 0.66; elementary school 0.62, middle school 1.02, upper school 0.88, group difference significant ($p<.001$). Ages 5-12 mean effect size 0.63, ages 13-18 it was 0.92, group difference significant at ($p<.01$). In terms of ability: low ability mean effect size 1.18, high ability 0.24, special needs 0.49, mixed-ability 0.57, group difference significant at ($p<.001$).

Type of peer tutoring: Same-age reciprocal mean effect size 0.74, same-age non-reciprocal 0.63, cross-age 0.44, group difference significant at ($p<.01$). Although cross-age peer tutoring appears to have the lowest effect size in this particular meta-analysis, in a different meta-analysis published one year later, cross-age tutoring is shown to have a much more similar effect size. Leung (2014) which is based on 72 studies and looks at many variables in peer tutoring provides the following findings:

Competing teams for group rewards mean effect size 0.30, not competing for group rewards 0.44. Structured tutoring mean effect size 0.40, non-structured 0.35. Same-age peer tutoring mean effect size 0.38, same-age reciprocal 0.38, and cross-age effect size 0.39. Striving to earn points mean effect size 0.30, and striving for tangible items 0.56, group difference significant at ($p<.001$). Same gender mean effect size 0.85, mixed gender 0.36, group difference significant at ($p<.001$).

Leung's (2014) study strengthens an important social interdependent view, i.e. the view that team competition does not necessarily produce higher effect sizes. Leung's study would have shed more light if it investigated the influence of reward formation to a greater extent, specifically whether goals or tangible rewards were set by teachers or those set by students differed, as they did in Rohrbeck, et al., (2003) study. One explanation why same-age tutoring appears to have similar effect size to cross-age tutoring in Leung (2014) could be that this meta-analysis includes very different student populations working in cross-age pairs, i.e. university students tutoring school students; in other words, the tutor is not necessarily a peer of similar background or age.

3.4.2 Other Reviews

Webb (1991) provides a systematic review of 17 empirical studies concentrating on student interactions in small group learning in mathematics. The review investigated the correlations between certain student interaction behaviours and their performance outcomes. The study finds that:

Firstly, out of 15 studies which investigated the relationship between observed content-explanations behaviour and student performance, 11 studies showed a significant positive correlation.

Secondly, out of 14 correlations investigating the relationship between receiving content explanations and performance only three collations were positive and significant, suggesting that in order for the explanations to be effective the receiver has to be able to understand them, internalise and apply them to the area of misunderstanding.

Thirdly, out of 8 correlations investigating the relationship between receiving the answer and performance, 5 were significantly negatively correlated to performance. Hence, when the students were given the answers directly, as opposed to first being provided explanations, the performance of the receiver decreased.

Fourthly, the three highest correlations in the relationship between receiving procedural information and performance were significant and negative. Hence, when students received help on procedural information their mathematics performance decreased.

Finally, the review also found that high ability students were more able to provide explanations and extroverted students were more likely to receive adequate help than introverted students.

Similarly, a recent systematic review by Roscoe and Chi (2007) attempts to explore the process of learning mechanisms behind peer tutoring, they concentrate on a sample of 18 studies. The studies all concentrated on linguistic-cognitive processes of learning mechanisms in peer tutoring with school and university students. However, not all of them provided the link between process of learning mechanisms and performance outcome. The following are their summarised conclusions:

- Tutor explanations are important, such as explanations in terms of concept understanding or problem solving that tutors provide when the tutee is stuck, or asks for help. The authors conclude that explanations help tutors to gain access to their previous knowledge and monitor, reshape and restructure knowledge in order to represent it to their tutee. Consequently retrieving and analysing their own information prior to presenting it to the tutee engages the tutor in meta-cognitive mental states. (Roscoe & Chi, 2007, p 547).

- Most studies that analysed explanation as a key process variable concentrated on mathematics for both tutor and tutee. The papers placed importance on conceptual/knowledge-building/elaborative explanations rather than knowledge-passing explanations; for example tutors making statements such as “this is how you answer the question”, or “this is why you answer it in this way”, or “this is how you link these ideas together or to real life”, rather than saying “this is the answer”. Training in conceptual help giving strategies is important.
- When a tutor asks a question they are automatically involved in meta-cognitive engagement, since most of the time they compare the information provided to what is already known (Roscoe & Chi, 2007, p 553).
- Self-questioning and answering, elaborative cognitive and meta-cognitive questions enhance performance. Thus instead of asking narrow questions such as asking for basic facts, or asking “what is the answer?”, the students can ask predictive, categorising, reviewing or knowledge connection questions such as “how does this relate to previous concepts?” or “how does this relate to real life?”.

The next section investigates the extent to which peer tutoring interventions have applied social interdependent/cooperative elements.

3.5 Methodical Literature Review

This section provides a methodical empirical theoretical evaluation for peer tutoring studies conducted in literacy, mathematics and science[5]. More specifically, the aim of this review was to evaluate the extent to which peer tutoring studies have taken a broad theoretical approach, specifically a social interdependent one. This was necessary in order to clarify the extent to which social interdependent elements had been implicitly or explicitly applied to peer tutoring. The review excluded interventions with disability students and concentrated on students of ages 4-18. The following were the methods, inclusion criteria and theoretical elements used to evaluate the studies, the findings and a short discussion:

3.5.1 Method

The search was mainly conducted on three online databases: PsycINFO, ERIC (Education Resources Information Centre), and ProQuest Dissertations. PsycINFO and ERIC, are considered robust search databases for systematic reviews and meta-analyses (Reed & Baxter, 2009). The following were the terms used to search the databases: “peer tutoring”, “peer learning”, “paired learning”, “cooperative learning”, “cross-age peer learning”, “reciprocal peer learning”, “peer assisted learning”, “classwide peer tutoring” and “PALS”. The search included all papers, i.e. books and journals. The searched age group was 4 to 18, and the published language was English. In total PsycINFO produced 18,623 results whereas ERIC produced 6,704, and ProQuest Dissertation produced 1,345. Other important hand-picked journals were the British Journal of Educational Psychology and The Journal of Educational Psychology. Studies included in books and academic conferences were also included (however, the number of these was very limited). Out of over 26,000 related publications that

[5] This review was aided by the knowledge and a body of literature accumulated through the MA in Research Methods Education.

emerged online, a lengthy manual effort reduced the number to 652 qualified studies relating to peer learning. The manual effort included the following steps: a) first reading the titles of each article, b) if the title contained any of the key words then the abstract was read, and c) if the abstract was relevant the article was then retained for further analysis. Of those 652, as many as 305 of the studies were reviews, comments, etc.; thus from the online sources 347 studies entered the next stage of the inclusion criteria.

3.5.2 Inclusion criteria

The following requirements were developed with the research aim in mind for this specific project, i.e. "how do past peer tutoring interventions compare to a broader theoretical position?" No restrictions were placed on the inclusion criteria in terms of the research methodology applied to each study, since the aim for this particular review was not to conduct a meta-analysis:

Type of work. To include only published work. This criterion was introduced mainly in order to ensure that the papers have already been reviewed one way or another. It must be pointed out, however, that this would introduce a publication bias.

Quality of work. Only journal articles, books or book chapters and conference papers were included. Trials published in newspapers, magazines, websites, schools, etc., were excluded. Again, this criterion was introduced mainly in order to ensure that the papers have already been academically scrutinised to some extent, although it is true that not all published work is peer-reviewed, or of good quality.

Time scale. Only the papers published since 1965 to 2013 were to be reviewed, hence the last 47-48 years of work. The main reason for this was that most of the theories in peer tutoring had only started to integrate in the last 30-40 years, and 40 years ago there were not many peer tutoring studies, as indicated by the Cohen, et al., (1982) meta-analysis.

Types of 'peer learning' studies. Only studies that contained peer tutoring were to enter the analysis. The logic for this criterion was that, as already mentioned, peer-learning is a broader umbrella encompassing group cooperative learning, group competitive learning, as well as peer tutoring. Hence peer tutoring that took many forms: 'Peer-Assisted Learning Strategies' (PALS), 'Cross-Age Tutoring' (CAT), 'Reciprocal Peer-Tutoring' (RPT), and etc., all entered the analysis.

Types of peer tutoring context. Only studies in which pupils had conducted peer tutoring in person, as opposed to virtual or online peer tutoring enabled via computerised programs. The logic for such choice was that virtual peers do not have the ability to show emotion such as praise, social and pedagogical skills, etc. Also, *peer counselling* or *peer education for health* were to be excluded, since the topic of such methods is not academic education. However, other forms of PAL, such as *peer modelling* or *peer feedback* with an academic interest while interacting face to face, were to be included due to their extensive similarities with peer tutoring.

Place of trial. Since peer tutoring had been proven to also work in extremely diverse cultures and countries, the studies could have been conducted in any country, as long as the studies were available in English. The disadvantage of the language criterion is that it too introduces a publication bias component in the findings.

Population characteristics. Studies conducted on students with learning disabilities were also excluded. There were two interrelated reasons for this: Firstly, this was due to the main aim of this paper, which concentrates only on the non-special education population. Also, since some students with learning disabilities require a mentor, the literature on mentoring and on this population is nearly as big as that of peer tutoring in the general population; therefore the task would have been unachievable within the time limit.

Participants' age*.* Although the search engine was set to include only ages from 4 to 18, studies with lower or higher ages still emerged. Therefore, further manual filtering was necessary. The reason for this age group was simply to concentrate on the elementary, secondary and high schools, which is the main interest of this research.

Area of concentration. Studies had to concentrate on one of the following areas: literacy, mathematics and science. This was mainly due to the reason that these are the core curriculum areas and expanding beyond them would have required more time.

Length. The studies had to be conducted over more than one session. Mainly for the reason that sometimes one session does not even suffice for training let alone for achieving any meaningful student interaction or positive outcomes.

In the end only 127 articles passed all the criteria, most of the articles which did not pass the inclusion criteria stage were those conducted on a special-education population and other topics.

3.5.3 Theoretical classifications and arrangements

This section identifies the benchmarks which were used when screening and evaluating the empirical work on peer tutoring. Each article was investigated according to the following 11 social interdependent benchmarks:

1.) Reward interdependence. In order for a study or peer tutoring condition to classify as implementing reward interdependence, *positively structured* tangible or non-tangible rewards need to be set either by the teacher or by the students themselves. The term *positively structured* is understood as the reward system being constructed in such a way that students within each group depend on each other in order to achieve the reward. The term positively structured acknowledges that there are many forms of structures in the peer tutoring settings.

2.) Goal-performance interdependence. Similar to reward-interdependence, the group needed to be *positively structured*, i.e. in such a way that in order for a student to achieve a performance goal there has to be cooperation rather than competition or individual work within each pair/small group. The difference between reward interdependence and goal interdependence is that the latter simply does not necessarily contain reward incentives to improve interdependence.

3.) Set-roles interdependence. The students needed to have set, non-reciprocal roles, unless the reciprocity took place after a period of 6 weeks, which is usually long enough for a role to become internalised.

4.) Interpersonal communication skills training interdependence. A paper which was coded as possessing interpersonal interdependence needed to have conducted training of students in social communication skills, or students being nice to one another.

5.) Informative guidance interdependence. In order for a paper to be coded as using guidance which concentrates on improving social interdependence, at least two of the following elements had to be included: a) reminding students of their *role,* 2) reminding students of their *goal*, 3) reminding students of their *reward*, 4) reminding students to use *positive communication skills*, 5) or reminding students to *praise*.

6.) Interdependence via praising. Although praising falls under "interpersonal communication skills training", praising here is treated as an element in its own right, since social interdependence places high emphasis on the need for the students to praise each other in order to improve motivation and interdependence. For a study to classify as using praise, the study had to mention that it had trained or encouraged students to praise each other while working in small groups.

7.) Cross-ability. A paper was coded as using cross-ability if it used students who ranged in ability from one another in the subject area within a small group; this did not exclude same-age students from being coded since they could also be cross-ability, as in the case of PALS intervention. By definition cross-age peer tutoring implies some level of cross-ability.

8.) Academic skills training. Papers coded as providing pedagogical skill training needed to illustrate that the students had been taught how to make use of the peer tutoring materials, when and how to ask questions, or give explanations and feedback.

9.) Advanced-deep/cognition. A study was coded as using elaborative or deep cognitive strategies if the materials contained any of the following: a) elaborative conceptual questioning, b) elaborative conceptual answers and explanations, c) predictive strategies, d) knowledge transformation, specifically applications of learned knowledge to real life.

10.) Meta-cognition. Studies are coded as using, meta-cognitive strategies if they meet one of the following: a) students use proposition questions, i.e. “how shall we approach this exercise?”, or “what is the best way to answer?”, b) students summarise or review their learned knowledge, c) students were asked to link new concepts to previously learned concepts or to categorise ideas, d) students’ materials were presented in multi-difficulty levels, so that the students could monitor each other’s knowledge and self-knowledge levels while working.

11.) Academic guidance. If the guidance made references to at least two of the following: a) using flash cards, b) deep/elaborative cognitive strategies, c) meta-cognitive strategies, d) when and how to ask or answer questions, e) when to give feedback, f) when to praise. This indicator is appropriate since the lack of such guidance could result in misunderstanding of sequences and conflict within the group.

3.5.4 Findings

Table 3 presents the percentage of various theoretical elements established for each academic subject: literacy, mathematics and science. This was calculated by dividing the total number of occurrences of each theoretical element by the total number of peer tutoring interventions within each subject and multiplied by a hundred. Several articles contained multiple peer tutoring intervention types. The following were the findings:

Table 3 shows that most peer tutoring studies have been conducted in literacy (82 studies, 139 interventions), followed by mathematics (43 studies, 81 interventions), and then science (14 studies, 32 interventions). Frequent theoretical elements applied in peer tutoring were:

1) *Academic skill training:* Up to 93% of science studies trained their students in academic engagement skills, followed by 84% in literacy and 81% in mathematics

2) *Cross-ability*: In literacy 78% of studies contained cross-ability groups, for mathematics 55% and science 48%.

3) *Role interdependence*: in literacy 63% of studies contained fixed roles, in mathematics 48%, in science 45%.

Table 3. *Percentage of theoretical elements for each subject*

Theoretical Elements	Subjects %		
	Literacy *82 Studies** *139 Interventions*	*Mathematics* *43Studies** *81Interventions*	*Science* *14 Studies** *32 Interventions*
1. Reward Incentives	29	26	10
2. Goal Setting	30	29	10
3. Fixed Roles	63	48	45
4. Cross-Ability	78	55	48
5. Social Script	31	19	0
6. Academic Script	63	29	48
7. Praise	48	39	24
8. Academic Training	84	81	93
9. Social Skills Training	14	13	55
10. Elaborative Cognition	26	28	55
11. Meta-Cognition	24	19	38

**Total of 127 articles, 10 articles concentrated on either two or three academic areas/studies. Totalling 252 peer tutoring conditions.*

Figures 3, 4 and 5 show the number of peer tutoring interventions by the number of theoretical elements for each academic subject:

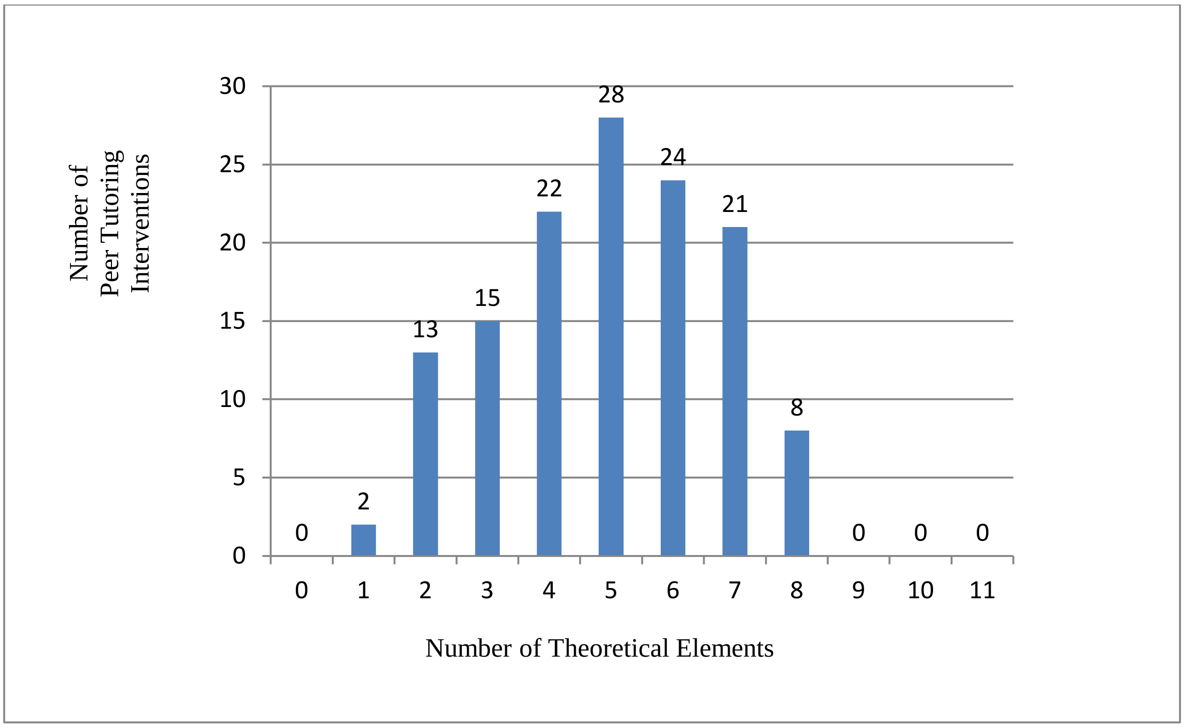


Figure 3. *Theoretical scopes in literacy*

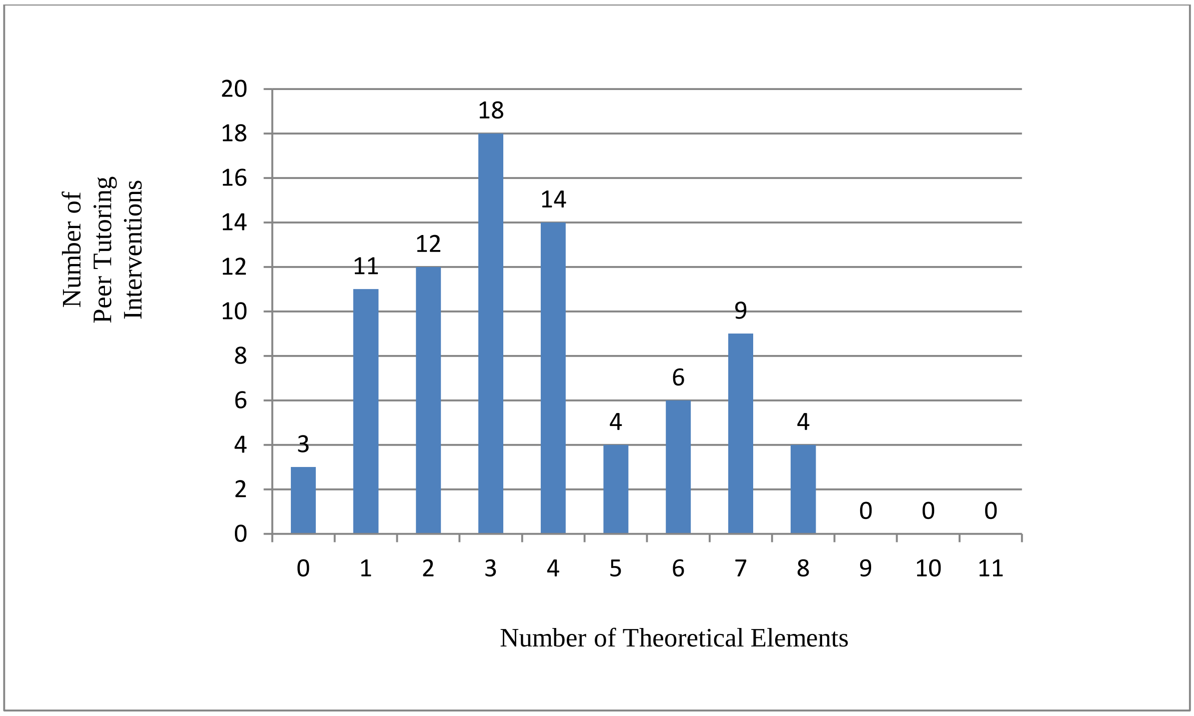


Figure 4. *Theoretical scopes in mathematics*

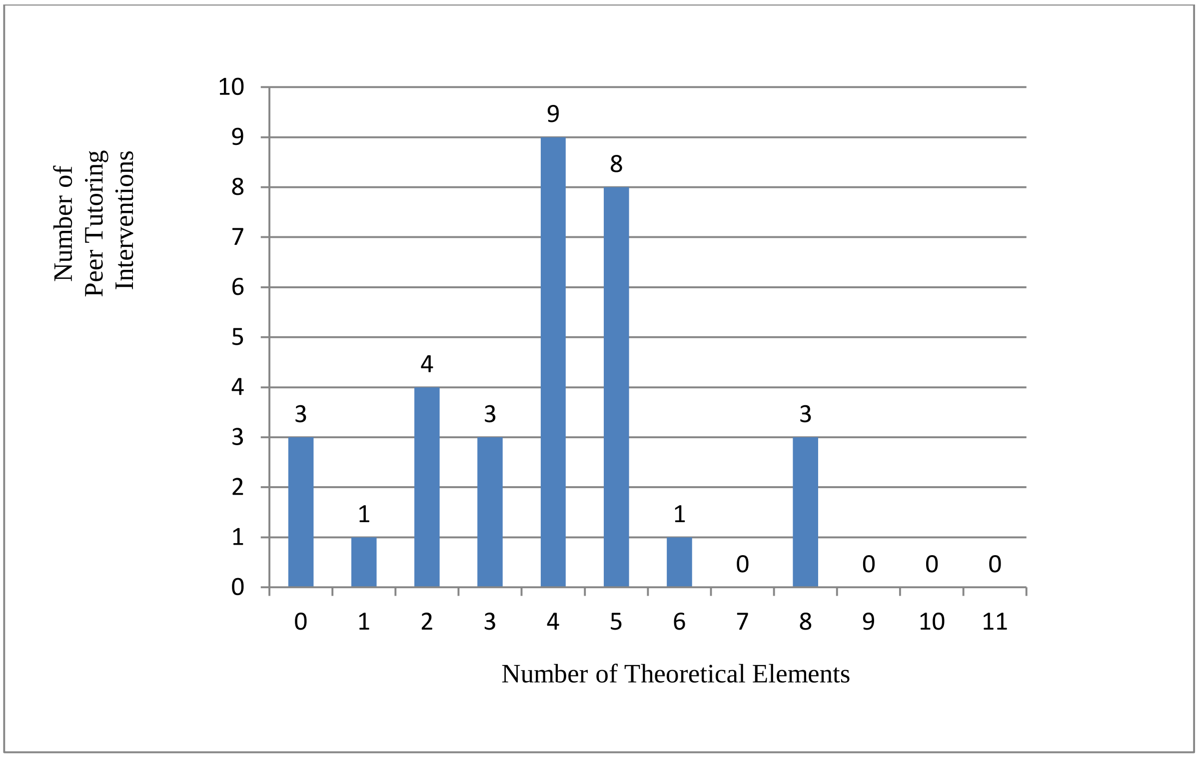


Figure 5. *Theoretical scopes in science*

The above figures show that in most peer tutoring interventions, in all academic subjects, most interventions do not apply more than up to 8 theoretical elements, most range between 3 and 7 elements. Only few studies belonged to the upper category of 7-11 theoretical elements, and within this category there were no studies which included more than 9 theoretical elements in any of the main academic subjects. This is the first and clear indication that there is a gap for peer tutoring interventions to take a broader theoretical applicability when implemented.

Table 4 on the next page shows the mean number of theoretical element for peer tutoring interventions within each subject, showing the total mean to be less than the median, 5.5.

Table 4. *Descriptive statistics of peer tutoring theoretical elements by subject*

	Mean	*Std.*	*n**
Literacy	4.92	1.732	139
Mathematics	3.65	2.169	81
Science	3.88	2.091	32
Total	4.381	2.013	252

*n=number of peer tutoring interventions, not studies.

3.5.5 Discussion

The overall total mean based on 11 theoretical elements used in academic subjects was 4.38 (Std. 2.01). This finding gives us a stance on the extent to which peer tutoring interventions have taken a broader theoretical approach when implemented, identifying the gap overall and within each academic subject. As is evidenced by Table 4, mathematics and science peer tutoring interventions have not had as many theoretical elements as literacy. This is not to be confused with the authors' actual theoretical stance, but rather what has taken place at the methodological/implementation level of peer tutoring[6]. Also, Tables 36, 37, and 38 of the raw data found in the appendix, show a strong pattern; i.e. when there have been set roles in peer tutoring (cross-age tutoring being a set role) there has not been any goal interdependence, rewards, or any other forms of social script/guidance interdependence to motivate the students. This further justifies the need for this particular research.

The findings from this broad review show that regardless of the positive social and academic effects of social interdependent theoretical elements shown by the meta-analyses cited above, when it comes to peer tutoring most researchers have been lacking behind in marrying up the most important elements of Social Interdependence Theory i.e. goal, reward,

[6] One reason why statistical tests were not conducted on the results was that the group sizes were very different to one another.

or informative structures (social scripts) which aid student autonomy, Table 3. This, however, does not imply that they are unaware of the theory, it simply means that when they have implemented peer tutoring, for various reasons, they have not been able to incorporate many social interdependence elements.

There is clearly a need to take a wider theoretical perspective of social interdependence when implementing peer tutoring in each of the academic areas, especially when considering that most meta-analyses investigating peer tutoring have shown to provide high effect sizes for social interdependence/cooperative elements. Tables 36, 37 and 38 in the appendix provide detailed information as to which studies and interventions employ each theoretical elements/benchmarks.

3.6 Systematic review of peer tutoring in mathematics

The what works literature has established that not only does peer tutoring rank at the top of the table when compared to other educational interventions, even within the peer learning and cooperative learning literature, peer tutoring performs positively. Consequently, an important question for policy makers, researchers and teachers, would be 'what works within peer tutoring in order to raise students' performance?', a question investigated in this section.

This review builds on that discussed in the previous section, specifically the theoretical empirical evaluation. The systematic review that follows is different from the one just conducted in the sense that it is more specific, it investigates peer tutoring only in mathematics and looks at effect sizes. The following sub-sections provide an outline of the inclusion criteria, how the effect size was calculated, the findings and a discussion:

3.6.1 Inclusion Criteria

Since the aim of this part was to investigate the impact of different types of peer tutoring interventions in mathematics on performance and attitude variables, the following 5 inclusion criteria were added on top of the criteria used previously for the methodical peer tutoring evaluation:

1) Studies had to be in mathematics.
2) Studies had to be of a quantitative nature.
3) Studies needed to have a single group, quasi-experimental design, matched design, or RCT.
4) Studies required data reported in a manner that allowed the calculation of an *effect size.* Some of the details the studies needed to report were the following: error (variation), mean pre/post test scores, standard deviation, T value or the F value.
5) Studies needed to have more than 10 participants per group, as very small sample sizes would bias the effect size (Slavin, Lake, Groff, 2009).

Out of 43 articles in mathematics from the previous review, 11 studies passed the extra criteria, two studies engaged in both same-age and same-age reciprocal (Menesses & Frank 2009, John, Fantuzo, King & Heller, 1992), and another engaged in same-age and cross-age (Tymms, et al., 2011): Most studies were excluded due to their qualitative nature and not providing the necessary data to calculate the effect size. The number of studies here is lower than that of 18 from Othman (1996) which compared peer tutoring in mathematics to other interventions, the reason being that the inclusion criteria are stricter in this review, specifically this review excludes unpublished or thesis papers, and does not include students with disabilities.

3.6.2 Calculating the effect size

Studies, which compared the control group to more than one form of peer tutoring method were considered to have their own effect size.

For each type of outcome measure, experimental group or sample within each paper, a standardised mean difference effect size was computed. This was done by dividing the group mean difference to a pooled estimation of *standard deviation.* Pooled estimation was used, as opposed to using only the *standard deviation* of the control group, due to it being reported to give a more robust measure (Coe, 2002). This is commonly known as *Hedge's g* (Hedge, 1981). Below are the formulas used for calculating the effect sizes:

$$g = \frac{\bar{x}_1 - \bar{x}_2}{s^*}$$

Equation 1. *Hedge's g*

$$s^* = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Equation 2. *Pooled standard deviation*

Alternatively *Cohen's d* was used for studies not reporting the necessary information. The procedures are similar to those above of *Hedge's g*. See below for more information:

$$d = \frac{\bar{x}_t - \bar{x}_c}{\sqrt{MSE\left(\frac{n_t + n_c - 2}{n_t + n_c}\right)}}$$

Equation 3. Cohen's d (using errors)

$$d \approx \frac{2t}{\sqrt{n-2}}$$

Equation 4. Cohen's d (to calculate Std. from t-values)

There were 59 effect sizes in total, including mathematics attainment and attitude, Tables 6, 7, and 8.

3.6.3 Findings

Tables 6, 7 and 8 provide descriptive information on the type of effect size for each study, specifically they provide the name of the studies, the length, instruments, descriptive data and effect sizes calculated via the formulas provided above.

Table 5 reports that for academic achievement, cross-age peer tutoring shows the highest mean effect size 0.59, especially when comparing it to same-age peer tutoring, mean effect size 0.47. For academic self-concept and social self-concept, the highest effect sizes were for same-age reciprocal peer tutoring: mean effect size 0.94 and 0.97 respectively; followed by cross-age peer tutoring, 0.19 and 0.71, and then same-age peer tutoring, 0.02 and -0.55.

Table 5. *Mean effect size by peer tutoring type*

	Performance			*Academic Self-concept*			*Social Self-concept*			*Total*		
	Mean Effect Size	Std.	n	Mean Effect Size	Std.	n	Mean Effect Size	Std.	n	Mean Effect Size	Std.	n
Cross-age	0.59	.34	9	0.19	.31	4	0.71	.55	2	0.48	.39	16
Same-age reciprocal	0.57	.54	21	0.94	.27	6	0.97	1.1	5	0.70	.62	32
Same-age	0.47	.47	7	-0.00	.02	3	-0.55	-	1	0.25	.50	11
Total Mean Effect Size	0.56	.48	37	0.47	.49	13	0.71	.99	8	0.56	.57	59

The total mean effect size for performance outcome was 0.56, for academic self-concept 0.47, and for social self-concept 0.71.

Table 6. *Same-age reciprocal peer tutoring*

Studies	*Length*	*Instruments*	*Descriptive statistics*	*Effect Sizes*
Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., Karns, K., & Dutka, S. (1997).	18 Weeks	Researcher Developed	LA – PMI Elaborated condition N=10 PMI Elaborated+conceptual N=10.	
			AAPMI Elaborated N=10 – PMI Elaborated+conceptual N=10	
			HA- PMI Elaborated N=10 – PMI Elaborated+conceptual N=10.	
			The control groups: LA N=20, AA N=20 and HA N=20.	
			For the first sub-test Operations:	
			PMI Elaborated	
			LA experimental Mean 18.8 (Std.=4.98), control Mean 14.30 (Std.=6.82).	**0.72**
			AA experimental M22 (Std.=8.22), control Mean 20.4 (Std.=7.03).	**0.22**
			HA experimental Mean 27.10 (Std.=7.54), control Mean 25.65 (Std.=7.99).	**0.18**
			PMI Elaborated+Conceptual	
			LA experimental Mean 19.10 (Std.=6.29), control Mean 14.30 (Std.= 6.82).	**0.72**
			AA experimental Mean 22.40 (Std.=7.78), control Mean 20.40 (Std.= 7.03).	**0.27**
			HA experimental Mean 27.90 (Std.=8.72), control Mean 25.65 (Std.=7.99).	**0.43**
			For the second sub-test 'concepts/application':	
			PMI Elaborated :	
			LA experimental Mean 27.30 (Std.=7.99), control Mean 21.95 (Std.=7.92).	**0.67**
			AA experimental Mean 33.10 (Std.=11.47), control Mean 33.35 (Std.= 8.90).	**-0.03**
			HA experimental Mean 42.00 (Std.=7.98), control Mean 42.50 (Std.=9.89).	**-0.05**
			- PMI Elaborated+Conceptual:	
			LA experimental Mean 26.80 (Std.=7.45) control Mean 33.35 (Std.=8.90).	**0.62**
			AA experimental Mean 34.10 (Std.=13.78), control Mean 33.35 (Std.= 8.90).	**0.07**
			HA experimental Mean 43.30 (Std.=9.41), control Mean 42.50 (Std.=9.89).	**0.08**

Ginsburg-Block. M., & Fantuzzo, J. (1997).	10 Weeks	Researcher modified Achievement test and Self –Perception for Children, Harter 1984.	Both experimental and control N=20 each. For achievement: experimental Mean 129.2 (Std.=48.7), control Mean 91.7, (Std. 55.92). For Self Concept: experimental Mean 3.03 (Std.=.39), control Mean 2.5 (Std. .45).	**0.72** **1.20**
Ginsburg-Block, M. D., & Fantuzzo, J. W. (1998).	7 Weeks	Performance (Researcher) Enjoyment Inventory from Gottfried 1999 & Self-Perception Profile for Children from Harter 1985.	All N=26 on achievement. All N=24 on self-concepts. ***Peer collaboration:*** *On achievement:* Computation experimental Mean 3.39 (Std.=1.01), control Mean 2.62 (Std.=.96). Word problem Mean .69 (Std.=.29), control Mean .61 (Std.=.38). *Self-concepts:* On motivation experimental Mean 48.90 (Std.=5.31), control Mean 44.41 (Std.=5.53). On academic self-perception Mean 2.94 (Std.=.71), control Mean 2.62 (Std.=.50). On social self-perception experimental Mean 2.67 (Std.=.42), on control Mean 2.27 (Std.=.49). ***Peer collaboration plus problem solving:*** *On achievement:* Computation experimental Mean 4.27 (Std.=1.31), control Mean 2.62 (Std.=.96). Word problem Mean .90 (Std.=.36), control Mean .61 (Std.=.38). *Self-concepts:* On motivation experimental Mean 51.44 (Std.=7.91), control Mean 44.41 (Std.=5.53). On academic self-perception Mean 3.32 (Std.=.62), control Mean 2.62 (Std.=.50).	**0.79** **0.24** **0.83** **0.52** **0.87** **1.10** **0.78** **1.0** **1.2**

			On social self-perception experimental Mean 3.55 (Std.=.46), on control Mean 2.27 (Std.=.49).	**2.6**
John, W., Fantuzzo, J., King, A., & Heller, L. R. (1992).	5 Months	Enright Diagnostic Inventory of Basic Arithmetic Skills from Enright, 1983	All N=16, with the exception of reward only self-perceived competence , N=14	
			Structure Only:	
			On Achievement, experimental Mean 4.5 (Std.=1.7), control Mean 5 (Std.=1.9).	**-0.27**
			On Self-Perception:	
		Self-Perception Profile for Children from Harter 1985	Scholastic Competence experimental Mean 2.8 (Std.=.3), control Mean 2.5 (Std.=.4).	**0.85**
			Social Acceptance control Mean 2.9 (Std.=.8), control Mean 3.1 (Std.=.6).	**-**
			Reward and Structure:	**0.081.50**
			On Achievement, experimental Mean 7.7 (Std.=1.5), control Mean 5 (Std.=1.9).	**1.10**
			On Self-Perception: Scholastic Competence experimental Mean 3.1 (Std.=.6), control Mean 2.5 (Std.=.4).	**0.18**
			Social Acceptance control Mean 3.2 (Std.=.5), control Mean 3.1 (Std.=.6).	
Fantuzzo, J. W., Polite, K.., & Grayson, N. (1990).	3 Weeks	Researcher	Experimental, N=12, Mean 3.7, control, N=5, Mean 3.5, t(15)=.23, giving a pooled Std.=.13.	**1.5**
			Experimental N=15. Control N=16.	**1.6**
Menesses& Frank. (2009).	5 Weeks	Curriculum Based Measurement Chosen Items.	Experimental, Mean 37.33 Std. 16.33 Control, Mean 17.50 Std.=6.95	

Table 7. *Same-age peer tutoring*

Studies	*Length*	*Instruments*	*Descriptive statistics*	*Effect Sizes*
Phelps, E., & and Damon, W. (1989).	2 year	Researcher Modified	The experimental group N=31, control N=41. For the experimental group Mean .44, for the control Mean .41, Mse=.02.	**0.43**
Krol , Janssen , Veenman & Linden (2004)	2 sessions	Researcher made instrument	Experimental N=12. Control N=8. Experimental Mean 34.83,Std.=4.53. Control Mean 31.13, Std.=5.14	**0.77**
Menesses & Frank. (2009).	5 Weeks	Curriculum Based Measurement Items	Tutor and tutee, each N=14. Control N=16. Control, Mean 17.50 Std.=6.95. Tutor Mean 25.71, Std.=12.94 Tutee Mean 34.07 Std.=19.05	**0.81** **1.1**
Tymms, P., et al., (2011).	2 years	PIPS (measures performance and attitude)	N=93 Schools for the younger students. N=119 Schools for the older students. (total 5179 students). The effect sizes were calculated by the authors via a similar method to that of Hedge's g:	
			Performance Younger Students	**-0.10**
			Performance Older Students	**0.00**
			Attitude Young Student	**0.01**
			Attitude Older Students	**-0.02**
John, W., Fantuzzo, J., King, A., & Heller, L. R. (1992).	5 Months	Enright Diagnostic Inventory of Basic Arithmetic Skills from Enright, 1983	All N=16, with the exception of reward only self-perceived competence, N=14	
			On Achievement, experimental Mean 5.4 (Std.=2.3), control Mean 5 (Std.=1.9).	**0.2**
		Self-Perception Profile for Children from Harter 1985	On Self-Perception: Scholastic Competence experimental Mean 2.5 (Std.=.5), control Mean 2.5 (Std.=.4).	**0**
			Social Acceptance control Mean 2.8 (Std.=.5), control Mean 3.1 (Std.=.6).	**-0.55**

Table 8. *Cross-age peer tutoring*

Studies	*Length*	*Instruments*	*Descriptive statistics*	*Effect Sizes*
Bar-El (1982)	4 Months	Form B of the Standard Arithmetic Test. And Teacher assessment Self-Concept Questionnaire Levine and Katz	All N 15	
			Standard Test	
			Tutee Mean 76.46 (Std.=10.77) Control Mean 65.88 (Std.=10.5)	**0.99**
			Tutor Mean 72.94 (Std.=6.67) Control Mean 63.33 (Std.=10.47)	**0.32**
			Teacher Test	
			Tutee Mean 7.22 (Std.=1.05) Control 6.84 (Std.=1.28)	**1.00**
			Tutor Mean 6.23 (Std.=1.22) Control Mean 5.51 (Std.=.64)	**0.74**
			Academic Self Concept	
			Tutor Mean 56.94 (Std.=70.3) Control Mean 58.12 (3.95)	**-0.20**
Fitz-Gibbon, C.T. (1990).	3 Weeks	Standard Progressive Matrices (SPM) test, Raven 1958.	***Experiment 1***	
			Tutors N=40, control tutors N=26. t=3.02. Tutors mean 34.6. Control, mean 34.	**0.79**
			Tutees' retention test (calculated by the author)	**0.80**
			Experiment 2	
			The experimental N=25, Mean 46 (Std.=8.6), control group N=30, Mean 43 (Std.=13.90).	**0.25**
Topping. K, Campbell. D., Walter, A., & Smith, A.. (2003).	5 Weeks	Me-As-Learner-Scale (MALS) Burden 1998 (Academic Attitude).	N=14,	
			Pre-test Mean 3.625 (Std.= 0.672), post-test Mean 3.971 (Std.=0.675).	**0.52**
		'Behaviour Indicator Of Self Esteem' (BIOS) Burnett 1998 (Social Attitude).	Pre-test Mean 3.32 (Std.= 0.87) to post-test Mean 3.60 (Std.= 0.86)	**0.32**
			Pre-test Mean 2.35 (Std.= 0.29) to post-test Mean 2.49	**0.48**

			(Std.=0.31)	
		‘Feelings’ questionnaire’ (academic attitude) (Researcher) What do you Think?’ (Researcher). (Social Attitude)	Pre-test Mean 2.13 (Std.= 0.22) post-test Mean 2.36 (Std.= 0.21)	**1.1**
Tymms, P., et al., (2011).	2 years	PIPS (measures performance and attitude)	N=93 Schools for the younger students. N=119 Schools for the older students (total 5179 student). The ESs are calculated by the authors via a similar method to that of Hedge’s d:	
			Performance Younger Students	**0.21**
			Performance Older Students	**0.22**
			Attitude Young Students	**0.02**
			Attitude Older Students	**0.12**

3.6.4 Discussion on the systematic review findings

There could be many reasons why same-age reciprocal tutoring shows a higher effect size than both cross-age peer tutoring and same-age normal on the attitude variables. Apart from the understandable benefit of having both students take both roles at some point, most reciprocal peer tutoring interventions also employ many social interdependent elements, such as goal interdependence, reward interdependence, and resource interdependence. In other words, apart from ensuring cross-ability within groups, a main component of Social Interdependence Theory, reciprocal peer tutoring interventions have also been able to implement positive interdependence within tutoring groups.

However, this systematic review did not investigate whether such differences were significantly different, due to the small and very unequal sample sizes, which leads to the p value being significant as a result of an increase of the between group error (Howell, 2010).

3.6.5 Chapter conclusion

This chapter has concentrated on the following four areas: a) providing a description of different Peer Assisted Learning methods, b) investigating the impact of peer tutoring from a social interdependence position on both processes of learning and academic performance outcomes by reviewing previous meta-analyses, c) conducting a methodical empirical review to evaluate the extent to which previous peer tutoring interventions had taken a broad social interdependent perspective when implementing peer tutoring, d) engaging in a systematic review of peer tutoring in mathematics.

The findings from these reviews have directly influenced the main aim and objectives of this research, which will be outlined in the next chapter. The following is a summary of the findings for the empirical review chapter:

***Peer assisted learning types*:** Peer assisted learning encompasses many small group interactions. Peer tutoring is one of the PAL strategies. It differs from other PAL strategies such as *peer education for health*, *peer counselling* and *peer monitoring*, in that peer tutoring concentrates mainly on the organised and structured learning of academic subjects and it involved a tutoring role (Topping & Ehly, 1998; Toping, 2007). There were four main peer tutoring strategies reviewed in this chapter: Cross-age peer tutoring, CWPT, RPT, and PALS (not to be confused with the overall term PAL) in the literature. The later three peer tutoring strategies share many similarities, and differ from cross-age tutoring since they are all conducted with same-age pairs/small-groups; all have some form of reciprocity between tutee and tutor, and all possess the main social interdependent components: cross-ability, goal, reward interdependence and resource interdependence.

Past reviews and meta-analyses. This chapter has reported several meta-analyses that have concentrated on the impact of peer tutoring in both process of learning elements as well as academic performance outcome. Social Interdependence has been the main explanatory theory, and through various meta-analyses which have also reported effect sizes for social interdependence variables, it has illustrated that peer learning influences both; a) process of learning outcomes such as interpersonal communication skills, behaviour towards peers and classroom actors and general self-attitude, and b) academic performance (Ginsburg-Block, et al., 2006; Leung 2014, Rohrbeck, et al., 2003; Roseth, et al., 2008). These meta-analyses have tried to provide some insight into important social interdependence elements such as reward, goal, structured interactions and autonomy from the teachers, specifically how these elements are associated with higher effect size in both process of learning as well as academic performance outcome. Other reviews and meta-analyses have implicitly looked at other elements associated with social interdependence, i.e. the learning process of mainly academic learning outcomes in the domains of psychological (academic self-concept, cognitive and meta-cognitive), linguistic (better academic communication skills), or better academic behaviour (Kohen, et al., 1982; Leung, Marsh, Craven, Yeung, & Abduljabbar, 2005; Roscoe & Chi, 2007).

The findings from various meta-analyses and systematic reviews all have something in common; they all show that peer tutoring interventions with social interdependent elements have a much higher effect size on social and academic results than peer tutoring interventions which did not do so.

Methodical literature review: Evaluating the extent to which past peer tutoring interventions had applied social interdependent elements. The aim of the extensive evaluation in this chapter was to assess and review the extent to which previous peer tutoring interventions had taken a broader social interdependent position when implementing the strategy. The review concentrated on three main academic subjects: literacy, mathematics and science, and presented three main findings:

a) Most peer tutoring interventions have taken place in literacy.

b) Theoretical elements such as *training in academic skills*, *using academic scripts*, *cross-ability* pairings or *fixed roles pairings* have had the highest applicability percentage on past peer tutoring interventions. The theoretical elements least applied to peer tutoring interventions were: *goal interdependence*, *reward interdependence*, *social script interdependence, training in interpersonal communication skills.* In other words, some of the elements which have shown to have the highest effect sizes in peer tutoring in terms of academic attainment as shown by the review of meta-analyses.

c) Most peer tutoring studies conducted in the three main academic areas, literacy, mathematics and science, have not taken a broader theoretical application at the implementation stage of peer tutoring, specifically most studies range between using 0-6 theoretical elements from the two dominant peer tutoring theories.

***Systematic review of peer tutoring in mathematics*.** Finally, cross-age tutoring in mathematics shows a higher effect size than normal same-age tutoring on all outcomes variables investigated here, such as mathematics performance, academic attitudes and social attitudes. Same-age reciprocal also shows higher effect sizes than normal peer tutoring on all variables. When comparing cross-age and same-age reciprocal, there was not much difference in terms of mathematic performance; however, reciprocal peer tutoring displayed a much higher effect size than cross-age peer tutoring on academic attitude variables, and slightly so on social self-concept. One explanation for the high effect size for same-age reciprocal peer tutoring on both attitude variables could be the fact that these interventions incorporate many social interdependence ideas, and therefore are more positively engaging than normal same-age or normal cross-age tutoring interventions.

4 Aims, Objectives, Research Questions and Significance

4.1 Introduction

This chapter concentrates on three inter-related areas: a) research aims/objectives, b) research questions and c) the significance of the study. This chapter has the following structure:

First, the link between research aims and research objectives is established.

The second part of the chapter looks at the link between the research objectives and research questions, identifying the research questions for each research objective and the hypotheses to be tested.

The final part explains the significance of the research in various areas, i.e. the significance of the research for mathematics, the importance of the study for teachers and students, as well as the importance for policy makers and academics.

4.2 Research aim and objectives

4.2.1 Research aim

The aim of educational research should be to test only ideas/theories that we are unsure of, rather than test what is already known (Tymms & Fitz-Gibbon, 2002). Social Interdependence Theory has proven to be a good explanatory framework in peer learning. Also, various meta-analyses in peer tutoring have shown that social interdependent elements provide the highest effect size. Yet the methodological review conducted here shows that most peer tutoring interventions do not apply social interdependence in a broader manner. As the empirical literature review illustrated, at the peer tutoring level most studies implement just a few of the theoretical elements from social interdependence, and those that do so have concentrated only on same-age peer tutoring.

Hence, even though we know that social interdependent elements have shown to provide high effect sizes as is the case for same-age reciprocal peer tutoring, it is not certain what impact social interdependent elements would have in a cross-age peer tutoring context. If cross-age tutoring interventions have shown high effect sizes, as have social interdependent elements, then the next step is to see if mixing these different elements will still provide an effective intervention. Consequently;

The primary aim of this research was to take a broader social interdependent position and test it in a cross-age peer tutoring setting**.** Specifically, the research aimed to do so by *developing, testing* and *interpreting* a peer tutoring intervention in a mathematical school context with year 6 to year 10 students. The secondary aim of the paper was to explore how differently achieving students worked under such conditions.

In other words, when it comes to peer tutoring among young students, one has to take into account many factors, such as: cross-ability, concentration on improving cognition and meta-cognition skills, providing pedagogical skills and academic structural engagement – goal interdependence, set role interdependence, social structural interdependence to manage social interaction, and interpersonal communication interdependence.

There is no doubt that individualized tutoring instruction designed to increase affective, cognitive and meta-cognitive elements can be beneficial (Harrison, 1976; Harrison & Cohen 1971), as are interdependent adjustment pairings (Johnson, D.W.,& Johnson, R.T., 1987), the classroom system (Fitz-Gibbon, 2000b; Royan & Deci, 2000), and individualised instructional training (Person & Graesser, 1999; Webb, 1991). However, what could also be beneficial is to test all of these elements simultaneously, and one theory which allows for that is Social Interdependence Theory. There is however a disadvantage to this approach, in the sense that the researcher is unable to pin down exactly what variable or element makes the highest contribution to the outcome, a disadvantage which is addressed via using randomised controlled trials that test straightforward ideas.

Hence, the argument made here is that in order to learn more about the effectiveness of Social Interdependence Theory we need to test it on many areas of peer learning, rather than only on groups of three and over, or one-to-one same-age contexts.

Figure 6, shows the learning processes through which ICAT, interdependent cross-age peer tutoring, would improve students' performance. One of the main limitations of the model below is that it does not distinguish between tutor and tutee learning processes, as is the case with most theoretical models in peer tutoring.

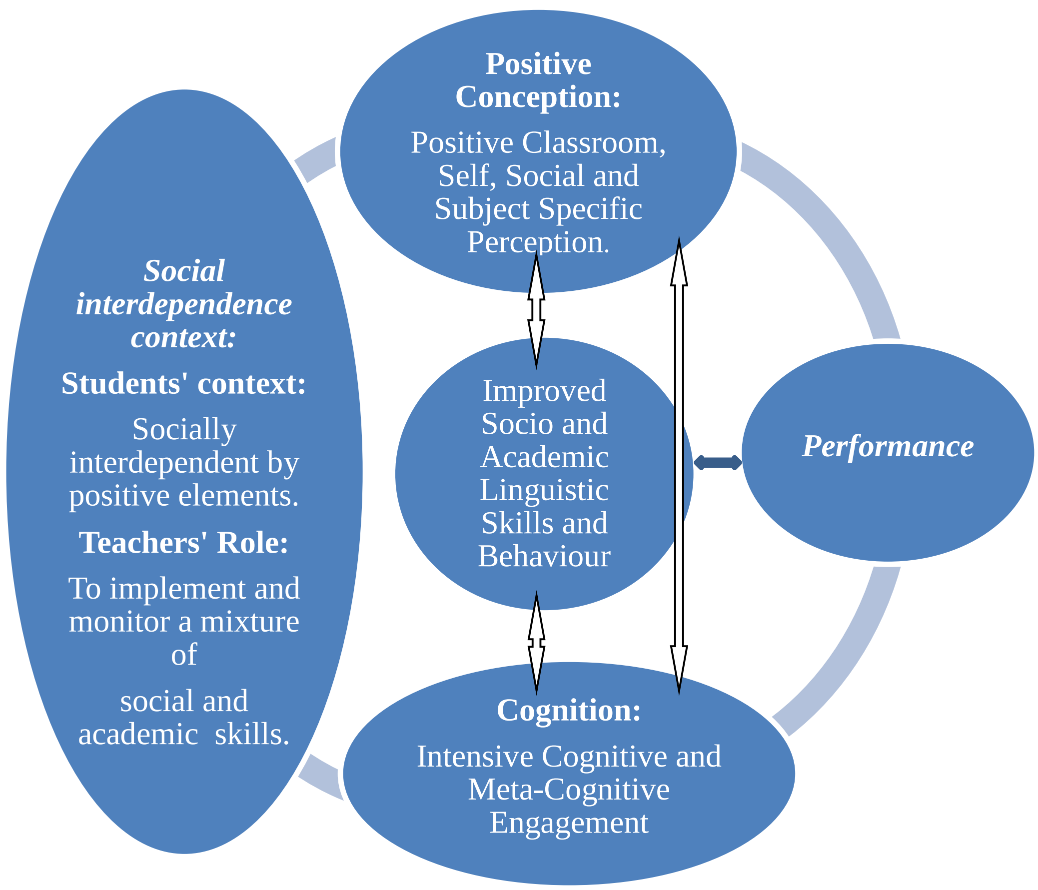


Figure 6. *Social interdependence learning framework for ICAT*

It should be made clear at this point that this study does not incorporate tangible reward interdependence. Rather, the reward the tutee and the tutor get in this intervention is that which comes with the role, i.e. for the older student feeling important and for the tutee having someone to look up to, or the feeling of accomplishment when achieving a goal which the students have chosen themselves.

4.2.2 Research Objectives

Most of the approaches in peer tutoring contain clear explanations as to the process through which peer tutoring enhances performance. However, most of the explanations present snapshots of the learning process in peer tutoring.

One way to understand the term *process* would be to consider not only how the *part* or the *whole* is constructed, but also how the *part* and the *whole* relate to one another, a task which can also be achieved by a Complexity Research Meta-Theory (Byrne, 1998, 2002, 2008; Cilliers, 1998, 2001; Elder-Vass, 2008; Harvey, 1992, 1996, 2002).

Therefore in order to investigate the wider theoretical understanding of the term "process" this study concentrates on the following objectives which are derived from the main aim:

A. Implementation fidelity:

1. To explore the extent of programme implementation fidelity at the school level, by investigating the extent to which peer tutoring lesson materials within each school correspond to programme specifications, as well as compare schools' level of implementation to one another. This investigation used the same process variables used at the group level analysis, with comparisons taking place between schools.

2. To observe the frequency of social interdependent behavioural variables across schools and see if they correspond to programme specifications, as well as to compare the schools to each other in terms of social interdependence behavioural variables.

B. Academic performance outcome:

3. To investigate the impact of a social interdependent peer tutoring framework on tutees' performance.

C. Social and academic psychological process of learning outcome:

4. To investigate the impact of social interdependent peer tutoring on the process of learning elements, such as tutors' and tutee's attitudes.

5. To conduct group level comparisons of tutees' lesson materials; in other words the objective was to explore process of learning elements, by comparing the lesson materials of those students who emerged as highest performing tutees with those of the lower performing tutees. The reason for such investigation was:

a) To shed more light on what social interdependent process variables were present in the higher performing tutee group; hence help understand what the important variables were.

b) Broadly speaking to understand better how different groups of tutees learned during ICAT. This type of analysis, did not seek to establish causation, explanation or prediction between certain independent variables and outcomes, the aim was simply to explore.

4.2.3 Justifying objectives

With the exception of the second objective, which investigates the impact of the peer tutoring intervention on performance, the remaining objectives explore the process of learning during ICAT or its implementation level. The following is the logic for such objectives:

Firstly, peer tutoring performance outcomes were measured in order to determine the size of the impact of the intervention, and see if the intervention actually improved mathematics performance as measured individually. Measuring performance outcomes was also necessary in order to compare the intervention with other similar and not-so-similar peer tutoring interventions.

Secondly, the process of examining learning variables, via analysing social and academic attitude questionnaires and completed/uncompleted peer tutoring lesson materials from the students would enable the exploration of beneficial process of learning variables in cross-age tutoring.

Finally, fidelity implementation data was important since not all aspects of an intervention are generally implemented accurately. Without fidelity analysis it is hard to be sure about the impact of an intervention **(**Arkoosh, et al., 2007; Bradshaw, Reinke, Brown, Bevans, & Leaf, 2008; Gresham, 1989). However, usually, high treatment implementation should produce a high performance outcome (Arkosh, et al., 2007; Noell, et al., 2005). One should add that this is the case if the intervention has been proven to be inherently effective.

Other broader and harder to measure important elements, which needed to be considered, were the *implementation*, *organization* and the *monitoring* of a peer tutoring intervention. As Fitz-Gibbon puts it: “learning need not be left to chance; it should be monitored” (1992, p258). In the study that follows implementation and organisation were conducted jointly with the teacher, and regular communication was held with the schools to monitor their experience.

4.3 Research questions

The following were the research questions and hypothesis for each objective**.** Some of the questions were very general at this stage and will be operationalized and further elaborated on in the method chapter.

Objective 1). To explore the extent of programme implementation at the school level, by investigating the extent to which peer tutoring lesson materials correspond to appropriately implemented interventions as well as compare implementation across the schools.

Research question. When comparing schools which implemented peer tutoring, were there any significant and substantive effect sizes on the following social interdependent process outcome variables:

1. Amount of peer tutoring lessons attended
2. Goals set by the students
3. Quantity of exercises attempted in the *practice test* section
4. Quantity of exercises attempted in the *connection* section
5. Quantity of exercises attempted in the *turn-taking* test
6. Quality of answers in the *practice test* section
7. Quality of answers in the *connection* section
8. Quality of answers in the *turn-taking* test
9. In terms of *total types of feedback*
10. In terms of feedback by ticks/crosses in the *practice test*
11. In terms of feedback by ticks/crosses in the *turn-taking* test
12. In terms of *checking if goal is achieved.*

When assessing the level of implementation according to programme specifications, the research question was: ‘to what extent did the schools implement the above points with the exception of feedback?’ As the level of feedback is hard to predict.

Objective 2). To observe the frequency of social interdependent behavioural variables across schools to further measure the extent of implementation.

Research question. Was there a significant difference and substantive effect size regarding the following observed social interdependent variables between schools?:

1. How often students set goals
2. The observed frequency of friendly body language and tone of voice
3. How often the tutor praised correctly
4. The number of observed meta-cognitive questions
5. The amount of time the tutee corrected him/her self
6. The number of times the tutee connected/categorised ideas
7. How often the tutor questioned the tutee
8. How often the tutee provided an answer
9. The number of times the tutee questioned
10. How often the tutor was observed providing an explanation
11. How often the students were observed engaging alone in the task
12. The amount of time a viewing was in-audible.

When assessing the level of implementation according to programme specifications, the research question was: 'to what extent did the schools implement the 12 variables above?'

Objective 3): To investigate the impact of a social interdependent peer tutoring framework on tutees' performance.

***Research question*:** Was there a statistically significant difference and substantive effect size for performance between the students undertaking peer tutoring and a comparison group?

Objective 4). To investigate the impact of ICAT on the social interdependent process of learning elements, such as tutors' and tutees' social and academic attitudes.

Research Question. Was there a statistically significant and substantive effect size between students taking peer tutoring and the comparison group on the following attitude variables?:

- Mathematic self-concept
- Enjoyment of mathematics
- Social self-concept
- Attitude towards others in mathematics classes
- Attitude towards having a choice in mathematics classes

Objective 5). To conduct group level comparisons of tutees' lesson materials to explore the extent to which social interdependent process variables were present in the higher performing tutee group in order to learn how different groups of students learned during ICAT.

Research question. When comparing the tutees who gained most from the peer tutoring (high performing group) to those who gained least, were there any significant and substantive effect sizes on the following social interdependent process outcome variables:

1. Amount of peer tutoring lessons attended
2. Goals set by the students
3. Quantity of exercises attempted in the *practice test* section
4. Quantity of exercises attempted in the *connection* section
5. Quantity of exercises attempted in the *turn-taking* test
6. Quality of answers in the *practice test* section
7. Quality of comments in the *connection* section, i.e. negative, broad or specific
8. Quality of answers in the *turn-taking* test
9. In terms of *total types of feedback*
10. In terms of feedback by ticks/crosses in the *practice test*

11. In terms of feedback by ticks/crosses in the *turn-taking* test
12. In terms of *checking if goal is achieved.*

4.4 Significance

This section looks at the study's significance. It does so by investigating the significance in three areas: in terms of the academic subject, students and teachers, policymakers and academics:

4.5 The significance of peer tutoring for mathematics as a subject

There is a need to increase academic performance in mathematics within the UK, as shown in the introduction chapter with the Key Stage 2, in relation to other core subjects mathematics is lagging behind. Considering the importance of mathematics in today's economy, it is only natural that a more substantial effort should be made to improve mathematic performance in the UK.

4.5.1 The significance of using ICAT for students and teachers

Firstly, ICAT would benefit the students. The main significance of applying a social interdependent approach to peer tutoring is that it goes to the heart of the governments' national strategy of 'Every Child Matters: Change for Children, 2004', regardless of whether the students' learning needs are social or academic. Cooperative learning strategies in the UK have not been tested to a sufficient degree (Jolliffe, 2007; 2011).

The dominant view in peer tutoring is that different peer tutoring interventions should apply to different classrooms, depending on the classroom needs, i.e. whether there is a need to improve students' social communication skills or educational communication skills (Yarrow & Topping 2001). However, this approach becomes problematic because in many classes there is a further subdivision taking place between those students who need urgent help with *social aspects* and those who need urgent *academic* help. Therefore, past peer tutoring interventions, with the exceptions of same-age reciprocal, have been unable to take into account the notion that classroom needs are complex, with many classrooms requiring social as well as academic help.

Regarding benefits for students, the literature review has already illustrated multiple advantages, especially in terms of process of learning and academic performance outcomes which indicate that students gain from peer tutoring or peer learning in general. Consequently, by applying a peer tutoring intervention shaped by Social Interdependence Theory, students should be able to gain in terms of social and academic attitudes as well as performance.

Also, interdependent cross-age tutoring would benefit the teachers. One of the reasons why some peer tutoring interventions are not fully successful may be the teachers' inability to manage the complexities involved and spend sufficient time and effort (Mahedy, 1998). However, enough incentives for the teachers, in terms of widening the benefits for a wider range of students, would make teachers more likely to implement peer tutoring fully. Three main benefits for the teachers can be outlined:

A) A wider theoretical perspective in peer tutoring would broaden teachers' knowledge of how and why different peer tutoring interventions work, enabling teachers to be innovative and create their own peer tutoring versions by picking and mixing.

B) A peer tutoring intervention shaped by leading peer learning theories would increase teachers' motivation and self-efficacy to apply peer tutoring in their classroom, since theoretically broad peer tutoring interventions would please most teachers by ticking the box on many process of learning mechanisms. Teachers are confused when their own goals for the tutoring intervention are not clear. Studies have shown that teachers' *expectancy of success*, or self-efficacy in being successful at implementing cooperative learning techniques, is important in determining whether they are going to make use of the method (Abrami, Poulsen & Chambers, 2004).

Some interventions concentrate purely on social and affective psychological outcomes and others on cognitive and meta-cognitive outcomes (Hoy & Tschannen-Moran, 1999). The benefits of a peer tutoring intervention shaped by broader perspectives are most evident when the intervention is implemented on a high scale, hence providing an incentive for multiple teachers' goals.

C) By concentrating on training and monitoring students in social interdependent elements which as mentioned also incorporate many social constructivist factors, the teachers improve their chances for a successful peer tutoring intervention, and professional development[7].

4.5.2 The significance for policy makers and academics

The significance for policy makers: The findings of this research are also aimed at influencing policy makers and funding bodies, providing them with structured and informative arguments,

[7] It is however very likely that most teachers simply take a pragmatic, rather than a theoretical view when implementing an intervention such as peer tutoring, and hence later forget why the method is powerful.

by testing a new intervention and exploring its process of learning elements[8], so that by applying cooperative ideas peer tutoring remains part of the 'what works literature'.

The significance for the academics: Although social interdependent ideas have already been applied to peer tutoring, as is the case with CWPT, PALS, or RPT, as identified they have not yet moved to the cross-age arena. The only study which incorporates interdependent elements to cross-age peer tutoring is that of DePaulo, Tang, Webb, Hoover and Litowitz (1989). The study however was different from the one conducted here in the following respects: a) the focus was a card-board drawing task b) it included only one session, c) the rewards and goals were imposed by the teachers/researchers, d) many other important interdependent elements such as praising, resource interdependence, training on social interdependent skills, were not evident, and e) the study was based on a laboratory setting format with students being recorded by audiotape. The study concluded that the tutors gained more than the tutees in learning.

Hence, a six week interdependent cross-age peer tutoring intervention in a core subject such as mathematics would be a new way of looking at the peer tutoring evolution and would initiate further discussions within the co-operative academic community.

4.6 Chapter conclusion

This chapter has engaged with the aim, objectives, research questions and the significance of the study. The following is a summary of each topic:

This work aims to apply social interdependent elements to cross-age peer tutoring in mathematics. The need for such aim is justified through and derived from a) the theoretical

[8] As mentioned in the acknowledgment the work is funded by the Economic and Social Research Council (ESRC), a body informing education policy in the UK. The data will therefore be made available to the ESRC.

review, b) the reviews on meta-analyses, c) the methodical empirical evaluation from a theoretical position, and d) a systematic literature review; all of which were elaborated in the previous chapter. The reviews helped to shape and justify the research aim taken here, for example:

- It is now clear what theory takes the broadest approach when explaining peer tutoring, i.e. Social Interdependence Theory.
- We have some idea what process of learning elements in peer tutoring have the highest effect sizes, those associated with Social Interdependence Theory.
- We have some understanding to what extent peer tutoring interventions have implemented the identified effective elements, i.e. most interventions implement only half of the effective theoretical elements of Social Interdependence Theory, and some of the most crucial elements, goal interdependence as set by students, reward interdependence as set by students, autonomous resource structure which aid students' interactions, have the lowest applicability in peer tutoring literature. These findings were noted across subjects, and those peer tutoring interventions which did so concentrated only on same-age peer tutoring.
- Since past systematic reviews have not explored the question of what works in peer tutoring mathematics, this review could assist in answering how cross-age, same-age and same-age reciprocal differ in terms of their mean effect sizes in students' performance, academic and social attitudes.

Different theories have come to different conclusions in terms of the benefits of peer tutoring on the process of learning outcomes. However, the problem is even more complicated; considering that every peer tutoring method is different, so are the socio-psychological processes through which they enhance performance (Yarrow & Topping, 2001). For example, same-age cross-ability is theoretically more beneficial to the tutor than the tutee, since the tutee being the lower performer during same-age tutoring in the class sometimes harms his/her self-esteem. Cross-age tutoring on the other hand does not acquire such degenerative elements (Fitz-Gibbon, 2000a) since the tutee is aware that the tutor knows more because he or she is older. Each tutoring strategy can be further sub-divided into different forms depending on who designed that particular strategy. Therefore generalization regarding the functioning of peer tutoring from one strategy to all tutoring strategies should be avoided (Yarrow & Topping, 2001).

However, this does not mean that the crucial elements of important theories could not be integrated in order to improve the explanatory power of academic performance outcomes, of process of learning outcomes, as well as the chances for the peer tutoring interventions to succeed; all of which would mostly benefit teachers and students. Therefore the kernel of the aim for this research was to test such broadened understanding and to interpret the findings from a social interdependent perspective.

In order to achieve this aim the study presents five main objectives, with each objective leading to a research question. All research objectives have the goal of testing and interpreting the new interdependent cross-age peer tutoring framework in mathematics, either by investigating: implementation fidelity in various ways (looking back at programme specifications or comparing implementation across schools) and through various methods (ICAT lesson analysis as well as observations), academic performance outcomes or process of learning outcomes (attitude gains and the comparison of higher performing and lower performing tutees' ICAT lesson materials).

As reported, the importance and significance of this study are multi-level and multi-purpose: a) the research introduces a redeveloped intervention for the improvement of mathematics as a subject in the UK; an area in which the UK is lacking behind internationally, and within England as a subject is not doing as well as the other subjects, b) the study findings should provide teachers and students in the English speaking world with more incentive, knowledge and success while using peer tutoring, c) the work aims to improve awareness of peer tutoring among policy makers, and d) hopes to open new academic discussions.

Although it is clear that a wider theoretical position in peer tutoring will provide more work on the teachers' side, if the teachers are able to deem a peer tutoring intervention useful, they will be more prone to accepting the intervention as they would perceive it to be successful, and hence be more willing to implement it (Abrami, et al., 2004).

PART II

5 Method

5.1 Introduction

This chapter outlines the research method. The paper takes a multi-method approach; specifically, it employs various data collection types and data analysis methods to evaluate and study ICAT.

Throughout this chapter it will be made evident how Social Interdependence Theory has shaped: a) the cross-age peer tutoring intervention, b) instruments/indicators, and c) data-collection and the data-analysis methods adopted here.

The section concentrates on the following areas: participants, school characteristics, intervention procedures and instruments used to collect information. The lengthiest part of this chapter will cover intervention procedures.

The intervention procedure section concentrates on: a) group creations, b) design and data collected, c) materials/exercises, d) ethical considerations, e) timetable, f) school assistance, g) pair and classroom set up, h) training and i) ICAT framework. The last four sub-headings concentrate on illustrating what a social interdependent cross-age tutoring intervention looks like.

5.2 Participants

The study included in total 841 participants from 5 different schools, out of which 306 students were from two schools in the North East of England, which only participated in testing and developing the attitude instrument; and 535 students from three different schools participated in the actual peer tutoring study. The three schools which accepted to take part in cross-age peer tutoring were located as follows: two in the North East of England and one in the South East. In the North East, one of the schools involved Year 8 (13.4 years old) tutoring Year 6 (11.7 years) which is the top end of primary school (School A), and the other Year 9 (14.2 years) tutoring Year 7 (12.3 years) (School B). The school in the South East involved Year 10 (15.4 years) tutoring Year 8 (13.2 years) (School C). Hence, the schools carried out cross-age peer tutoring with different age groups.

In terms of identifying the sample size required for the project, a power analysis through an online calculator by Professor Daniel Sopers was conducted. The analysis showed that for an expected effect size of 0.56, which was the mean effect size of peer tutoring in mathematics as shown by the systematic review in this paper, with a significance level of 0.05 (one tailed t-test), a power of 0.8, and two groups, the total sample requirement for each school needed to be 84 students, or 42 per each ***clustered*** group. In order to account for attrition rates on tests it was necessary for the clustered groups to include over 50 participants. However, for school B the numbers were smaller as there were not enough students available.

There were no exclusion criteria for the participants, all students were invited in the project.

The following were the number of participants for each school:

School A contained a total of 228 students, in which Year 8 tutors Year 6. The peer tutoring group contained 54 Year 8 and 54 Year 6 students, the control group contained 62 Year 8 and 58 Year 6 students, in total 8 classes.

For school B, in which Year 9 tutored Year 7, the total number of students was 153, again in total there were 8 classes. The peer tutoring group for Year 9 contained 36 students, as did the peer tutoring group for Year 7; control group Year 7 contained 39 students, control Year 9 contained 42 students.

Finally, the total number of participants for School C was 154, in which Year 10 contained 80 students and year 8 contained 74 students, also totalling 8 classes, however as will be shown further down, this school decided to drop the control group.

The schools for the project were selected as follows: emails reporting the opportunity for the project were sent from the school authorities of three different English County Councils, North Tyneside, Leeds, and Medway. Schools in Durham were contacted through the Centre for Evaluation and Monitoring (CEM) at Durham University. Over 70 schools were contacted. Nine schools expressed the wish to participate in the project giving a response rate of 13%. Out of the 9 schools three schools agreed to the initial terms and conditions in respect to the design and time frame.

5.3 Schools' characteristics

The last two Ofsted reports, 2008 and 2012, stated that schools A and C were both in the category of "good schools". The Ofsted report of 2008 for school B was "satisfactory school" and for the year 2012 was "good school". Schools B and C were both secondary schools, and school A taught Years 5-8. The national average of good primary and secondary schools in 2013 was 55.5%. One member of staff from school A had in the past received some training in cross-age peer tutoring in literacy, which she, however, did not previously make use of. In terms of socio economic background, of the students who participated in the project, school A consisted of 26.6% of its students receiving free school meals. School B was estimated to have up to 54.5% of its students in the project receiving free school meals, and school C, with 28.5% of the students.

School C contained slightly more boys, 55.1%, than girls, - for school A the percentage of boys was 50.6, and in school B 53%. School A included a higher percentage of students with English as a second language, 7.9%, than school B 0%, and school C 1.6%. In terms of special educational needs, school A consisted of 5.2% of their students with special educational needs, school B 11.3%, and school C 13.7%. For the ages on which peer tutoring was conducted, school C included the highest number of students with disabilities, with 16 students, school B with 7 students and school A, 2 students. Students with disabilities participated in the project, however they were removed from the analysis, since meta-analysis (Bowman-Perrot, Davis, Vannest, Williams, 2013) have found that students with disabilities benefit more from peer tutoring than normal population, which would have in return biased the overall findings. Finally, in terms of average Key Stage 2 points the schools were similar, with school A being slightly higher, 28.8 points, followed by school B, 28 points and then school C, 26.5 points.

Table 9. *School characteristics*

	Mean Age For Each Year					*School Age Range*	*Total Stud-ents*	*% Boys*	*% Girls*	*% SEN**	*% ESL**	*% Free School Meal*	*Average KS2 Point*
Schools	***6***	***7***	***8***	***9***	***10***								
A	11.7		13.4			9 -13	478	50.6	49.4	5.2	7.9	26.6	28.8
B		12.3		14.2		11 – 16	415	53	47	11.3	0	54.5	28
C			13.2		15.4	11 – 19	1301	55.1	44.9	13.7	1.6	28.5	26.5
National Average[9]							978	51.0	49.0	7.7	15.9	16.8	28.4

**SEN=special educational needs. ESL=English as second language*

[9] Data gained from Ofsted School Data Dashboard

5.4 Intervention procedures

5.4.1 Group creation and group characteristics

All the groups were chosen by the mathematic heads of each school. The selection of groups was influenced by the timetable and practicality within the schools. Therefore, there was insufficient control for the researcher to make sure that the groups were equal in terms of performance, age, and ethnicity. However the schools were advised to at least ensure that the groups were balanced in terms of the number of younger students' free school meals on each group, which they did.

Table 10 below shows the gender composition for each group/age within each school.

Table 10. *Gender composition for each group by school and year*

School A				*School B*				*School C*			
Year 6	Control	Girls	24	Year 7	Control	Girls	22	Peer Tutoring	Year 8	Girls	31
		Boys	34			Boys	17			Boys	43
	Peer Tutoring	Girls	22		Peer Tutoring	Girls	23				
		Boys	32			Boys	13				
Year 8	Control	Girls	41	Year 9	Control	Girls	18		Year 10	Girls	36
		Boys	21			Boys	24			Boys	44
	Peer Tutoring	Girls	25		Peer Tutoring	Girls	15				
		Boys	29			Boys	21				

All schools managed to have similar numbers of students receiving free school meals in each group, control and treatment; this was only for the tutees, namely the younger students. In terms of free school meal, School A Year 6 students for the control group consisted of 8 students, peer tutoring group contained 6 students. For school B, Year 7 students, in the control group there were 17 students with free school meal, and in the peer tutoring group 15 students. Finally, for school C year 8, there were a total of 19 students on free school meals.

5.4.2 Design and data collected

This was a *multi-methods* research. Table 11 below provides a summary of the overall design and data collected from each school and table 12 provides the pre-test scores.

During the initial discussions with the schools an agreement was reached that the design would be a two stage pre-post-test quasi-experimental, with the second stage commencing after the summer holiday and the classes switching their conditions.

As can be seen from table 11, School C decided not to include a control group and all schools, for various reasons, dropped their second stage[10]. In terms of performance data, the initial aim of the research was not to include the data for the tutors, since the control group for the tutors was not going to cover the same mathematics materials as the peer tutoring group. However, for school B, performance data for the tutors were also collected for two reasons: Firstly, the teachers expressed a strong wish to do so, and secondly the teachers expressed a concern that the tutors' absolute mathematics ability did not differ much to that of the tutees; therefore it was necessary to see how the tutors compared to the tutees on performance for this school. Also, it has to be pointed out that the post-test performance and attitude data for the tutors in school B took place two months later than expected.

Finally, in terms of lesson scripts, only two out of four classes from school A contained the students' names on the lesson scripts. This was due to a communication breakdown within the school.

[10] The reasons are discussed in the limitation chapter.

Table 11. *Design and data collected*

Schools	*Length & control groups.*	*Performance*	*Attitude Data*	*Observation*	*Lesson scripts for school level peer tutoring implementation for 6 weeks*
A	Six weeks, pre-post-test quasi experimental design.	Pre-post for all Year 6s.	Pre-post, for all Year 6s and Year 8s.	Two classes.	Two classes out of four.
B	Six weeks, pre/post-test quasi experimental design.	Pre-post for all Year 9s and Year 7.	Pre-post, for all Year 9s and Year 7s.	Two classes.	Yes.
C	Six weeks, pre-post-test design.	Pre-post for the Year 8s.	Pre-post, for all Year 10s and Year 8s.	Two classes.	Yes.

Regarding group difference, as can be seen from table 12 next page some of the variables were not equal on the pre-test results, especially Year 6 student's from school A and Year 9 student's from school B ($p<.05$). There were also differences on pre-test scores on some of the attitude variables, especially for mathematics self-concept Years 6, 8 and 9 student's, as well as variable choice for Years 8 and 9 student's ($p<.05$).

Table 12. *Pre-test score differences by variables and schools*

Student Performance	***Pre-test mean peer tutoring***	*Std.*	*n*	*Pre-test mean control*	*Std.*	*n*	*Sig two tailed*
Year 6 (Tutee) School A	20.02	5.94	46	23.51	5.4	49	.004
Year 7 (Tutee) School B	11.54	5.09	35	13.69	5.49	26	ns
Year 9 (Tutor) School B	17.04	5.18	25	12.82	7.86	17	.042
Student Attitude							
School A Year 6 (Tutees)							
Mathematics self-concept	17.91	5.02	46	19.81	4.09	48	.047
Mathematics enjoyment	11.85	4.89	46	13.60	4.79	48	ns
Relating to people in mathematics classes	19.15	5.61	46	20.29	4.25	48	ns
Social self-concept	14.41	4.66	46	14.35	3.84	48	ns
Choice	11.20	4.6	46	10.83	4.60	48	ns
School B Year 7 (Tutees)							
Mathematics self-concept	16.65	5.37	34	16.36	5.96	28	ns
Mathematics enjoyment	10.21	4.73	34	10.42	5.55	28	ns
Relating to people in Mathematics classes	17.62	4.58	34	19.04	5.44	28	ns
Social self-concept	13.09	3.5	34	14.04	3.99	28	ns
Choice	12.94	4.19	34	12.14	4.61	28	ns
School A Year 8 (Tutors)							
Mathematics self-concept	19.65	3.8	48	17.75	4.13	59	.016
Mathematics enjoyment	13.58	4.88	48	12.32	3.80	59	ns
Relating to people in mathematics classes	19.25	3.91	48	18.64	5.06	59	ns
Social self-concept	13.81	4.12	48	14.54	4.33	59	ns
Choice	12.77	2.77	48	10.67	3.99	59	.003
School B Year 9 (Tutors)							
Mathematics self-concept	18.00	3.24	28	16.20	2.46	25	.029
Mathematics enjoyment	11.14	5.13	28	9.28	3.93	25	ns
Relating to people in mathematics classes	18.17	4.97	28	14.28	3.72	25	.002
Social self-concept	14.16	3.82	28	13.96	3.95	25	ns
Choice	13.04	4.30	28	10.88	3.10	25	.043

5.4.3 Materials/exercises

The topics on which the students worked were chosen by the schools in order to reflect their lesson plans and so that the control group tutees also concentrated on the same topics. The exercises were chosen and created by the mathematics teachers from each school, together with the researcher and supervisors' help. Many of the exercises were influenced by MathsLinks, 1, year 7 Practice Book (Ray Allan, 2008); since it provided a good illustration of how to arrange mathematics exercises in different complexity levels. The following in table 13 were the topics covered by each school:

Table 13. *Lesson topics by schools*

School A	*School B*	*School C*
1. Number patterns and sequence	1. Mean, median, range, mode	1. Measurements
2. Fractions	2. Data interpretation and representation	2. Probability
3. Understanding measures	3. Factors, multiply, fractions	3. Transformation
4. Properties of shape	4. Sequences	4. Enlargements
5. Data interpretation and representation	5. Mental methods, multiply/divide	5. Area
6. Written methods	6. Equations	6. Equations

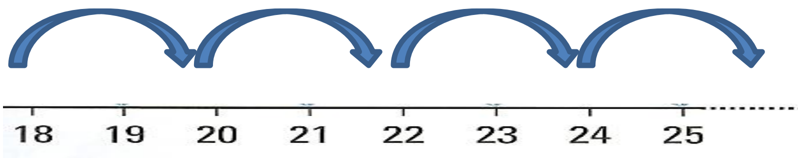
The decision to allow the teachers to choose the topics and have input into the exercises was intended to have the following benefits: Firstly, to enhance the possibility of producing exercises at the appropriate level of ability within each school. Secondly, to make teachers feel part of the project, so that in return they would give their best effort in implementing the project, and consequently reduce dropout rates. Finally, due to the complexity of the task it would have been extremely time consuming for one researcher to design all the exercises at the correct levels of student ability.

The following were some of the exercises covered by the students on ‘number patterns and sequence’ with Year 6 tutees, school A:

1. Find an even number in this sequence: 1, 2, 3, 5, 7, 9, 11,

2. Find an odd number in this sequence: 2, 4, 6, 7, 8, 10, 12,

3. Write the sequence shown below: ……………………………..



4. What should the next number be on the line above? ……..

5. This is a sequence of numbers starting with **100**. If we **subtract 45** each time. Write the next **two** numbers in the sequence.

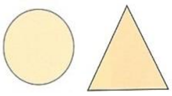
100, 55, 10, ☐ ☐

6. What is the next number in this sequence? 10, 7, 4 , 1, …….

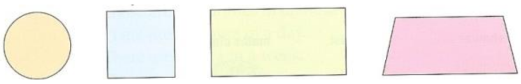
7. Describe the sequence in question 6.

The following exercises on the next page were on ‘properties of shape’ with Year 6 tutees, School A:

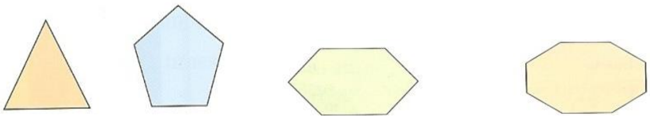
1. What are these shapes called?

………………………………………………… 

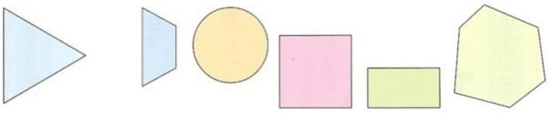
2. Identify the square from these shapes?



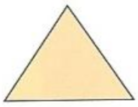
3. A pentagon has 5 sides, a hexagon has 6, and an octagon has 8: Tick which of the shapes below is a pentagon?



4. *A shape is regular if all its sides have equal lengths.* From the shapes below which ones are regular?



5. *To find the perimeter add the lengths of each side*. This triangle has equal sides, what is its perimeter: …………….



1.1cm

6. Which of these shapes has a larger perimeter if all their sides are equal?

9mm 1cm

7. Draw a square with a perimeter of 200mm:

Below is also an example of exercises in "factors, multiples and fractions" used with Year 7 tutees school B:

Write all the factors of:

1). 32

2). 60

3). 28

Write the first two multiples of:

4). 8

5). 16

Simplify the following fractions:

6). $\frac{8}{16}$

7). $\frac{24}{84}$

The following two examples on the next pages were used with the Year 8 tutees, school C. The first is on 'probability' and the second on 'equations':

Probability

1) For each outcome write whether it is impossible, certain, even chance, likely or unlikely.

1. It will snow in London in January
2. When you flip a coin it will land on tails
3. The sun will not rise tomorrow
4. You will get older
5. You will roll a three when you roll an ordinary six sided dice

2) There are 30 cars in a car park. 5 are silver, 12 are red, 6 are green and 7 are black. A random car leaves the car park. Work out the probability of the car being:

a) Red

b) Black

c) Green

d) Silver

e) Blue

3) Sarah has a bag with 20 coloured marbles in it. Five of the marbles are red. Four of the marbles are black. Seven of the marbles are blue. The rest are pink. Sarah takes one marble at random from the bag. What is the probability that the marble she takes is:

i) red ii) black iii) not blue iv) pink v) white

Equations

Find the missing numbers or the number represented by a letter:

1. $8 + ? = 12$

2. $3x = 18$

3. $b - 3 = 12$

4. $2x + 4 = 12$

5. $3a - 4 = -10$

6. $4x + 2 = 2x + 18$

7. $6(a + 2) = 4(a + 4)$

5.4.4 Ethical considerations

Information sheets with the details of the project, including data collection methods and storage, were presented to the students and to their parents/guardians, with the option to withdraw from the project or from any data collection methods at any stage. Also, a further written consent form was taken from all students on the questionnaire data-collection and the short-tests; such consent was taken twice, prior and after the intervention when the tests were conducted. Ethics approval was also gained from Durham University to conduct the intervention, collect and analyse the data[11].

In terms of the ethical dilemma of all students benefiting from the intervention, as mentioned schools were informed prior the intervention that they had the option of a second stage, i.e. applying the intervention to the control group, especially if the academic results of the first stage were positive.

Also, it was agreed with the head-teachers in advance that once the project had ended they would receive a USB stick with all ICAT resources, so that each school could apply ICAT to the control groups if the findings were positive, or for future school improvement. At the end of the intervention all the schools were provided with the USB sticks.

Finally, no harm was reported during the project.

[11] Please see appendix for all the consent forms and ethical approval confirmation emails from Durham University School of Education.

5.4.5 Timetable

Every school agreed to 35-40 minutes of cross-age peer tutoring each week for six weeks, as well as 45 minutes of student training prior to the intervention. The project started at the last term of the 2013 academic year after the Easter holiday, 3rd of June. Initially all schools conducted their pre-tests, training of students, and their first lesson at different times from one another during the first week.

Each school then proceeded at different times in order to fit their timetables. School A conducted their peer tutoring every Monday at 9am, school B every Thursday at 1pm, and school C every Wednesday at 10am. The post-tests were collected during the last week of peer tutoring, and last week of school, after all the peer tutoring sessions were completed.

5.4.6 School Assistance

School assistance is important in the implementation of a cooperative learning intervention (Jolliffe, 2005; 2011; 2015; Jolliffe & Hutchinson 2007). School assistance for the intervention took many forms:

Teachers working together: Prior to the intervention teachers within each school were asked to work together and with their head-teachers in order to create the exercises that were to be used for the intervention. Each teacher was given one or two mathematic topics to work on and exchange ideas. This was essential in order to give the teachers a sense of belonging to a group and part of the organising body, as opposed to being isolated.

In-house facilitator: For two of the schools, schools A and B, there were in-house facilitators in order to provide teachers with any queries or simply boost their confidence. For school C, the mathematics head teacher was chosen from the school's head to also serve as the facilitator. Research shows that an in-house facilitator is extremely important for the successful implementation of any cooperative intervention, with school teachers highly appreciating and requiring the support from the in-house facilitators (Jolliffe, 2005).

Supporting teachers in implementing cooperative learning: Once all the mathematics materials were chosen, each of the teachers from each of the schools sent the exercises to their mathematics head teacher, who then send them over to the researcher. The researcher then double-checked the exercises with the supervisors to ensure that the exercises within each lesson were ranked from easy to hard in terms of the level of complexity. Each lesson consisted of two sets of ranked exercises, one for the students to understand the concepts, titled 'Practice Test', and the other titled 'Turn-Taking Test', totalling 36 sets of ranked exercises for the three schools.

The second step was to integrate all the 12 sets of exercises for each school into the ICAT framework, print all the materials (most in colour) and place them in the relevant folders for the teachers of each peer tutoring class; for each of the schools there were 4 folders. This was to aid the teachers in efficiently managing the peer tutoring classrooms and in the collection of the peer tutoring lesson scripts at the end of each lesson. In many peer tutoring interventions the teachers are expected to choose the exercises on their own during the intervention as well as manage the paperwork. However research suggests that an effective way to implement cooperative learning is to provide as much support as possible to the teachers, including administrative support (Jolliffe, 2011).

Schools were also supported during and after the intervention. During the intervention regular email exchanges took place to remind and advice the head-teachers and the teachers. After the intervention came to an end the head-teachers were briefed on the results and how to move forward to implement peer tutoring in the future.

5.4.7 Pair and classroom set up

The pairings were arranged by the heads of the mathematics department based on previous individual classroom assessments.

The pairings of the students were conducted in the following way: firstly, each class was divided into two groups of boys and girls; secondly the highest performing older age boy was paired with the highest performing younger age boy, and so on down the line. This was repeated for the girls as well. Also, each of the younger and older classes had to be split into two in order to accommodate for space.

Due to the structure of the ICAT, the only requirement was that the tutor needed to sit on the tutee's right hand side for the training and the first lesson, in order for the students to become familiar with ICAT quicker. Once the students were paired, they worked together for the duration of the project, unless there were major disagreements within the pair. Schools A and B both contained a pair each changing their partner after the first training session. School C consisted of three pairs changing their partners after the training session. The students were not removed from the analysis.

5.4.8 Training

Training was provided for the facilitators and the mathematics head teachers for schools A and B, and for the head teacher for school C who was appointed to also be the facilitator by the school. Each received a total of three hours of training on the ICAT intervention. Also, the school facilitator from school A was a keen advocator of peer tutoring, and had recently attended a cross-age peer tutoring in reading programme. Therefore she was already confident with many of the aspects around peer tutoring in general.

Each training session was conducted either one-to one or one-to-two with the appointed persons, hence the training was intensive. Also, because the head teachers and teachers produced most of the exercises, the intervention involved two months of intensive preparations, consultation and interaction with the researcher, during which time most of the practicalities and arrangements were discussed.

The training concentrated on the following three topics:

1. ***Theory***. The basics of why and how peer tutoring works, and the literature that exists on peer tutoring. The training also explained the benefits of peer tutoring in terms of social and academic development as well as performance outcomes. Theoretical training is also important in the effective implementation of a cooperative intervention (Jolliffe, 2011).
2. ***Practice.*** Role-play with the head teachers and the facilitators on the ICAT framework.
3. ***Training the teachers and the students***. This part concentrated on how to train the teachers and the students to concentrate on the content of the ICAT and on the interpersonal skills, specifically how to give praise, to listen and combine the tone of voice with the body language and the context overall.

The head teachers and the facilitators were advised to arrange role play with the teachers and the students on both, interpersonal communication skills and the ICAT framework. They were also advised to train the teachers to be vigilant in the classrooms and keep an eye on monitoring both interpersonal as well as academic skills and not to disrupt the interactions during the lesson but wait until the end, unless there was a major issue.

A minimum of 12 mathematics teachers received training, four from each school; this included the mathematics head teachers as well, who also carried out the tutoring intervention themselves. Although the number of teachers who received training in school C is not for certain, as mentioned the school decided to extend the intervention to the entire Years 8 and 10.

Prior to the intervention both teachers and students received one full session of 45 minutes, which consisted mostly of role playing. The training was conducted in a cascade model, specifically the researcher trained the mathematics head teacher and a facilitator in each school, the head teacher and the facilitator then trained the teachers, and the teachers finally trained the students.

The materials for the student training were prepared and printed in advance by the researcher to facilitate the training taking place. Studies have found that training of the students is helpful in peer tutoring (Harrison & Cohen, 1971; Barron & Foot, 1991), as also illustrated by the meta-analyses mentioned.

The teachers were advised to conduct the 45-minute teacher and student training as follows:

The first 20 minutes of the training to concentrate on interpersonal communication skills and building rapport. Specifically, the pairs were asked to engage in role playing: First they had to sit far away from each other, interrupt, not listen or maintain a neutral face. Secondly, the pairs then had to sit close to each other, smile, listen and be nice to one another. And finally to discuss, a) why the second option is better, b) the importance of listening, and c) the importance of synchronising body language with the tone of voice. The training session on social skills was short since in one-to-one interactions the social dynamics are easier to manage than in higher groups of 3, 4 or 5 students (Abrami, et al, 1995; Johnson, Johnson, Holubec & Roy, 1980).

The remaining 25 minutes of the training, concentrated on the ICAT scripts, going through what to do at each of the four parts of the method. The students were told to bear in mind that in order to learn the method quickly all they had to do was to concentrate on four aspects of the peer tutoring script:

1. First, that the method contained very different but linked parts.
2. Second, that each part had the tutees' and the tutors' roles written down, with the tutor's roles written on the right hand side and the tutee's on the left hand side.
3. Third, that most parts contained numbers, in order to guide the pairs to navigate through each section while simultaneously employing the correct social and pedagogical skills.
4. Finally, that they had to praise correctly, i.e. praise only when the answer was correct and to praise kindly and mean the praise.

The reason for choosing a short training period for the actual framework is because the ICAT framework was designed from the beginning with the intention to be informative, and reduce the complexities of involving numerous materials. In other words ICAT does not contain separate praising cards, separate diagrams with instructions, separate question sheet, separate answer sheet. Therefore one strength of ICAT, is that it combines all these in one. The following section goes into detail regarding the ICAT framework.

5.4.9 ICAT framework.

The ICAT framework, consisted of four parts, as can be seen in Figures 7, 8 and 9:

Part 1, Goal Setting. In the first part, 'Goal Setting', the students chose together in pairs a number threshold that they wished to reach, a threshold which became their goal to be achieved. The emphasis at this stage was that both the tutor and the tutee take the decision *together*. They then were required to write the number threshold inside a circle, so that they would not forget.

Part 2, Practice-Test. In the second part, ‘Practice Test’, the tutor prompted the tutee to answer a range of mathematics questions. This was the part in which mathematics concepts were introduced to the tutee, ranging from very easy to very complex. Maximum interaction was expected at this stage, as the tutees would struggle with the new concepts, and the tutors were expected to provide help in different ways, at first implicitly then explicitly.

In the practice part the tutor started by prompting the tutee with a question, such as ‘How can we think about these exercises and solve them?’. The tutee then tried to answer the questions. These kinds of operational questions have been used to improve students’ meta-cognitive abilities (Mavaresh & Kramarski, 1997), and also used with the ‘Shared Maths’ project run by the CEM centre.

Answers were not given directly to the tutee. If the tutee could not answer the questions he/she first had to ask for hints from the tutor, who then had to explain or direct the tutee to the answer, however not actually give the answer. Only if the tutee still did not know the answer was he allowed to ask for the solution. The reason for this was that many studies in peer tutoring in mathematics have shown that providing indirect, procedural or elaborative help improves the performance of both the student who provides the explanation and the one who receives it; on the other hand, when the peers give the answer directly, the performance of the receiver or the giver does not improve, (Fuchs, et al., 1996; Native, 1994; Swing & Peterson, 1982; Webb, 1989; 1991; Webb, Troper & Fall, 1992). These ideas have been used in reciprocal peer tutoring, especially by Ginsburg-Block and Fantuzzo (1997, 1998).

Students were also provided with numbers of the steps they needed to take in order to set accountability and guide them through their interactions, specifically number 1) tutor askes the question ‘How would you approach and solve the questions?’, number 2) tutee responds and tries to solve the questions, number 3) tutee asks tutor for hints, number 4) tutor provides hints, however does not give the answer, number 5) if tutee still does not know the answer then he asks the tutor for the answer by putting operational meta-cognitive questions such as ‘how would you answer it?’, 6) tutor gives the answer.

Also, every time the tutee got a question correct, the tutor had to praise in a kind manner using words such as ‘well done’, ‘very good’, ‘excellent’, ‘brilliant’. All these steps were intended to make the students’ interactions more informative and provide students with more instructional autonomy in terms of knowing their specific roles and task at each stage, similar to same-age reciprocal peer tutoring (Ginsburg-Block and Fantuzzo, 1997, 1998).

Part 3, Connect. High interaction was also expected at the third stage, ‘Connect’. At this stage the tutee was prompted by the tutor on two issues: 1) to connect the new concepts to previous mathematics concepts, and 2) to connect the concepts to real life events. While the first question is influenced by the IMPROVE group learning method, (Mavaresh & Kramarski, 1997), the latter is influenced by ‘Duolog Mathematics’ used by Tymms, et al., (2011) and developed by Topping, Kearney, McGee (2004) with parents tutoring their siblings. This part was aimed to ensure advanced cognition and meta-cognition development. Again, the tutor was asked to provide help in different ways, first implicitly then explicitly.

Also, the tutor and the tutee were accountable for specific roles, which were similar to the ‘Practice Test’ section. In order to make the structured interactions more informative for the students, as well as enforce accountability, the roles and interactions were numbered from 1-6 in a similar manner to the practice test section. The tutor was also asked to provide praise if the tutee answered the questions correctly, again using praise words such as ‘very good’, ‘well done’, ‘excellent’ and ‘brilliant’.

<table>
<tr><td colspan="3">GOAL: Together, choose a number between 5–15. Write the number inside this circle:
This is the number of points you both want to win. Together you will do a little test at the end. Each correctly completed exercise will earn you 4 points.</td></tr>
<tr><td colspan="2">PRACTICE TEST: Tutee's Role:</td><td>Tutor's role:</td></tr>
<tr><td colspan="2">2). Try solving:

1. 44 + 22 = ?

2. 25.3 - 13.5 = ?

3. 14.8 - ? = 12

4. 30 x ? = 150

5. (4 x 5) + (10 x ?) = 50

6. (? x 2) x (8 x 1) = 32

7. (10 x 10) ÷ (2 x ?) = 5

Clues:</td><td>1). Ask:

"How can we think about these exercises and solve them?"

Direct only with nods.
If correct praise kindly:

Well Done!
Very Good!
Excellent!
Brilliant</td></tr>
<tr><td>3). Ask: Can you give me any clues?

For an answer:</td><td colspan="2">4). Respond with only hints and clues. If correct, praise kindly:
Well Done! Very Good!
Excellent! Brilliant!</td></tr>
<tr><td>5). Ask: How would you answer?</td><td colspan="2">6). Try giving answers. Once written down, praise kindly:
Well Done! Very Good!
Excellent! Brilliant
If you are not sure ask the teacher</td></tr>
</table>

Figure 7. *ICAT lesson script framework: Parts 1 and 2*

<table>
<tr><td>CONNECT: Tutee's Role:</td><td>Tutor's Role:</td></tr>
<tr><td>2). Summarise ideas below:

Clues:</td><td>1). <u>Ask:</u>

"Can you think of other things you have done in maths like this?

"Can you think when you might do this kind of maths in your everyday life?"

Direct only by nods.

If correct, praise kindly.

Well Done!
Very Good!
Excellent!
Brilliant</td></tr>
<tr><td>3). Ask: Can you give me any clues?

For an answer:</td><td>4). Respond with only hints and clues. If correct, praise kindly.
Well Done! Very Good!
Excellent! Brilliant</td></tr>
<tr><td>5). Ask: How would you answer?</td><td>6). Try giving ANSWERS.
If written down,
praise kindly.
Well Done! Very Good!
Excellent! Brilliant!
If you are not sure ask the teacher.</td></tr>
<tr><td colspan="2">TURN TAKING TEST:

1) Ask the teacher for the test.
2) Do the exercises in the test by taking turns.
3). Tutor marks tutee's exercises and gives feedback.
4) Together, add up the points for every correct answer, and check if you have won.
5) High-five each other for the great effort. ☺</td></tr>
</table>

Figure 8. ***ICAT lesson script framework: Part 3***

Solve the exercises below by taking turns:

1. $8 + ? = 21$
2. $3c = 21$
3. $b - 3 = 15$
4. $2x + 4 = 18$
5. $3a - 4 = -13$
6. $4x + 2 = 2x + 22$
7. $4(c + 4) = 6(c + 2)$

Figure 9. *ICAT lesson script framework: Part 4*

Part 4, Turn-Taking Test. The final part contained slightly different interaction types to the 'Practice Test' and 'Connect' section. In this section the students took turns to complete the exercises in order to determine whether they had achieved their self-set performance goals.

This final part was also marked by numbers 1-5 to guide the students' interactions and reinforce students' accountability. However, the ordering of the interactions is different in this part; number 1) students asked the teacher for the test, 2) students took turns in doing the exercises, number 3) the tutor marked tutees' exercises and gave the feedback to the tutee, 4) both tutor and tutee added up the points for the correct exercises and checked if they had won, number 5) students gave a high 5 to each other for the effort.

By making the tutor engage in exercises at the end, it was expected that the tutor had to take stage 1, 2, and 3 seriously, as he/she would need to have a good engagement with the mathematics concepts as well. By taking turns, rather than doing the exercises alone in different papers, the students had to be more engaged with one another due to resource interdependence. This also enabled the tutor to model the way he/she answered the questions to the tutee. Also, students were given learning autonomy at this stage as the final exercises were not marked by the teachers, but by the students themselves, so that students would not feel pressurised.

The next section outlines the instruments and the data collection process that took place:

5.5 Instruments and data collection process

In total four types of data were collected: mathematics performance tests, attitude questionnaire, observations and students' lesson scripts[12].

[12] The next Chapter concentrates on instrument development and what each instrument measures.

The performance tests and attitude questionnaires were given to the students together at the same time. First the students completed the attitude questionnaires then the mathematics tests. Students spend 10 minutes on the questionnaire and 35 minutes on the short test.

A timetable with all data-collection times was provided to each school. The questionnaires and the short tests were all administered and collected by the teachers with explanations from the researcher in terms of timing and how to operate the questionnaire.

The observations were conducted during the last two lessons of the project, so that the students had plenty of time to familiarise themselves with peer tutoring.

ICAT lesson scripts were collected by the teachers at the end of each lesson and were placed back in the folder provided to them.

5.5.1 Performance

Students’ mathematics performance was measured via three researcher-made instruments, reflecting both the students’ age and the mathematics topics covered during peer tutoring. The tests contained 3-4 questions for each of the six curriculum topics covered during the last 6 weeks of the year, for each of the schools.

5.5.2 Attitude

The questionnaire consisted of five sub-scales taken from various instruments:

The mathematics self-concept sub-scale was taken from Marsh’s ‘Academic Self-Description Questionnaire 1’ (ASDQ-1) (1990), with his consent.

Mathematics enjoyment, relating to people in mathematics classes, and choice of how to do mathematics in class were all taken and modified from Ryan and Deci's (2012) 'Intrinsic Motivation Instrument' (IMI) (IMI website), with their consent.

The social self-concept scale was a researcher adapted instrument, however, inspired by Harter (1985), with her consent.

5.5.3 Social and academic behavioural observations

This study employed general school observation as well as pair observation. In order to measure the extent of peer tutoring implementation two classes were randomly chosen from each school. At the beginning of the observations the names for the tutee and tutor were taken for each desk for organisational purposes. Each pair was observed through 5 windows, specifically the researcher stopped five times a meter away behind each pair. The procedure was similar to the observations conducted by Topping, et al., (2011).

It must, however be made clear that there are many issues with observation methods in general. For example, one of the main issues with observations is that it is difficult to gain a representative sample of the participants or the process of learning in in general (Cohen, Manion, & Morrison, 2000). Consequently, a researcher could end up observing participants which belong to a background with certain characteristics, or observe only certain learning processes.

Moreover, the following have also been identified to be some of the limitations associated with observation techniques: a) influencing students' behaviour, as they could perceive the observations as evaluations, b) not being able to know what goes on in students' mental engagement, and c) limited resources and time, together with evaluating the broader context have implications for reliability and final interpretations (Cohen, et al., 2000).

5.5.4 Lesson materials

Each of the four parts in the ICAT framework reported above in the intervention section, were used to: a) conduct analysis at the tutee group level, specifically between the higher performing group and the lower performing group of tutees, and b) conduct implementation analysis at the school level.

5.6 Chapter conclusion

This chapter has reported information on the participants, schools' characteristics, the intervention procedures such as group creation and characteristics, the design and data collection, exercises, training, and ethical issues, the ICAT framework. Also, the chapter briefly pointed out the instruments used for the intervention.

Three schools participated in the study, one school consisted of Year 6 students tutored by Year 8, Year 7 tutored by Year 9, and Year 8 tutored by Year 10. The schools were all reported as good schools by Ofsted, however, they differed from one another in terms of the percentage of students receiving free school meals, and the percentages of children in SEN and EFL categories.

The design for the intervention was the same for two out of three schools, with one of them deciding to implement ICAT to the entire school. The two schools with the same design, adopted a pre-post-test quasi-experimental design, and the remaining school a pre-post-test single group design. For the two quasi-experimental schools the groups prior to the intervention consisted of similar groups in terms of free school meals, however not in terms of previous performance, the better predictor.

The intervention was the same across the schools, one session of training and 6 sessions of ICAT. All teachers received the same amount of training in ICAT. Exercises in all schools were chosen by the teachers of each school to reflect their students' ability levels.

All schools applied the same data collection techniques, the same instruments in terms of attitude questionnaires, collecting lesson materials, observation techniques, and similar performance test, differing only in the topics and levels.

The next chapter provides more information on instrument development and coding.

6 Instrument Development and Data Coding

6.1 Introduction

This chapter reports instrument development. Specifically it looks at data entry procedures, how the performance tests, questionnaires, social interdependent variables, and observations were developed. Therefore, there are five sections to this chapter: The first section looks at data entry procedures. The next section then reports the Cronbach's Alpha for various performance tests developed for measuring students' mathematic performance for different ages. The third section looks at the attitude questionnaire and development, reporting the main reliability coefficients via Confirmatory Factor Analysis. Section four reports on the methods of coding for social interdependent variables used for the analysis of the lesson materials. The final part reports the development of the observation instrument.

6.2 Data Entry

All the data were double checked once entered into SPSS, this included performance tests scores, attitude questionnaires, the data from the lesson materials and the observation. Data for the lesson materials and the observations were also entered into Microsoft Excel.

The first step in entering the attitude data into SPSS was to create question items in the same order as they appeared on the questionnaire, so that the data entry would be quicker. The second step was to cluster all the questions that belonged to a particular variable. The final stage was to deal with missing data within each questionnaire, which accounted for less than 2%. Missing data were replaced with the mean of each student for a particular variable, if for example on the 'mathematics attitude' variable student **x** missed answering one of the questions on the questionnaire, the missing value was replaced with the mean of that student's 'mathematics attitude'. This process was carried out manually.

6.3 Mathematic tests reliability

Cronbach's Alpha is a statistic which measures instrument internal reliability for performance instruments; it measures the instruments' ability to produce consistent results (Blake, 2009). The method looks at the extent of correlation between various performance indicators and can then be used to make suggestions as to whether particular items could be dropped in order to improve the instrument's reliability. Cronbach's Alpha internal consistency is usually categorised as excellent (when is over 0.9), good (when is between 0.7-0.9), acceptable (when is 0.6 to <0.7), poor (when is between 0.5 and <0.6) and unacceptable (<0.5) (George & Mallerey, 2003). Although figures above 0.9 are categorised as 'excellent' such high figures could indicate that each items was measuring an identical feature and that the scale might lack diversity

Table 14 below provides the information for item reliability. For the performance tests in this analysis the lowest Cronbach's Alpha was for pre-test Year 8 tutees from school C, Alpha 0.36 and its post-test Alpha 0.59. As table 14 suggest none of the tests were excellent when investigated for pre or post test results. Two of the tests provided acceptable Cronbach Alphas, school A Year 6 post-test Alpha 0.67, and school B Year 7 pre-test Alpha 0.66.

Four of the tests provided good Cronbach Alpha coefficients, two of which were just about good; those of school A Year 6 pre-test, Alpha 0.7, school B post-test also Alpha 0.7, while for school B Year 9 pre-test Alpha 0.81 and post-test Alpha 0.77.

Table 14. *Cronbach's Alpha reliability coefficient for performance measurement*

Schools	*Year*	*Cronbach's Alpha Based on Standardized Items*	*n of Items*	*n*
A	Year 6 Pre-test	.70	6	98
	Year 6 Post-test	.67	6	97
B	Year 7 Pre-test	.66	5	64
	Year 7 Post-test	.70	5	63
	Year 9 Pre-test	.81	5	72
	Year 9 Post-test	.77	5	49
C	Year 8 Pre-test	.36	6	49
	Year 8 Post-test	.59	6	46

When analysing tutees' tests for school C, only 40% of the exercises were attempted. Hence, the students either found the test too hard, or not enough time was given to them, which is the main explanation why their Cronbach Alpha was unacceptable. However, the fact that half the performance instruments can be categorised as good without prior testing is an accomplishment. Nevertheless, caution needs to be taken with coming to final conclusions.

6.4 Attitude questionnaire reliability

This section applies confirmatory factor analysis (CFA) in order to evaluate the reliability of the attitude instrument. The evaluation was conducted twice, first with the attitude data from the two trial schools and then with the data derived from the ICAT schools.

Confirmatory factor analysis (CFA) as a statistical tool is usually used to test how well a construct or a model is represented by the collected data, in other words it measures how well the data fits a predicted model. CFA is usually used in two cases, to test a well formed model within a particular literature, or as a follow up after exploratory factor analysis. The moment changes are made to a model after using CFA procedures, one then enters exploratory factor analysis (EFA) areas. CFA can therefore also be used as a *measurement model development* the moment any changes are made to the initial model (Brown, 2006).

Since the attitude questions used in this work derive from various sub-scales of various theories, it was necessary to do both, to use CFA first as measurement model development and then as a confirmatory tool.

The initial aim was not to conduct CFA with the students' data for this peer tutoring project. However, the initial CFA coefficients used to test the attitude variables from the trial schools were not fully satisfactory, therefore the instruments needed to be further double checked.

After testing and revising the initial questionnaire of 32 items and 6 subscales used with the two trial schools, the final questionnaire for the actual peer tutoring intervention was reduced to 23 items, measuring 5 subscales, for reasons that will be explained shortly.

Many of the omitted items belonged to the omitted sub-scale *pressure*. Therefore, in both instances, using the data from the trial schools and the data from the peer tutoring schools, the measurement model was reshaped in order to strengthen the final model, which consisted of 17 items of the original 32. The following five sub-scales were used for this research, and each served as an additional dependent variable to the performance variable:

1. Mathematics self-concept.
2. Mathematics intrinsic motivation (enjoyment).

3. Level of choice when doing mathematics in class.
4. Social self-concept.
5. Relating to people in mathematics classes.

6.4.1 Trial schools

This section looks at EFA, developing a measurement model, and CFA employing the data from the two trial schools. The CFA model was tested with 167 Year 7 students and 139 Year 10 students, attending two secondary schools in the north East of England.

Developing a measurement model with Year 7 students. The tested instrument initially contained 32 items with the following six sub-scales:

- Mathematics self-concept (8 items).
- Enjoyment in mathematics (6 items).
- Social self-concept (5 items).
- Relating to people in mathematics classes (4 items).
- Choice in how mathematics is done in class (4 items).
- Pressure while doing mathematics in class (4 items).

The wording of the intrinsic motivation instrument by Royan and Deci (2012) was altered in order to make it accessible to younger students, as the instrument was initially designed to measure the intrinsic motivation of older students.

The following model was predicted: mathematic self-concept and mathematic intrinsic motivation, also known as enjoyment, would have a high correlation as they are similar in nature (Roseth, Johnson, D & Johnson, R., 2008). Social self-concept and relating to people in mathematics classes would also have a high correlation with each other since they both measure social issues; and finally relating to people in mathematics classes, choice and pressure would have a high correlation with intrinsic motivation in mathematics (with pressure negatively correlating to intrinsic motivation) (Royan & Deci, 2012).

Some of the main statistical assumptions for conducting CFA (Hair, Anderson, Tatham & Black, 1998) were the following:

1. The model needed to be over-identified, this was analysed in at least two ways: a) The degree of freedom needed to be over 0, in this analysis it was df=449; b) There should have been at least three items per scale, again this was achieved.
2. There should have been a ratio of at least five participants to an item, in this case it was 167÷32=5.22.
3. The data should have been normally distributed. Observations of Skew and Kurtosis coefficients did not indicate violation of this assumption, indicating that the data was normally distributed.

Figure 10 is a screenshot of the questionnaire and figure 11 presents the model, Table 15 provides the model fit summary for the first measurement model:

Absolute Fit Indices. The Absolute Fit Indices relate to the first part of table 15 under the heading CMIN. The Chi-square compares the actual structure of the data with that which is predicted, Chi-Square needs to be non-significant. However, in CFA the Chi-square is often significant due to the high degrees of freedom (Brown, 2006). The default model, under the heading CMIN in the table, shows that Chi-square = 756.635, df=449, ($p<.001$). Therefore conclusions could not have been made at this stage, other coefficients needed to be scrutinised.

Another Absolut Fit Indices coefficient which assesses the fit of the model is the CMIN/DF. The CMIN/DF coefficients for the default model are usually compared with those of the independent model, comparing the structure in the data with the hypothesis that there is no structure, the opposite of the Chi-square. The coefficients for the independent model need to also be non-significant and most importantly the p value needs to be higher than that of the default model (Hair, Anderson, Tatham & Black, 1998). Table 15 shows that the model was acceptable, i.e. CMIN/DF=1.685, ($p<.001$), while CMIN Independent Model CMIN/DF=6.552, ($p<.001$), although one cannot see the difference in the p value, taking away 1.685 (default model coefficients) from the coefficient 6.552 (independent model) gives a difference of 4.87 points. Which is considered to be a good difference, however, again no overall conclusions regarding the model could have been stated at this stage.

Comparative Fit Indices-TLI/CFI. TLI and CFI relate to the second part of table 15. This compares the data between the predicted model and a more restricted model. Here we look at Comparative Fit Indices (CFI) and Tucker Lewis Index (TLI), they both need to be .95 and above (Brown, 2006). Table 15 shows that this test was not passed as both CFI and TLI were under .95, CFI=888 and TLI=877. Therefore, the data did not fit the model.

Parsimony correction-RMSEA. The RMSEA correction refers to the last part of table 13. This looked at the 'root mean square error of approximation' (RMSEA), looking at whether the model fits the population data reasonably well, RMSEA is sensitive to model complexity but robust to variations in sample size, and needs to be less than .05 for the model to be considered as good (Brown, 2006). Table 15 shows that RMSEA=.064, also indicating that the model did not fit the data.

Please answer every question:	Not at all True		Somewhat True			Very True	
Compare to others my age I am good at MATHEMATICS classes	1	2	3	4	5	6	7
I find it easy to make FRIENDS	1	2	3	4	5	6	7
I feel like I have a choice in MATHEMATIC classes	1	2	3	4	5	6	7
I think MATHEMATICS is quite enjoyable	1	2	3	4	5	6	7
I feel very tense while doing MATHEMATICS in class	1	2	3	4	5	6	7
I am satisfied with how well I do in MATHEMATICS classes	1	2	3	4	5	6	7
I trust PEOPLE in my mathematics classes	1	2	3	4	5	6	7
I get good marks in MATHEMATICS classes	1	2	3	4	5	6	7
I know how to make many FRIENDS	1	2	3	4	5	6	7
I do MATHEMATICS in class because I have a choice	1	2	3	4	5	6	7
I enjoy doing MATHEMATICS very much	1	2	3	4	5	6	7
I feel pressured while doing MATHEMATICS in class	1	2	3	4	5	6	7
I like PEOPLE in mathematics classes	1	2	3	4	5	6	7
I learn things quickly in MATHEMATICS classes	1	2	3	4	5	6	7
I find it easy to become POPULAR	1	2	3	4	5	6	7
I feel like is my choice to do MATHEMATICS in class	1	2	3	4	5	6	7
MATHEMATICS is fun to do	1	2	3	4	5	6	7
I am relaxed while working on MATHEMATICS classes	1	2	3	4	5	6	7
PEOPLE are nice in the mathematics classes	1	2	3	4	5	6	7
Work in most MATHEMATICS classes is easy for me	1	2	3	4	5	6	7
I know how to get PEERS to accept me	1	2	3	4	5	6	7
I have a choice about doing MATHEMATICS in class	1	2	3	4	5	6	7
When doing MATHEMATICS I think how much I enjoy it	1	2	3	4	5	6	7
It is important to me to do well in MATHEMATICS classes	1	2	3	4	5	6	7
I feel nervous while doing MATHEMATICS	1	2	3	4	5	6	7
PEOPLE in mathematics can be trusted	1	2	3	4	5	6	7
I have always done well in MATHEMATICS classes	1	2	3	4	5	6	7
I know how to make CLASMATES like me	1	2	3	4	5	6	7
MATHEMATICS is a boring subject	1	2	3	4	5	6	7
I am calm in MATHEMATIC classes	1	2	3	4	5	6	7
I am hopeless when it comes to MATHEMATICS	1	2	3	4	5	6	7
I think MATHEMATICS is boring	1	2	3	4	5	6	7

Figure 10. *Attitude questionnaire for the trial schools*

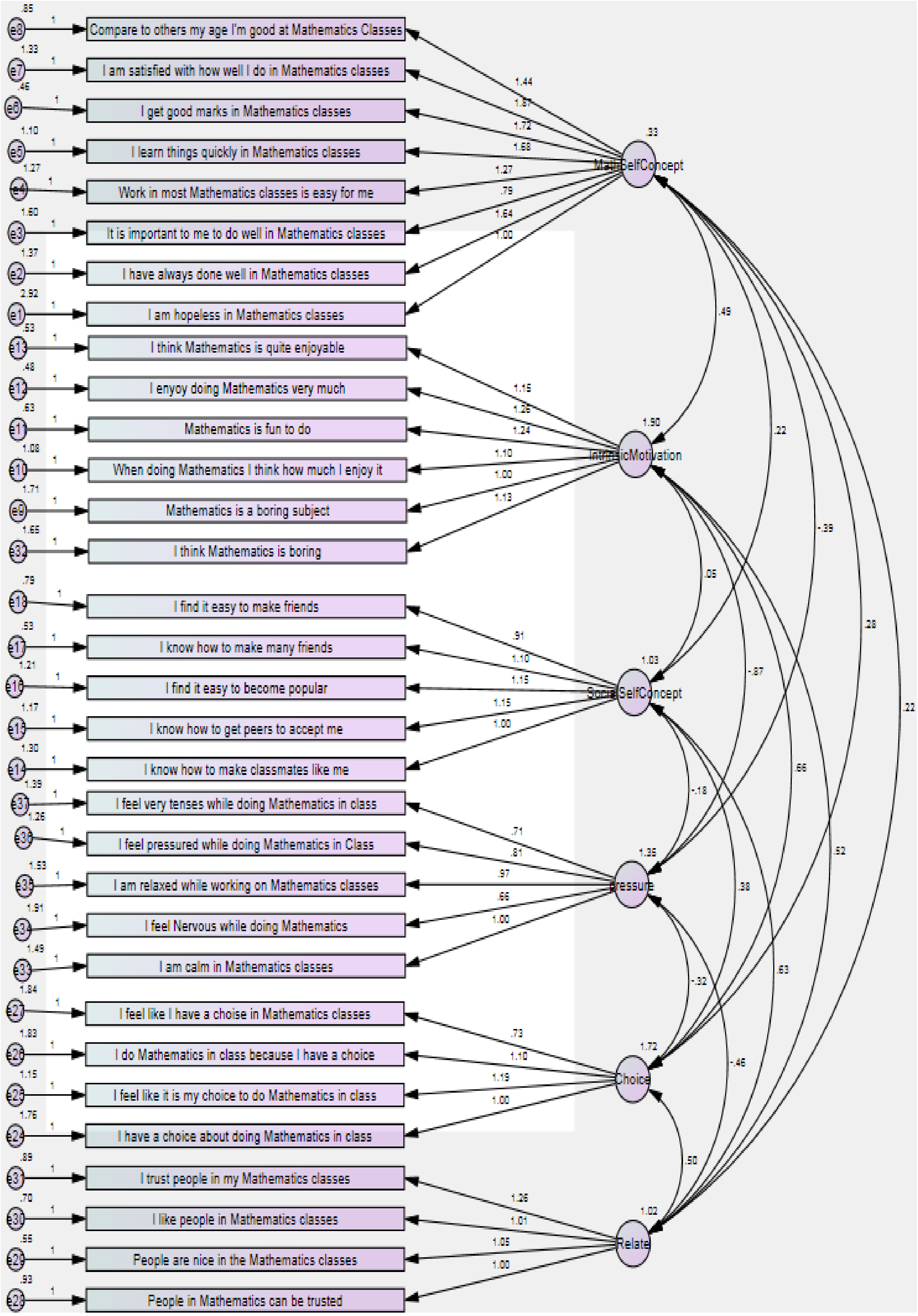


Figure 11. *First model outcome, Years 7 students: Trial school data*

Table 15. *First model fit summary with trial schools*

	CMIN				
Model	*NPAR*	*CMIN*	*DF*	*P*	*CMIN/DF*
Default model	79	756.635	449	.000	1.685
Saturated model	528	.000	0		
Independence model	32	3249.678	496	.000	6.552
	TLI/CFI				
Model	*NFI*	*RFI*	*IFI*	*TLI*	*CFI*
Default model	.767	.743	.890	.877	.888
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000
	RMSEA				
Model	*RMSEA*	*LO 90*	*HI 90*	*PCLOSE*	
Default model	.064	.056	.072	.002	
Independence model	.182	.176	.188	.000	

Further problems with the predicted model arose. First, apart from omitting the sub-scale ‘pressure’; e33 to e37 were adding complexity to the model as they negatively correlated to other scales. Second, the wording on the sub-scale ‘choice’ was misleading; feedback from a teacher from one of the trial schools pointed out that the students were not sure how to answer the questions from this sub-scale. Third, two of Marsh’s new items, e3 from the model graph ‘I am satisfied with how well I do in mathematics classes’, and e7 ‘It is important for me to do well in mathematics classes’ showed that their error variance correlates highly with other subs-scales. Finally, some of the items were very closely worded. In the social self-concept subscale, this was the case between e17 and e18, and in enjoyment between e9 and e32, in relating to people e29 correlated very highly to e30. Finally, the wording on the sub-scale ‘choice’ e-24 to e27 was not very clear.

Due to the above outlined limitations, the following changes were made to developing the model:

- Removed the sub-scale ‘pressure’.
- Removed Marsh’s two new items.
- Removed the similar items from sub-scales ‘enjoyment’ and ‘social self-concept’.
- In the sub-scale ‘relating to people’, e29 which correlates to e30 were maintained by joining the errors. The reason for maintaining both of them is due to the low number of items in this scale as compared to others.
- Even though some items in the ‘mathematic choice’ sub-scale were not worded correctly, the sub-scale was nevertheless maintained and the items were re-worded for the actual intervention.

Confirming the model: Year 7. Figure 12 on the next page is the final measurement model with the first tested data after the above changes, and table 16 provides the model fit summary:

As can be seen from table 16 above all model-fit indices showed improvement and now fitted the model. Chi-square = 295.029, df = 219, ($p<.001$). The difference between the default model CMIN/DF and independent model showed an increase from initially 4.87 to 7.61, improving by 2.7. TLI and CFI were also shown to improve to .956 and .962, respectively. And finally RMSEA was reduced to .046.

Assessment of normality. All items for both Skewedness and Kurtosis were within a range of +/-3, indicating that the data were normally distributed.

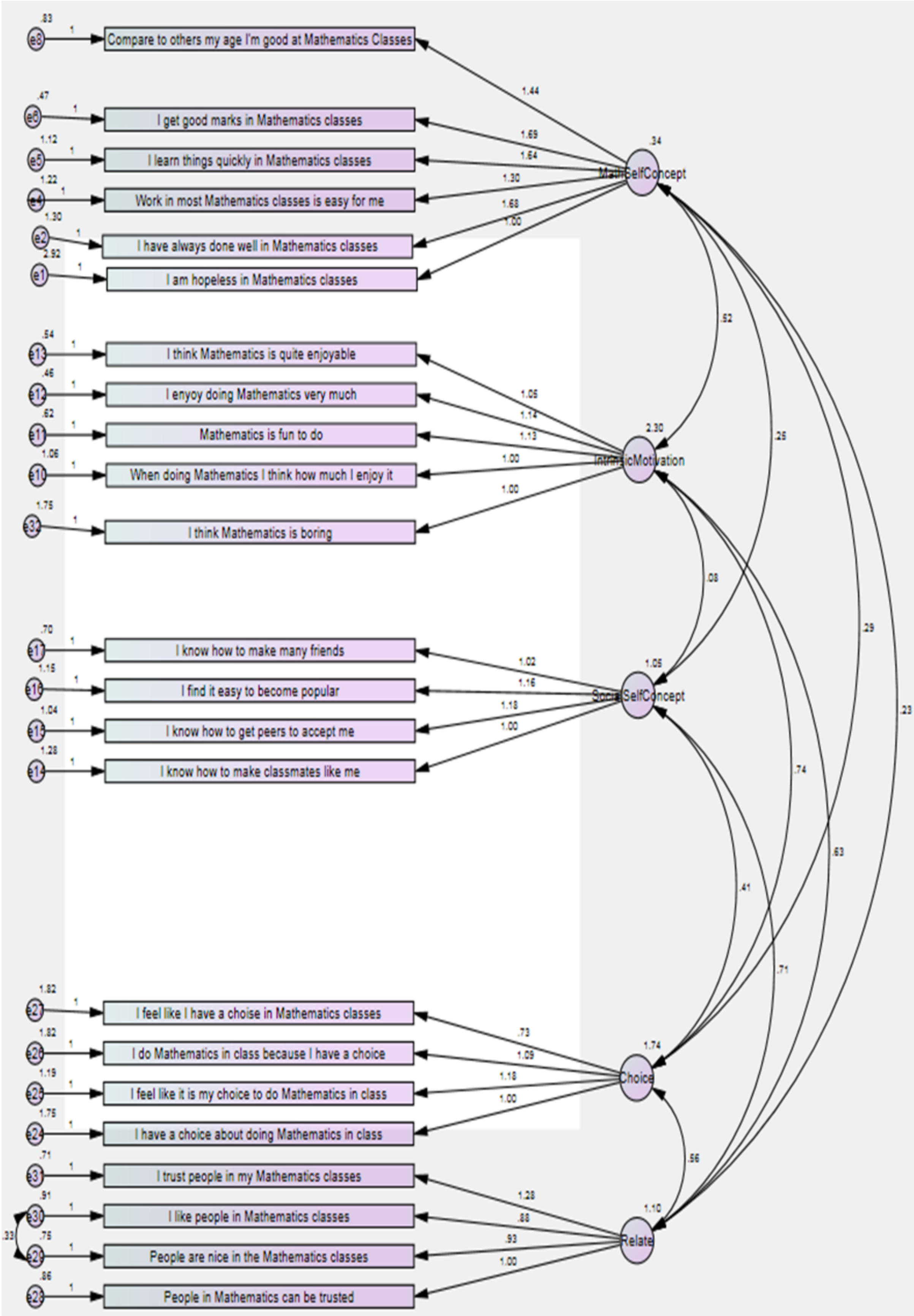


Figure 12. *Confirming the model, Year 7: Trial school data*

Table 16. *Confirming the model with trial schools, year 7: Model fit summary*

Model	*NPAR*	*CMIN*	*DF*	*P*	*CMIN/DF*
Default model	57	295.029	219	.000	1.347
Saturated model	276	.000	0		
Independence model	23	2267.16	253	.000	8.961

TLI/CFI

Model	*NFID*	*RFI*	*IFID*	*TLI*	*CFI*
Default model	.870	.850	.963	.956	.962
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

RMSEA

Model	*RMSEA*	*LO 90*	*HI 90*	*PCLOSE*
Default model	.046	.031	.058	.700
Independence model	.218	.210	.227	.000

Confirming the Model: Year 10. Figure 13 on the next page is a confirmation of the revised model used with the year 10 students, and table 17 provides the model fit summary:

Looking at the coefficients from table 17 for the year 10 CFA, it is clear that the model was still not great and could not be fully confirmed. However, this could have been mostly due to the item wording. Therefore, other CFAs with the data from the project were necessary. Hence for the actual peer tutoring project the wording was taken into account and for the subscale 'mathematics choice', some of the items had to be changed.

Assessment of normality. All items for both Skewedness and Kurtosis were within a range of +/-3, again, indicating that the data were normally distributed

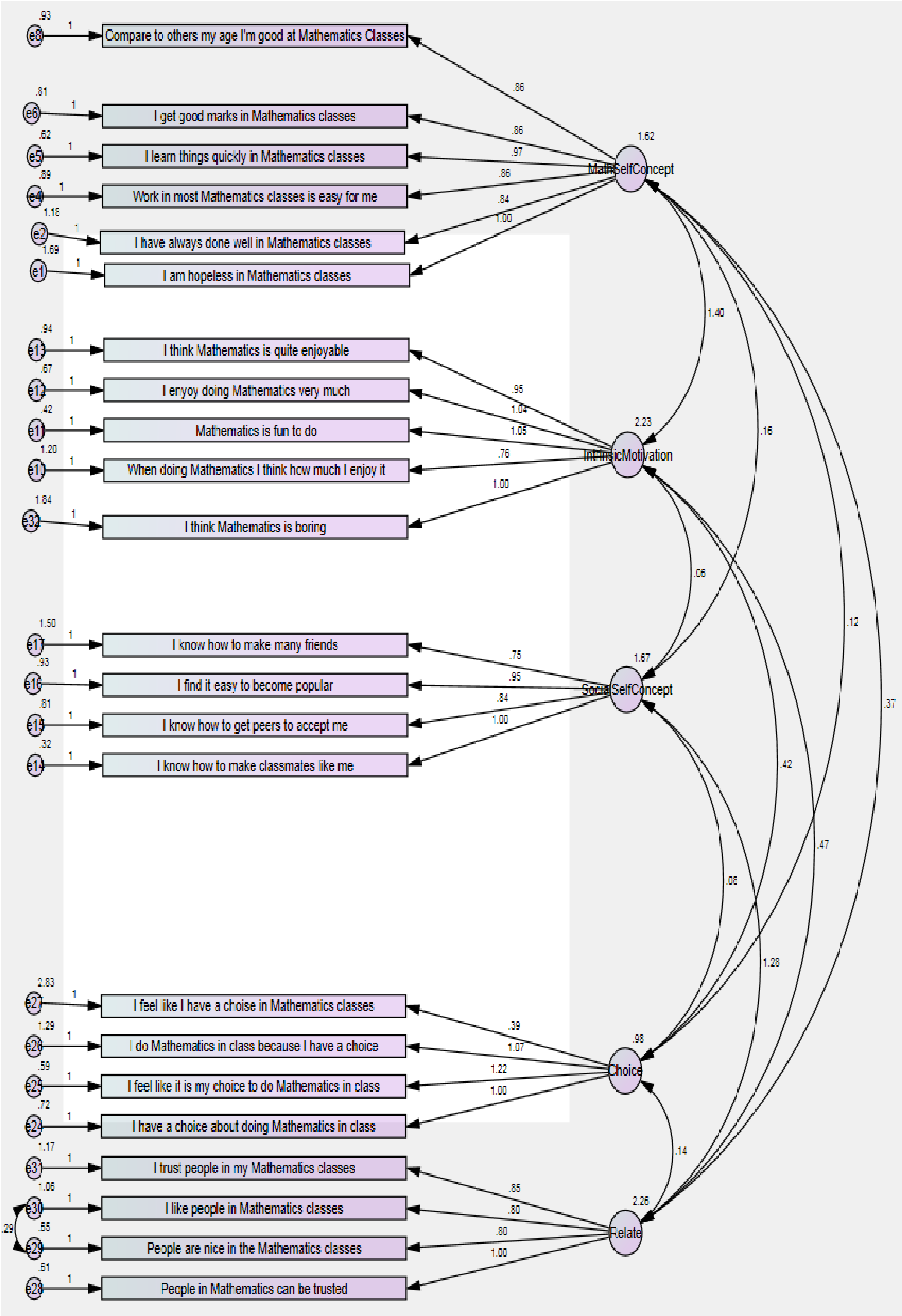


Figure 13. *Confirming the model, Year 10: Trial school data*

Table 17. *Confirming model trial school Year 10: Model fit summary*

	CMIN				
Model	*NPAR*	*CMIN*	*DF*	*P*	*CMIN/DF*
Default model	57	454.204	219	.000	2.074
Saturated model	276	.000	0		
Independence model	23	2215.590	253	.000	8.757
	TLI/CFI				
Model	*NFI*	*RFI*	*IFI*	*TLI*	*CFI*
Default model	.795	.763	.882	.862	.880
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000
	RMSEA				
Model	*RMSEA*	*LO 90*	*HI 90*	*PCLOSE*	
Default model	.089	.077	.100	.000	
Independence model	.238	.229	.247	.000	

6.4.2 ICAT schools data: Model development and CFA

This section carries on with EFA and CFA with the ICAT school data, in order to ensure that the instrument is reliable for the final analysis:

School A: Year 6 – Pre-test data: The final instrument given to the students in the ICAT project consisted of 23 items (5 sub-scales). Most of the assumptions required to proceed with CFA were achieved. Figure 14 below is a screenshot of the questionnaire showing the model items (questions) that entered the model, and figure 15 shows the measurement model:

Please answer every question:	Not True			Somewhat True			Very True
Work in most MATHEMATICS classes is easy for me	1	2	3	4	5	6	7
I know how to make many FRIENDS	1	2	3	4	5	6	7
I think MATHEMATICS is enjoyable	1	2	3	4	5	6	7
I trust people in my MATHEMATICS classes	1	2	3	4	5	6	7
It is my choice how I do MATHEMATICS in class	1	2	3	4	5	6	7
I have always done well in MATHEMATICS classes	1	2	3	4	5	6	7
I find it easy to become POPULAR	1	2	3	4	5	6	7
I enjoy doing MATHEMATICS	1	2	3	4	5	6	7
I like people in MATHEMATICS classes	1	2	3	4	5	6	7
In class I do MATHEMATICS the way I want	1	2	3	4	5	6	7
I learn things quickly in MATHEMATICS classes	1	2	3	4	5	6	7
I know how to make PEOPLE accept me	1	2	3	4	5	6	7
MATHEMATICS is fun to do	1	2	3	4	5	6	7
People are nice in the MATHEMATICS classes	1	2	3	4	5	6	7
I have a choice in how I do MATHEMATICS in class	1	2	3	4	5	6	7
Compared to others my age I am good at MATHEMATICS classes	1	2	3	4	5	6	7
I know how to make CLASSMATES like me	1	2	3	4	5	6	7
I think MATHEMATICS is boring	1	2	3	4	5	6	7
People in MATHEMATICS classes can be trusted	1	2	3	4	5	6	7
I am hopeless when it comes to MATHEMATICS	1	2	3	4	5	6	7
MATHEMATICS is a boring subject	1	2	3	4	5	6	7
I choose the way I do MATHEMATICS	1	2	3	4	5	6	7
I get good marks in MATHEMATICS classes	1	2	3	4	5	6	7

Figure 14. *Peer tutoring attitude questionnaire*

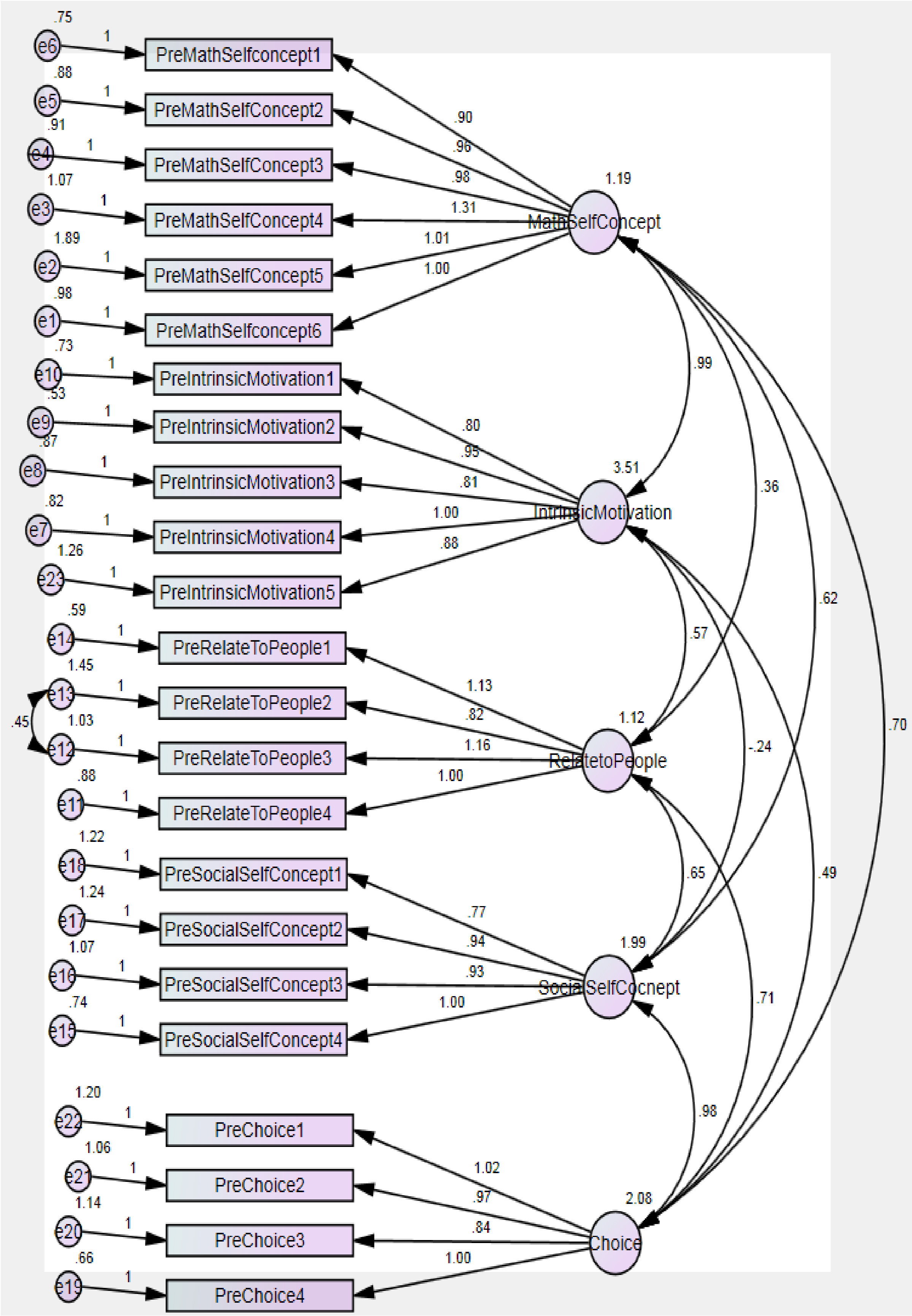


Figure 15. *First measurement model, Year 6: Peer tutoring school data*

Table 18. *Measurement model development, peer tutoring school, Year 6: Model fit summary*

	CMIN				
Model	*NPAR*	*CMIN*	*DF*	*P*	*CMIN/DF*
Default model	57	345.702	219	.000	1.579
Saturated model	276	.000	0		
Independence model	23	1723.779	253	.000	6.813
	TLI/CFI				
Model	*NFI*	*RFI*	*IFI*	*TLI*	*CFI*
Default model	.799	.768	.916	.900	.914
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000
	RMSEA				
Model	*RMSEA*	*LO 90*	*HI 90*	*PCLOSE*	
Default model	.076	.060	.091	.004	
Independence model	.241	.230	.252	.000	

*Assessment of normality***.** All items for both Skewedness and Kurtosis were within a range of +/-3. Therefore, indicating that the data were normally distributed.

*Absolute Fit Indices-CMIN***.** Chi-square = 345.702, df= 219, ($p<.001$). Table 18 shows that the model passed that test, CMIN Default Model CMIN/DF=1.579, ($p<.001$), while CMIN Independent Model CMIN/DF=6.813, ($p<.001$), with a difference of 5.23, suggesting that the model fit the data.

Comparative Fit Indices-CFI/TLI: Table 18 showed that this test was not passed, as both CFI and TLI were under .95; CFI=.914 and TLI=.900, suggesting that the model was slightly under the threshold.

Parsimony Correction-RMSEA: Table 18 above showed that RMSEA=.076, hence also showing that the model did not fit the data.

Overall it can be concluded that the model did not fit the data as it should have. Hence, the model needed to enter exploratory stage again, EFA, to be re-confirmed after the exploratory stage.

The following problems arose with the model above: The first issue to notice from the graph is that social self-concept correlated negatively to mathematic enjoyment. Secondly, some of the item errors correlated highly with items from other subscales. The main reasons for this is that they were; a) too long (mathematics self-concept, item 4 (e3), b) negatively worded (mathematics self-concept item 5 (e2), mathematics enjoyment, item 4 (e7) and 5 (e23)), and c) still wrongly worded, such as (choice, item 2 (e21), and social self-concept, item 3 (e16)).

There was also a high sub-scale correlation, specifically between social self-concept and choice, which was not predicted by the model specification set at the beginning of this section, this was puzzling. However, further investigation of the model on other samples within the peer tutoring population did not show this relationship. Please see model diagrams in the appendix for more information.

The following changes took place to improve the model:

- The negative relationship between social self-concept and mathematics enjoyment was removed from the model. This should have taken place at the data-testing stage, as there was no theoretical justification for this relationship.

- The items which are wrongly worded, too long, or negatively worded were also removed.

The following questions and variables entered the final model and the analysis stage:

Mathematics self-concept variable

Item 1) Work in most **MATHEMATICS** classes is easy for me

Item 2) I have always done well in **MATHEMATICS** classes

Item 3) I learn things quickly in **MATHEMATICS** classes

Item 6) I get good marks in **MATHEMATICS** classes

Mathematics intrinsic motivation variable

Item 1) I think **MATHEMATICS** is enjoyable

Item 2) I enjoy doing **MATHEMATICS**

Item 3) **MATHEMATICS** is fun to do.

Relating to people in Mathematics class variable

Item 1) I trust people in my **MATHEMATICS** classes

Item 2) I like people in **MATHEMATICS** classes

Item 3) People are nice in the **MATHEMATICS** classes

Item 4) People in **MATHEMATICS** classes can be trusted

Social self-concept variable

Item 1) I know how to make many **FRIENDS**

Item 2) I find it easy to become **POPULAR**

Item 4) I know how to make **CLASSMATES** like me

Having a choice in how mathematics was done in class variable

Item1) It is my choice how I do **MATHEMATICS** in class

Item 3) I have a choice in how I do **MATHEMATICS** in class

Item 4) I choose the way I do **MATHEMATICS**

Figure 16 on the next page was the final model used to analyse the findings:

*Assessment of normality***.** All items for both Skewedness and Kurtosis were within the +/- 3 range. Therefore, indicating that the data were normally distributed.

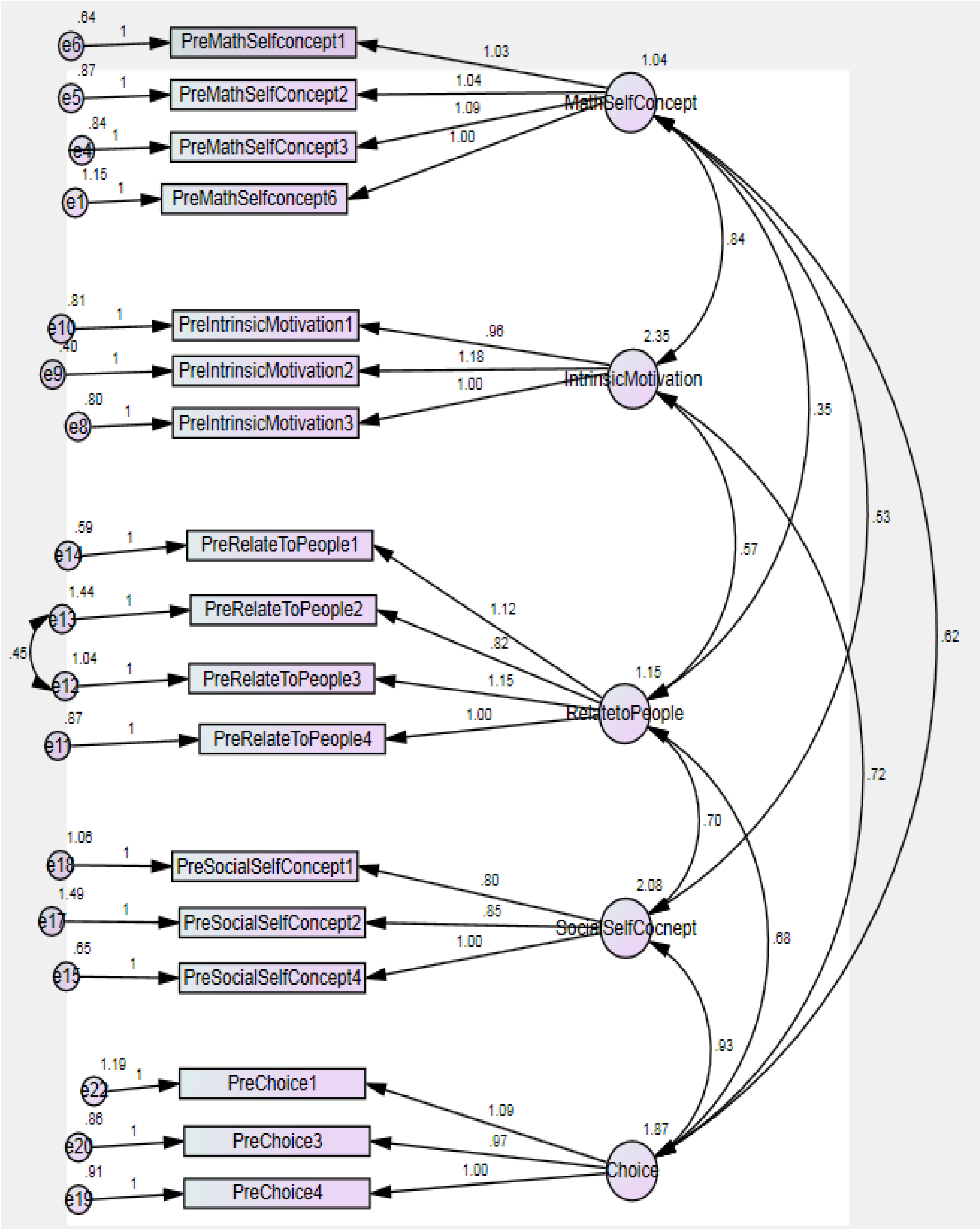


Figure 16. *Confirming the model, Year 6: Peer tutoring school data*

Table 19. *Confirming the model, peer tutoring schools, Year 6: Model fit summary*

	CMIN				
Model	*NPAR*	*CMIN*	*DF*	*P*	*CMIN/DF*
Default model	44	131.964	109	.067	1.211
Saturated model	153	.000	0		
Independence model	17	1006.709	136	.000	7.402
	TLI/CFI				
Model	*NFI*	*RFI*	*IFI*	*TLI*	*CFI*
Default model	.869	.836	.974	.967	.974
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000
	RMSEA				
Model	*RMSEA*	*LO 90*	*HI 90*	*PCLOSE*	
Default model	.046	.000	.072	.578	
Independence model	.253	.239	.268	.000	

As illustrated by table 19, each of the three indices showed improvement. The model was confirmed. There was still a strong correlation that was not predicted, that between social self-concept and choice. However, as mentioned when confirming the model on the Year 6 post-tests and other ages, such high correlation was absent (please see diagrams in appendix). In order to show how the model compares in relation to all ages, table 20 below provides an overall conclusion:

Table 20. *Instrument development and CFA coefficients by year and school*

	Assumptions				*CMIN*		*CFI*	*TLI*	*RMSEA*
Instrument Development	*Nor-malit-y*	*DF*	*N*	*Chi*	*Default Model*	*Indepe-ndent Model*			
Year 7 (First model development)	Yes	449	168	756.635 p<.001	1.685	6.55	.89	.88	.06
Year 7 (CFA Test)	Yes	219	168	295.029 p<.001	1.347	8.961	.96	.96	.05
Year 10 (CFA Test)	Yes	219	138	454.204 p<.001	2.074	8.757	.88	.86	.09
Tutees:									
Pre-test year 6 (First model development)	Yes	219	101	345.702 p<.001	1.579	6.813	.91	.90	.08
CFA: Confirming the Model									
Pre-test Year 6 (Final model 17 items)	Yes	109	101	131.964 p=.067	1.21	7.40	.97	.97	.05
Post-test Year 6	Yes	109	99	165.558 p<.001	1.519	8.769	.95	.93	.07
Pre-test Year 7	Yes	109	66	149.793 p=.006	1.374	4.955	.92	.91	.08
Post-test Year 7	Yes	109	66	182.470 p<.001	1.674	6.521	.90	.88	.10
Pre-test Year 8	Yes	109	43	139.954 p=.024	1.284	4.798	.94	.93	.08
Post-test Year 8	Yes	109	44	162.029 p=.001	1.487	4.893	.90	.88	.11
Tutors:									
Pre-test Year 8	Yes	109	117	130.005 p=.083	1.193	7.140	.98	.97	.04
Post-test Year 8	Yes	109	116	129.497 P=.088	1.188	8.072	.98	.97	.04
Pre-test Year 9	Yes	109	68	170.449 p<.001	1.564	4.915	.89	.86	.09
Post-test Year 9	Yes	109	57	155.634 p=.002	1.428	5.236	.92	.90	.09
Pre-test year 10	Yes	109	72	149.780 p=.006	1.374	5.181	.93	.91	.07
Post-test Year 10	Yes	109	65	188.933 p<.001	1.733	6.420	.89	.87	.11

Table 20 above clearly shows that for most ages the model did not entirely meet the criteria, however, they were close to being confirmed. There were at least four reasons regarding this outcome: Firstly, the nature of mathematics self-concept and mathematic enjoyment were relatively similar, as predicted. Secondly, social self-concept and relating to people were also relatively close. This could have been due to the reason that the social self-concept taken from Harter was transformed, in order to fit the overall instrument format, hence losing many of its properties. Finally, the timing of the post-tests was conducted at the end of the school year and therefore there is a strong reason to believe that the students did not take the instruments as seriously, considering that most of the post-tests have poorer coefficients than the pre-tests. Other limitations were found in some of the graph models, especially with the Year 9 tutors from school B. Specifically some of the items and sub-scales for the mathematic self-concept had a negative correlation as shown in the appendix.

Finally, although in CFA literature the coefficients benchmarks are not fixed, as different authors quote different statistical benchmarks (Brown, 2006), instrument limitations need to be taken into account in the final analysis and conclusion.

6.5 Group comparison instruments/variables

Most of the process analysis research which seeks to explore how students work together concentrates on observed behavioural interactions and their test outcomes, with the interactions usually video recorded. This study, by contrast, concentrates on exploring whether there were any differences in the lesson materials between differently achievement groups of tutees that received the peer tutoring.

Table 21 provides a map showing what social interdependent areas each indicator measured as part of instrument development, (for the exact questions for each indicator please refer to the research questions, section 4.3):

The total number of lessons which entered the analysis was 507 ICAT lesson scripts from 102 students. For school A there were a total of 149 lessons, from school B 152, and from School C, 206 lessons.

Table 21. *Area of measurement by indicator type*

Indicators	*Area measuring*
Amount of peer tutoring lessons	This served as an overall social interdependent variable
Amount of goals set	Goal interdependence
Attempted exercises in the *practice test* section	Cognitive engagement, task interdependence, cross-ability
Quality of answers in the *practice test* section	Indicator for cognitive engagement, and cross-ability
Attempted exercises in the *connection* section	Cognitive engagement indicator, task interdependence
Quality of answers in the *connection* section, i.e. negative, broad or specific statements.	Indicator for meta cognitive engagement, cross-ability
Attempted exercises in the *turn-taking* test	Cognitive indicator, task interdependence
Quality of answers in the *turn-taking* test	Cognitive indicator, cross-ability
Amount of *total feedback*	Interpersonal interdependence, cognitive and meta cognitive indicator, cross-ability
Feedback by ticks/crosses in the *practice-test*	Interpersonal interdependence, cognitive and meta cognitive indicator, cross-ability
Feedback by ticks/crosses in the *turn-taking* test	Interpersonal interdependence, cognitive and meta cognitive indicator, cross-ability
In terms of *checking if won*	Goal interdependence

The following sections provide more information as to how the indicators were coded:

6.5.1 Arrangement of the lesson materials and the group threshold

All lesson materials with student names on were clustered into mini-portfolios, each portfolio was given a student ID. Only these lesson materials were considered for analysis since only the materials with the names on could be linked to tutees belonging to a particular group.

A benchmark of two points increase, taken from the researcher made mathematics performance test data, was used to divide the lower performing tutees from the higher performers. In other words, if a tutee achieved a mathematics performance of more than two points from the pre-test to the post-test, then that student's lessons were placed on the higher performing end of the peer tutoring group for comparison purposes. Hence 507 lessons from 102 students from three schools were split into two groups, the higher performing group, 60 tutees, and the lower performing group 42.

There were two reasons for choosing this score cut off was chosen: Firstly, the cut off represented, or was close to the median 2.5, since the performance gains ranged from -5 to 11. Secondly, considering the short period of the intervention, 6 weeks, any higher test gains due to maturation were unlikely.

6.5.2 Question 1: Total peer tutoring lessons (Indicator 1)

The total peer tutoring lessons were counted for each student portfolio, and then the total lessons within each group of tutees.

Separation of lessons between the higher and the lower-performing tutees, was then maintained for the remaining questions and indicators.

6.5.3 Question 2: Goal interdependence (Indicator 2)

For this question the lessons with set goals were counted for each student within each group, and then for the total group. The following figure (17) is an example of a student having set a goal:

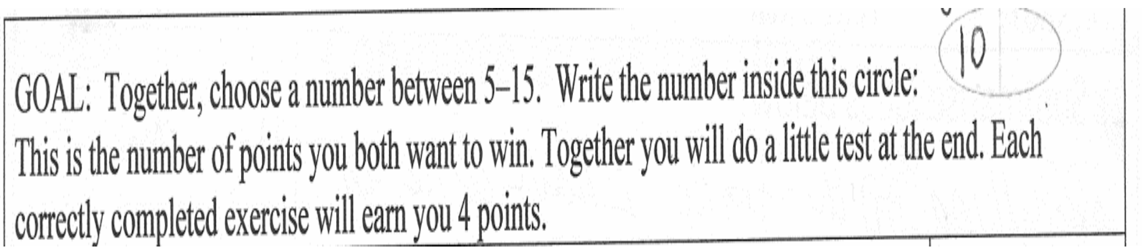

GOAL: Together, choose a number between 5–15. Write the number inside this circle: 10
This is the number of points you both want to win. Together you will do a little test at the end. Each correctly completed exercise will earn you 4 points.

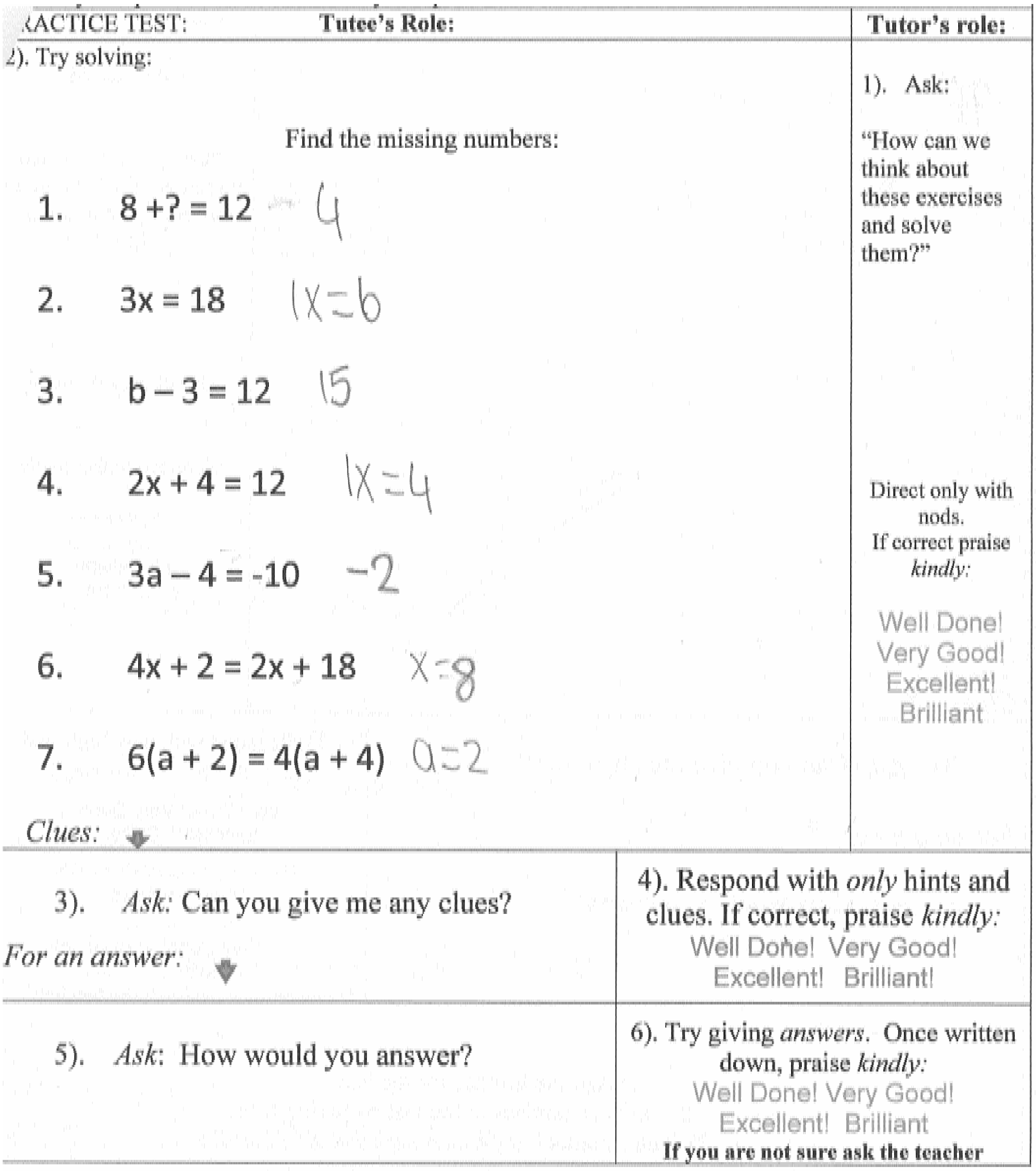
Figure 17. *Example of student set goal*

6.5.4 Questions 3, 4 and 5: Quantity of exercises attempted by each group of tutees (Indicators, 3, 4 and 5)

Cognitive engagement was indirectly measured by the number of attempts in the “Practice-Test” and “Turn-Taking Sections”. This also indirectly aimed at investigating whether there was a difference in ability between the tutee and the tutor within each group of tutees, since tutors who are confident and know the materials would have pushed their tutees to tackle more questions. Also, the quantity of exercises attempted served as task interdependence, the more questions the pairs attempted the more they would have been interdependent by the task.

The “Connect” section, measured the amount of meta-cognition, rather than cognition that emerges for the tutee; it was structured in such a way that students were required to monitor the knowledge they engaged with.

In order to identify the data for the '*quantity of attempted exercises*' variables, the attempted exercises within each part of the lesson-script were added up for each student within each tutee group, and then for the total group. The following figure (18) is an example of a full attempt in one of the ICAT sections (Practice section):



ACTICE TEST: **Tutee's Role:**	**Tutor's role:**
2). Try solving: Find the missing numbers: 1. $8 + ? = 12$ 4 2. $3x = 18$ $x = 6$ 3. $b - 3 = 12$ 15 4. $2x + 4 = 12$ $x = 4$ 5. $3a - 4 = -10$ -2 6. $4x + 2 = 2x + 18$ $x = 8$ 7. $6(a + 2) = 4(a + 4)$ $a = 2$ *Clues:*	1). Ask: "How can we think about these exercises and solve them?" Direct only with nods. If correct praise *kindly:* Well Done! Very Good! Excellent! Brilliant
3). *Ask:* Can you give me any clues? *For an answer:*	4). Respond with *only* hints and clues. If correct, praise *kindly:* Well Done! Very Good! Excellent! Brilliant!
5). *Ask*: How would you answer?	6). Try giving *answers*. Once written down, praise *kindly:* Well Done! Very Good! Excellent! Brilliant **If you are not sure ask the teacher**

Figure 18. *Example of student practice section attempt*

6.5.5 Questions 6, 7, and 8: Quality of answers by each group of tutees (Indicators, 6, 7, 8)

The quality of answers, such as correct or incorrect, in each part in the ICAT measured similar indicators as those in the *quantity* section. Analysing the quality of attempts in the “Practice Test” and “Turn Taking” Sections aimed at measuring the amount of cognitive engagement as well as whether there was a difference in ability between the tutee and the tutor, since common sense would entail that tutors who were confident and knew the materials would have helped the tutee to answer correctly by providing better explanations.

The quality of answers in the students’ lesson materials could be partially explained by the quality of explanations provided by the tutor. For example Webb (1980; 1982) has shown that higher ability tutors’, or older tutors’ explanations had an impact on tutees’ performance.

Further, the quality of explanations provided by the tutor has been found to be linked to their ability level, Fuchs, Fuchs, Karns, Hamlett, Suzanne Dutka and Katzarof, (1996). That experimental study was conducted in mathematics with students of different ages. It paired a higher ability and average ability tutor with the same tutee at different times, showing that the higher ability tutor was able to provide more procedural explanations than the average ability tutee, effect size 0.67, more conceptual explanations, effect size 0.64, and more social affect/attention to the task, 0.39.

The Fuchs, et al., (1996) study also investigated the quality of the answers provided by the tutee on mathematics problems during the lessons. They illustrated, by investigating tutees’ lesson materials, that when the tutees worked with the higher achieving tutor the tutees completed correctly 91% (standard deviation 0.12) of the problems they attempted, while when working with the average achieving tutor they completed only 75% (standard deviation 0.26), giving a significant effect size of 0.84, at (p.<.05).

Similarly, the ‘Connect’ section, also measured the mathematics cross-ability levels between the tutor and tutee. However, as mentioned this part was better suited to measuring the quality of meta-cognition within the pair.

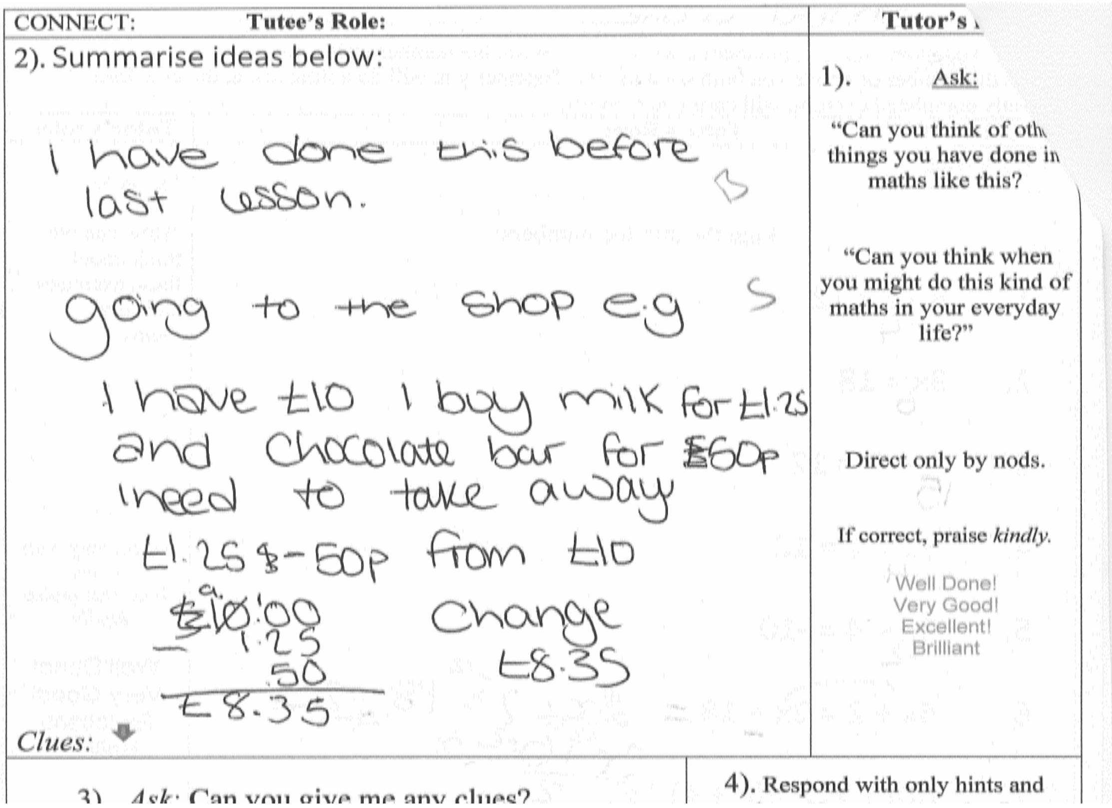
The analysis for the ‘Practice-Test’ (question 6) section and the ‘Turn-Taking Test’ (question 8) section followed the general trend of steps discussed so far, specifically adding up the correct answers given by each student in each lesson within each higher or lower group.

In the ‘Connect Section’, question 7, the analyses were conducted slightly differently due to the nature of the exercise, such as providing statements and making links. A better comparative picture was to code whether the statements provided by the tutee were, ‘specific’, ‘broad’, or ‘negative’; then count the types of coded comments for each tutee within each group.

The expectation was that the higher performing group would have had a higher ratio of ‘specific/elaborative’ rather than ‘broad’ or ‘negative’ comments than the lower performing group. This expectation was influenced by the Fuchs’, (1997) study, which showed that students who in a peer tutoring group provided conceptual/elaborative explanations, which were elaborative or specific, performed better than students in a peer tutoring group who did not provide such explanations.

Examples of the coding for question 7 are provided below:

Broad and specific examples can be seen in the following lessons: (broad was coded with a ‘B’ by the researcher, negative with an ‘N’ and specific with an ‘S’).



CONNECT: Tutee's Role: | Tutor's

2). Summarise ideas below:

i have done this before last lesson.

going to the shop e.g

I have £10 I buy milk for £1.25 and Chocolate bar for £60p i need to take away £1.25 £-50p from £10

£10.00
-1.25
.50
£8.35

Change £8.35

Clues:

3) Ask: Can you give me any clues?

1). Ask:

"Can you think of oth things you have done in maths like this?

"Can you think when you might do this kind of maths in your everyday life?"

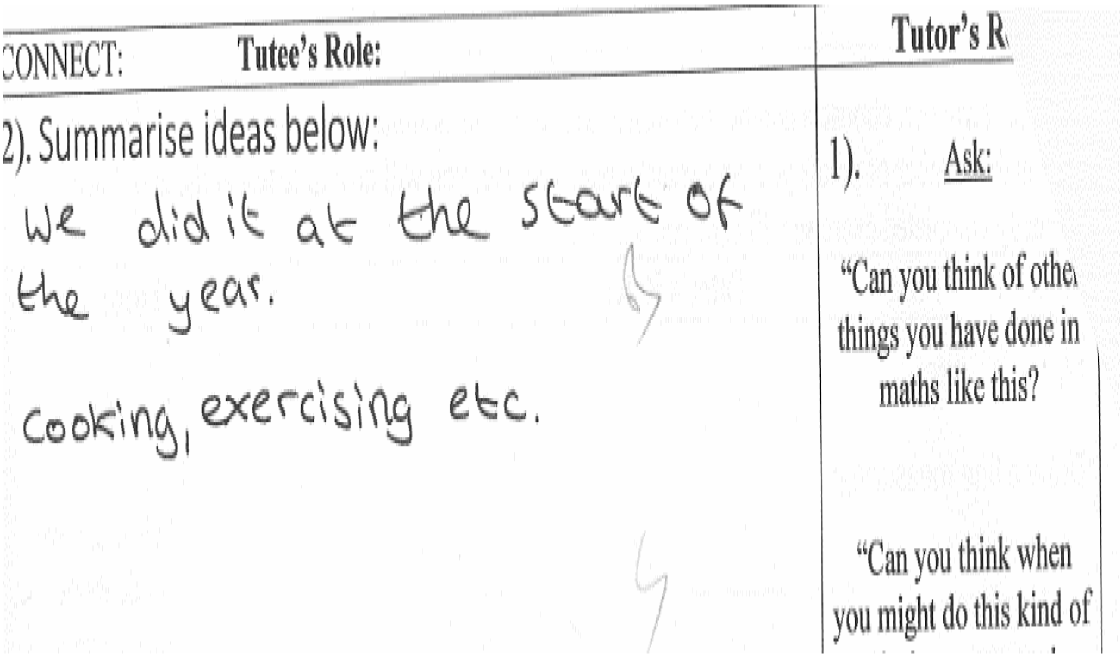
Direct only by nods.

If correct, praise *kindly*.

Well Done!
Very Good!
Excellent!
Brilliant

4). Respond with only hints and

Figure 19. *Example of a lesson in functions in the connect section*



CONNECT: Tutee's Role: | Tutor's R

2). Summarise ideas below:

we did it at the start of the year.

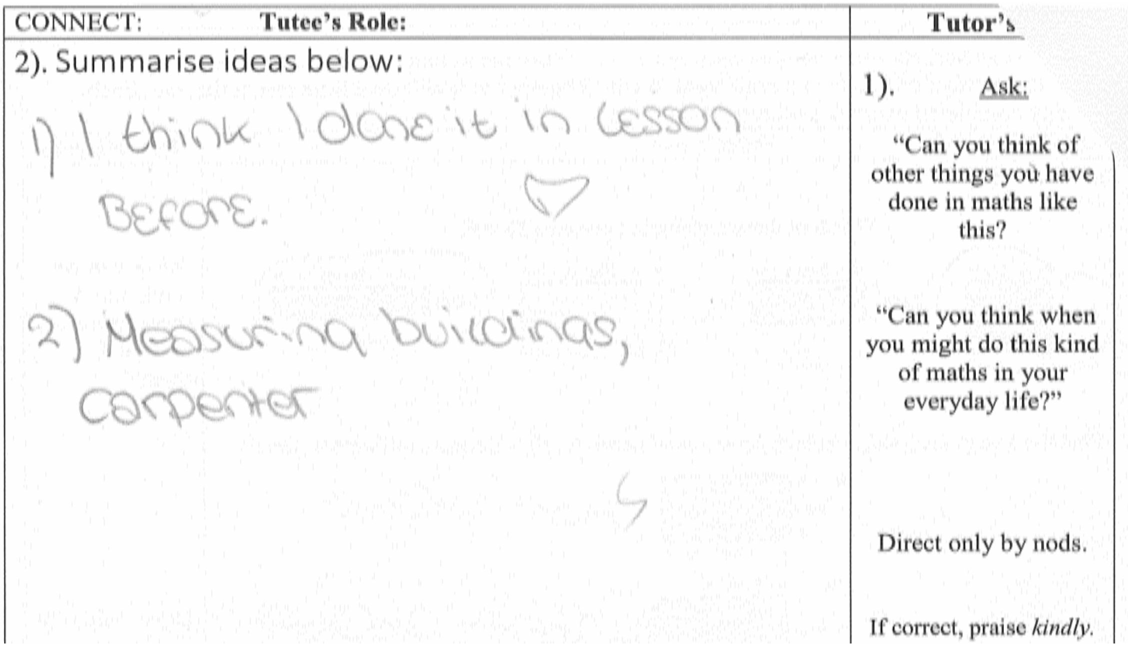
Cooking, exercising etc.

1). Ask:

"Can you think of othe things you have done in maths like this?

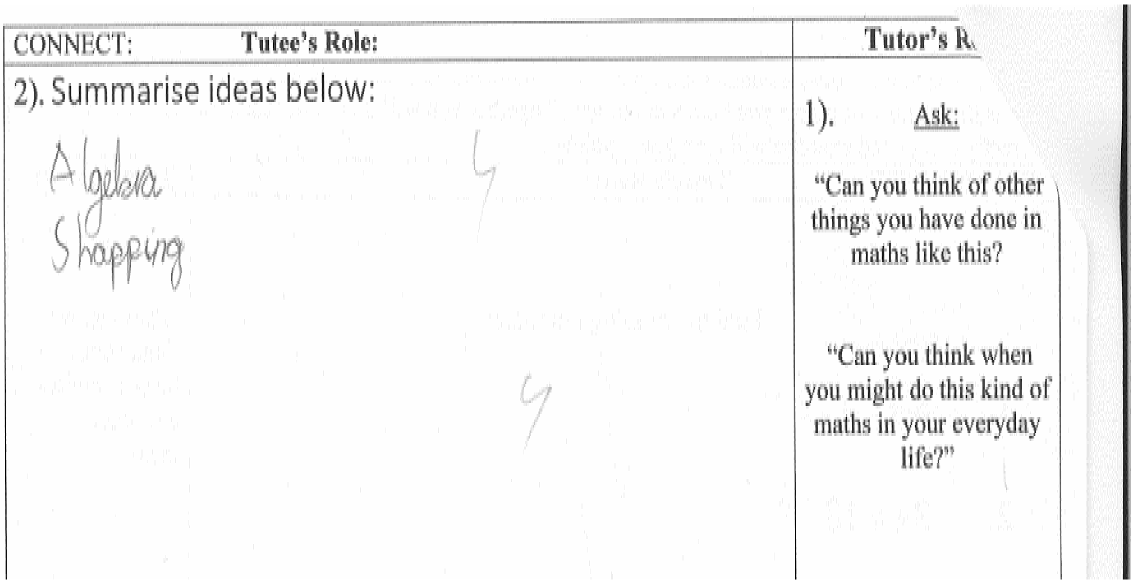
"Can you think when you might do this kind of

Figure 20. *Example of a lesson in measurement in the connect section*



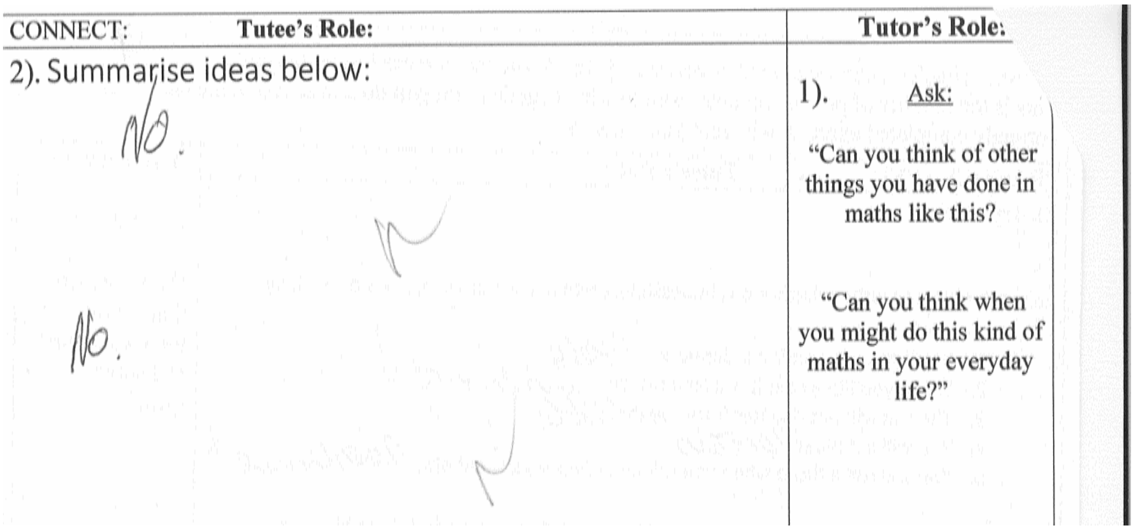
CONNECT: Tutee's Role:	Tutor's
2). Summarise ideas below: 1) I think I done it in lesson before. 2) Measuring buildings, carpenter	1). Ask: "Can you think of other things you have done in maths like this? "Can you think when you might do this kind of maths in your everyday life?" Direct only by nods. If correct, praise *kindly*.

Figure 21. *Example of a lesson in area and volume in the connect section*



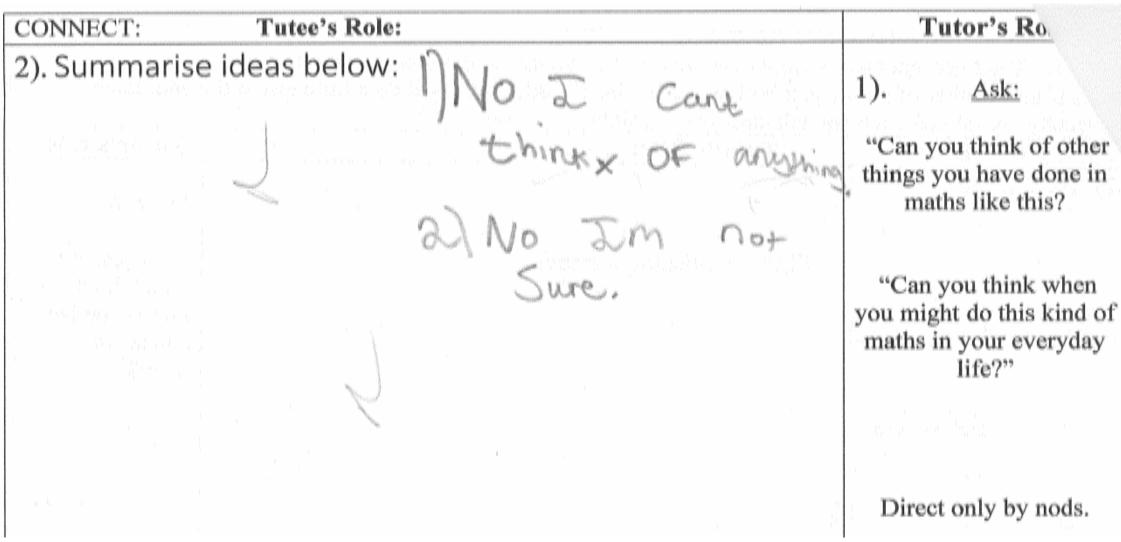
CONNECT: Tutee's Role:	Tutor's R
2). Summarise ideas below: Algebra Shopping	1). Ask: "Can you think of other things you have done in maths like this? "Can you think when you might do this kind of maths in your everyday life?"

Figure 22. *Example of a lesson in functions in the connect section*



CONNECT: Tutee's Role:	Tutor's Role:
2). Summarise ideas below: No. No.	1). Ask: "Can you think of other things you have done in maths like this? "Can you think when you might do this kind of maths in your everyday life?"

Figure 23. *Example of as lesson in probability in the connect section*



CONNECT: Tutee's Role:	Tutor's Ro
2). Summarise ideas below: 1) No I Cant think of anything 2) No Im not Sure.	1). Ask: "Can you think of other things you have done in maths like this? "Can you think when you might do this kind of maths in your everyday life?" Direct only by nods.

Figure 24. *Example of a lesson in function in the connect section*

The following three tables provide information regarding the quality of the statements made by students in the connect section for each of the three schools:

Table 22. *School A: Examples of students' answers in the connect section*

Lessons	*Specific*	*Broad**	*Negative**
Training Equations	- If you worked in a shop you would have to work out how much money you earned. – Money question, £1.50-.23p=1.27. – When counting money. – Money, accountant, shop owner, algebra, mathematicians, builder. – Money, job.		- No.
Lesson 1 Number pattern/ sequence	- Locations coordinates, area, perimeter. - A job in a bank. – Accountant, done sequences in tests. - Owing money. – Adding money, temperature. –Adding up money, recipe when baking, adding and taking away. – You could use it with money to work out how much you earn.	- I have used these things in lessons - I have done this in maths	
Lesson 2 Fraction	- Dividing up a pizza with your friends. - Joe has a cake, he shares it with three friends, how much of the cake would each friend get? - Pizza fractions. – Revision for the SATS, sharing sweets. – Fractions, pie maker. – When I'm doing percentages, cutting a cake. –Dividing slices of cakes, pie charts, we do fractions in maths. – Homework, lessons, tests, Sat tests. - Putting things into groups/sorting.	- In class	- No
Lesson 3 Measures	– I tell the time at home to see if I'm late for school, If I was making a project to make a wooden car I would need to measure the wood to the exact length. If I was going on a trip I would need to measure how long it would take to get there. If I was doing sewing I would need to measure the length of thread. - To build or measure something to see how long it would take, to see if you had enough time to do it. - Architects. – I use this when I tell the time. -Sats lessons, time/to no be late, building houses, how may minutes to 8 o' clock. –Builder, teacher, joiner, doctor. – Every job, banking science. – Running distances, bus timetable.	- In class - Use time every day.	- I don't think I have done these before.
Lesson 4 Shapes	- If I worked in a farm I would need to measure the length of the field to see how many sheep I could fit in. - Architect, Preparing for the SATS. – Creating towers with building blocks when I was younger. -Builder, bricks, size of a room, area + perimeter in a		- No

	maths lesson. - What is the area of a triangle, Sats. – Recognising shapes, drawing and measuring shape.	
Lesson 5 Data interpretation /representation	- Student shows a table of shop items with prices. -Pie charts and bar charts, working out how much you spend in shops.- Shopping, survey. – Job, math tests, Sats. -When working out how much money you have earned after a certain period of time, (student shows a graph representing the speed of a car in hours).	- Yes
Lesson 6 Written method	- Revision for SATS, Money. –Shop keepers, how much you are spending. Doing number lines and columns in maths. -Adding money, subtracting money. -Adding money, .40 + 2.10=£2.50, take away money 2.40-2.10=£.30, dividing money share it to make it equal. - Adding and taking away decimals using different mathematic methods, column, mental, number line.	

Note. **School A did not have many broad or negative comments.*

Table 23. *School B: Examples of students' answers in the connect section*

Lessons	*Specific*	*Broad*	*Negative*
Lesson 1 Mean Median Range	- Yes, for when you are counting money, seeing what coins you have. - Learned mean, mode, median and range better than before. - This is range, mean, mode and median. - Mode/most, medium/middle, range/the difference between the highest and the lowest. - Going shopping. – Maths teacher, till. -When you are adding things up. -I learned how to calculate the range. - Range, take the lowest number from the highest.	- I already knew the rest. - Yes. - This term.	- No - No I can't think of anything.
Lesson 2 Data Interpretation and Representation	- Yes, in a test and in our workbooks. If you had sweets and you wanted to divide them with friends, or if you had a party and you wanted to find out which food should go to each person. - They are money charts to show things like numbers, people, time. -I can now read a pie chart better. - ICT teacher, math teacher, business. - Drew graph, discussed ideas for work. - What a picture graph is. -We have learned how to put data into a graph using information that has been given.	-lots of other things.	
Lesson 3 Factors Multiply Fractions	- With factor I see if I can divide the numbers in a chronological order. With fractions I divide as much as I can. -Simplify fractions, divide highest common factor - Maths teacher, sweet shop, most shops, computer, science. - Test, job. I learned how to simplify fractions better. - Counting money. - Times tables, math teacher, IT teacher, Science. - I used the multiplication chart to help.	- Yes, but you will probably not realise	-No

Lesson 4 Sequences	- I learned how to write sequence rules better.	-	Don't understand questions -Nothing -No I can't.
Lesson 5 Mental methods Multiply/Divide	- Shop till, math teacher, PE teacher. - Yes I learned it the other day; x by decimal number. - Told me what factors mean. - Math teacher, guessing method column.	- Yes - Could already do it.	- Don't have a clue. -Nothing
Lesson 6 Equations	- I used multiplication - Math questions on fractions, in shops, business. - I learned how to do fraction algebra. - Maths teacher.	- Yes, in a lesson. - Trial and error. - I used good ideas.	- No

Table 24. *School C: Example of students' answers in the connect section*

	Specific	*Broad*	*Negative*
Training Fractions	- You would use it at a supermarket. - Shopping, money, exams, reception. - Money, shopping, wages, bills. - Adding, decimals. – I use maths in shops. - Prices and adding. – Jobs. – Taking away decimals. - Equations, shops, sums, - If someone was a carpenter.	- Everything in the front I have done in maths. - Everything really - I have done these questions in my math lessons.	- No - I've never done this before.
Lesson 1 Measures	- Distances and conversions, converting a distance. - Cooking, measuring things. - When you measure things. – Cooking, transport, shopping. - Shapes, areas, perimeter, building. - Units converting. – Exercises. - Doing measurements. – Shops, measurements, money. – Area and perimeter of shapes, building things. – Shapes, chef, shops, builder. – When we are cooking.	- We did it at the start of the year. - I have done this before - Yes I did this on my maths class.	- No - don't know
Lesson 2 Enlargement	- You could do it in art. - Architect – Planning. - I have been completing angles and factors. - Scale drawing. – Construction, architects	- In a lesson	-No - No I can't think of anything
Lesson 4 Areas	- Architects and fabricators, because they can see if they can fit things into rooms and what they are fabricating. – Building homes, garden planner, interior designer. - Shape work, building a garden. - Architect, builder. – Areas, perimeter and building. - Yes I have done this before, this is area. - Flooring. – Garden, building. – Carpet, flooring. Shapes, measuring areas of rooms. - Measuring buildings, carpenter.	- I have done this before in math. - Yes, lots of thing in the year. - I think I done it in lesson before. - Yes, maybe.	- No.

Lesson 5 Probability	- Probability of picking an even number from 1-20, 2,4, 6, 8, 10, 12,14, 16, 18, 20, even chance. - Probability, lottery. - When you do fractions, buying something. - If you toss a coin.	- Yes I have done this before. - Yes - Year 7.	- No I can't. - No I haven't done it in maths. - Nope.
Lesson 6 Fractions	-Going to a shop, e.g., I have £10 I buy milk for £1.25, and chocolate bar for .50p, I need to take away £1.25 and .50p from £10, change £8.35. - Banking, teacher. - I learned how to add and subtract, and finding missing numbers. -Have done algebra in the past. - Accountant. – Algebra, shopping.	- Yes - Yes we have done it before. - I have done it in maths. - I have done this before, last lesson. - Yes I have done this before.	- No I can't. - You wouldn't use this in everyday life.

Lesson 3 on transformations were not collected by the teacher

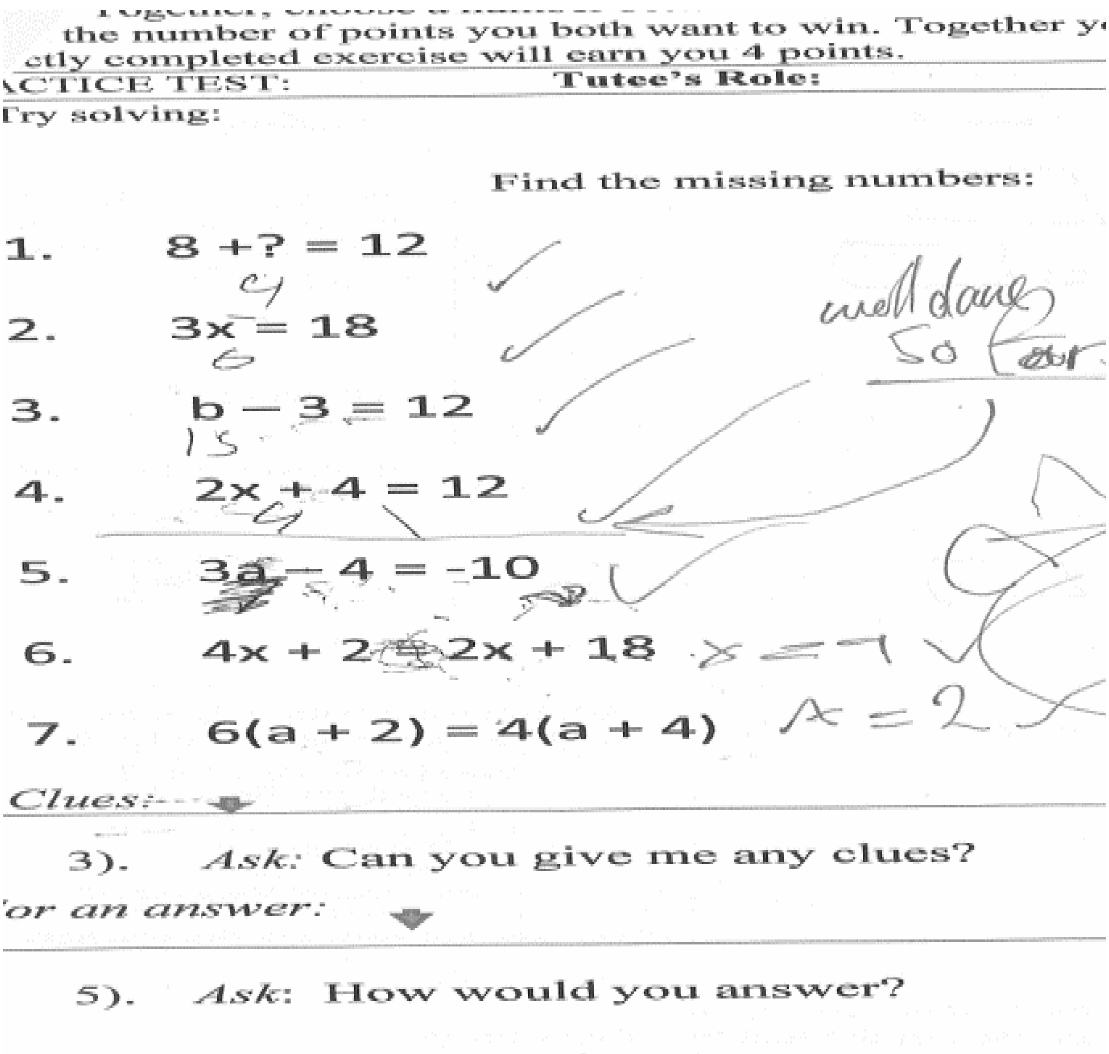
6.5.6 Questions 9, 10, 11, and 12: Quantity of different feedback types (Indicators, 9, 10, 11 and 12)

Feedback takes many forms and, this section looks at the following: total feedback, feedback by ticks/crosses in the 'Practice' part of the ICAT, feedback by ticks/crosses in the 'Connect' part, feedback in terms of checking if the goal was achieved.

The feedback indicators aimed at measuring the following areas: the level of cognitive engagement, the level of meta-cognitive engagement (due to the overall structure of ICAT), the level of interpersonal interdependence, and the level of cross-ability.

According to Higgins, Kokotsaki and Coe (2011), and Higgs, et al., (2014) review of meta-analyses, effective feedback, such as quick, personalised and elaborative feedback, provides the highest performance gains. Feedback by ticks is not necessarily elaborative. However, in this context it was quick and personalised. Higgins, (2011; 2014) report that peer tutoring was one of the leading interventions, not only in enhancing performance, but also in terms of accommodating effective feedback.

The following figure is a form of feedback by ticks:



the number of points you both want to win. Together y
ctly completed exercise will earn you 4 points.

CTICE TEST: Tutee's Role:

Try solving:

Find the missing numbers:

1. $8 + ? = 12$ 4
2. $3x = 18$ 6
3. $b - 3 = 12$ 15
4. $2x + 4 = 12$ 4
5. $3a - 4 = -10$
6. $4x + 2 = 2x + 18$ $x = 7$
7. $6(a + 2) = 4(a + 4)$ $x = 2$

well done
50 four

Clues:

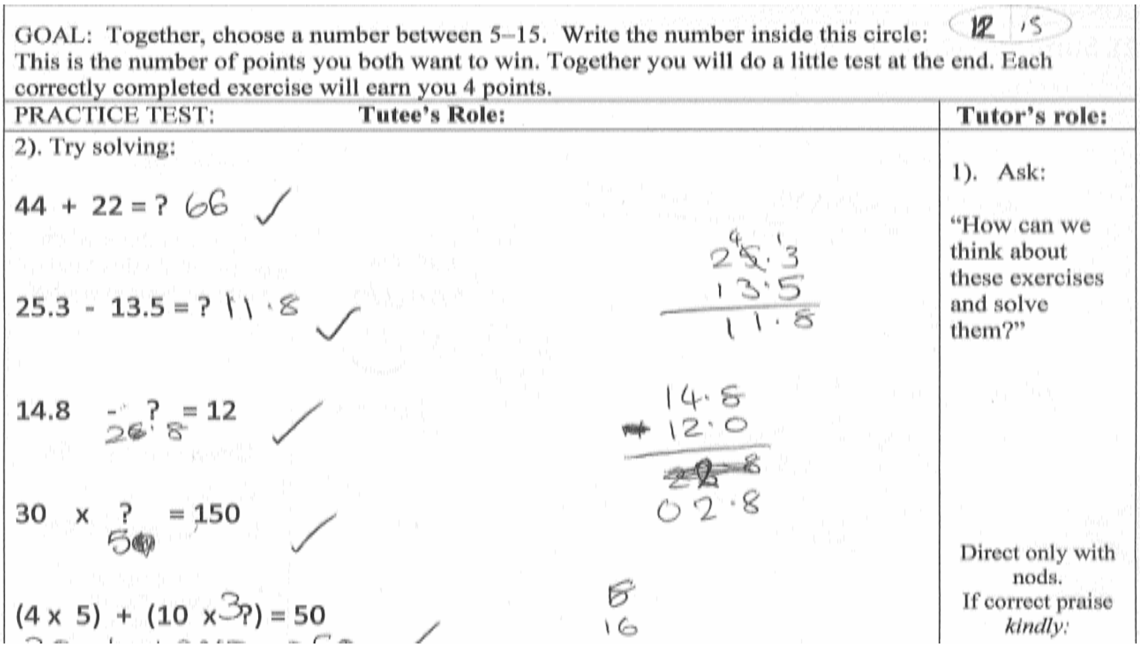
3). *Ask:* Can you give me any clues?

or an answer:

5). *Ask:* How would you answer?

Figure 25. *Feedback by ticks/crosses*

The following is an example of students checking if they had achieved their aim:



GOAL: Together, choose a number between 5–15. Write the number inside this circle: 12 15
This is the number of points you both want to win. Together you will do a little test at the end. Each correctly completed exercise will earn you 4 points.

PRACTICE TEST: Tutee's Role:	Tutor's role:
2). Try solving: 44 + 22 = ? 66 ✓ 25.3 - 13.5 = ? 11.8 ✓ 25.3 13.5 11.8 14.8 - ? = 12 26.8 ✓ 14.8 12.0 02.8 30 x ? = 150 50 ✓ (4 x 5) + (10 x 3?) = 50 8 16	1). Ask: "How can we think about these exercises and solve them?" Direct only with nods. If correct praise *kindly:*

Figure 26. *Checking if goal is achieved*

Students' sum of feedback was arrived at by simply counting each item of feedback for each student in each group for all their lesson materials, and then adding up the total for the group.

6.6 Observations

Observations were conducted in the following way:

General school observation paid attention to issues such as classroom size, seating, available materials, and whether the teachers were knowledgeable about ICAT and monitored the classes properly.

In terms of pair observation, which consisted of the majority of time, the following was the procedure: First the name of the teacher and the topic of mathematics was recorded on the observation sheet. The mathematics areas observed were: for school A 'written mental methods' and 'data interpretation and representation', for school B 'factor multiplying' and ‘equations’, for school C ‘area and volume’ and ‘equations’. The observations took place during the second half of peer tutoring, so that the students had enough time to master the ICAT framework. For school A the observations took place during the last two ICAT lessons, Monday mornings, both at 9am. For school B the observations also took place during the last two ICAT lessons, Thursdays at 1pm. For school C the observations took place during the fourth and the sixth session, Wednesday mornings at 10am.

To begin, the researcher was first introduced to the class in an informative manner and the class was informed that the researcher would just walk around the classroom to learn more about how the students worked together. If any of them did not want to be observed then they just had to say so. The older students, the tutors, were then politely asked to all sit on the right hand side of their tutee, in order to make the observations easier for the researcher. During the observation the researcher maintained a positive facial expression. During the pair observations the researcher walked slowly behind each pair, stopping two steps away, so that the students were not disturbed, in order to minimise the chances for the students to alter their behaviour.

Each pair of students was observed in a structured manner throughout five windows of time. For a Year 8 student tutoring a Year 6 student, and a Year 9 student tutoring a Year 7 student each window lasted approximately 30 seconds, 5 seconds to adjust, 15 seconds to observe and 10 seconds to record. For the school of Year 9 student tutoring a Year 7 student the window was longer, as there were only 9 pairs per class, 40 seconds long. On a few occasions the students had to work in groups of three as their tutor or tutee was not present on the day, those students were not observed.

Figure 27 below shows the observation items, influenced by Social Interdependence Theory (Argyle, 1976; Allen & Feldman, 1976; Johnson & Johnson 1975; Johnson 1990). However, some came from other related theories such as Social Constructivism (Fitz-Gibbon

2000, Roscoe & Chi, 2007; Topping, 2011). All together they measured the level of interdependence, meta-cognition and cognition. The only difference being that the observations took place at the individual level, concentrating more specifically on tutors' and tutees' questions and answers rather than overall pair interaction as previously conducted by Topping, (2011). Table 25 on the next page provides a more detailed picture of what each indicator aimed at measuring and their respective observatory method.

School:	Teacher:			Date:	
Topic:					
Pair's Name:	Tutee: Tutor:				
	1	2	3	4	5
Goal Interdependence					
Tutor B-Language/T-Voice					
Tutor Praises Correctly					
Tutor M/C Questions					
Tutee Self Corrects					
Tutee Connects/Categorises					
Tutor Questions					
Tutee Answers					
Tutee Questions					
Tutor Explains					
Task Engagement					
In-Audible					

Figure 27. *Pair observation sheet*

Table 25. *Area of measurement and method by observation indicator*

<table>
<tr><th>Indicators</th><th>Method</th><th>Measuring</th></tr>
<tr><td>Goal Interdependence</td><td>The pair had written down the points they wanted to achieve. The point could have been written on the ICAT format or a separate paper if there was one. If a pair did not write down a point they wanted to achieve, however discussed and agreed on this was also marked as goal interdependence.</td><td rowspan="2">Goal interdependence and interpersonal interdependence.</td></tr>
<tr><td>Tutor B-Language/T-Voice</td><td>Tutor had a relaxed body language and tone of voice, and sat close to his/her tutee.</td></tr>
<tr><td>Tutor Praises Correctly</td><td>When tutor praised, the praise had to be warm and genuine. When the praising was associated with a smile this was marked as correct praising.</td><td>Interpersonal interdependence and cognitive engagement (as praising is a form of feedback).</td></tr>
<tr><td>Tutor M/C Questions</td><td>Tutor read the questions which were written in the ICAT, or asked how questions. A closer eye for meta-cognitive questions was paid to the Connect section which was designed specifically to make students reflect, connect, categorise.</td><td rowspan="3">Meta-cognitive engagement.</td></tr>
<tr><td>Tutee Self Corrects</td><td>Tutee reflected back on own work to correct him/herself.</td></tr>
<tr><td>Tutee Connects/Categorises</td><td>Tutee connected or categorised ideas to other previous ideas or to real life. Again, this usually took place at the Connect section. However if in the other sections the tutee reflected back to previous knowledge this was also marked as connect/categorise.</td></tr>
</table>

Tutor Questions	General questions from the tutor, taking place throughout the entire peer tutoring interaction.	Cognitive engagement.
Tutee Answers	General attempts to answer by the tutee. Tutee answers were marked if the tutee provided a verbal answer, or started to write straight after the tutor explained something.	
Tutee Questions	Tutee asked tutor a question	
Tutor Explains	Tutor provided spoken explanations; the explanations could have been conceptual or procedural.	
Task Engagement (alone)	Tutee worked alone on the exercise, while tutor watched.	
In-Audible	Too much noise coming from the rest of the class when observing a pair.	Noise.

Table 25, shows that most of the indicators measure cognition. The remaining measured meta-cognition and interdependence levels between the tutee and the tutor. Certain indicators such as ‘praise’ fell into two categories, interdependence as well as cognition, since praising stimulates both affection (interpersonal interdependence) as well as serving as a feedback mechanism for knowledge construction/acquisition (cognition).

6.7 Chapter conclusion

This chapter has looked at how each instrument used in this research was developed. Half of the instruments were researcher made; they were those measuring mathematic performance and group-level analysis to investigate social interdependent variables (and school level implementation/fidelity variables). The remaining instruments, the questionnaire measuring attitudes, social and academic, and the observation items, were all researcher modified instruments.

All the instruments measured different aspects of social interdependence: For example, the observations were mainly to be used as implementation fidelity. They measured students' behaviour in terms of goal and interpersonal interdependence, cognitive and meta-cognitive elements at the student level, and general teacher observations at the classroom level. The questionnaires on the other hand measured attitudes, again both social and academic. Finally, in terms of analysing the lesson materials, the instruments used to investigate tutees' group differences also measured and aimed at exploring the level of social interdependence, cross-ability, cognitive and meta-cognitive elements. Lesson materials were also used to investigate implementation fidelity at the school level; the indicators used for this analysis were the same indicators as those identified for the tutee group comparison purposes.

The strength of using the lesson materials is that one could gain a view of the learning process during the entire peer tutoring process, rather than concentrating on one or two lessons as was the case with observations, or with the 10 minute attitude responses, as was the case with the attitude questionnaires.

7 Analysis and Reporting Findings

7.1 Introduction

This chapter concentrates on two topics: a) explaining how the analyses were carried out, and b) the findings. It does so for three areas: 1) implementation comparison of the intervention across the schools, 2) students' mathematics performance and attitude variables 3) group comparison of higher performing and lower performing tutees' lesson materials for different social interdependent variables.

7.2 Analysis

7.2.1 Implementation

As mentioned previously in the method and instrument development chapters, implementation was analysed by concentrating on two areas: lesson materials and observations.

Percentages were used to investigate the extent to which the schools implemented ICAT according to programme specifications. To compare implementation across the three schools, for both lesson materials analysis and observations, one way ANOVAs were applied.

7.2.2 Performance and attitude analysis

Mathematics performance and attitude variables for the quasi-experimental design school data were analysed via two statistical methods: Analysis of Covariance (ANCOVA) and t-test residual gain analysis. Both ANCOVAS and t-test residual gain analysis were used with pre and post-test quasi-experimental designs, due to their strength to control for the pre-test differences. Both tests were used to analyse differences between the control and the peer tutoring group for the following variables: mathematics performance, and all the attitude variables described in the instrument development chapter such as, mathematics self-concept, mathematics enjoyment, choice in how mathematics was done in class, relating to people in mathematics classes, and social self-concept. SPSS 20 was used to analyse the data. A combination of both statistical tools was applied in order to strengthen the analysis, and provide a further form of data analysis triangulation within the statistical analysis for the thesis**.** The single group school data for performance and attitude variables were analysed via a dependent t-test. SPSS did not conduct one sided tale t-tests, which is required when testing rather than exploring hypothesis; therefore in order to take account of this shortfall the p value coefficients were multiplied by 2.

The effect sizes for the quasi-experimental designs were manually calculated by taking the ANCOVA coefficients and using Cohen's d technique with the square root of the MS Error as the denominator:

$$Effect\,\mathrm{size} = \frac{Mean\ of\ gains\ experimental - Mean\ of\ gains\ control}{\sqrt{MsError}}$$

Equation 5. *Cohen's d for ANCOVA*

For the independent t-test residual analysis and the dependent t-test the effect size was calculated via using Hedge's g (1981) pooled standard deviation method, mentioned in section 3.6.2:

Firstly, for the performance data collected from the quasi-experimental schools, ANCOVA was also conducted at the class level, in order to ensure that there was no class effect on performance and thereby no statistical difference in mathematics performance data *between* peer tutoring classes or *between* control classes for the two schools which contained quasi-experimental designs. In other words, in school A and in school B, the peer tutoring classes were compared to one another within each school, as were the control classes. Consequently, this showed whether a particular statistical difference in performance within the control and peer tutoring groups was due to the actual peer tutoring intervention or whether there was a classroom effect within a particular group which would have biased the results.

Also, for every statistical comparison, statistical requirements/assumptions were checked, and will be reported and interpreted if found to be violated.

Finally, for the quasi-experimental school data, Bonferroni correction was used on ANCOVAs. Bonferroni correction provides more conservative results (Field, 2003) in order to counteract Type One Errors (failing to reject the null hypothesis), which are likely to occur from repeated analysis of the same population (Field, 2003). Consequently, since the same sample was used for different analyses, such as performance and attitude, Bonferroni correction was considered to be more appropriate. This is due to the fact that ANCOVA, having one dependent variable at a time - rather than MANCOVA, having multi-variables simultaneously – was used for the analysis.

The reason for not having used MANCOVA, looking at performance and attitude all in one statistical analysis method simultaneously, was due to the notion that MANCOVAS are not able to cope with correlated dependent variables (Field, 2003). The decision not to use MANCOVA was based on the initial findings that were produced by the EFA and CFA analysis on the attitude instruments with the trial school data, which showed that many attitude variables were correlated to one another.

Also, for the t-test residual gain analysis, residual gains derived from the regression slope of pre and post-test performance data, the analysis took the regression slope which best fitted the data, such as linear, quadratic or cubic; so that the degree of bias that would derive from the regression slope was minimised with the best line fit. Linear, quadratic and cubic refer to different regression relationships between the pre and post-test variables. The three methods can be used to search for the best model that fits the data scores between the variables. For example, if we were to simply impose a linear model on a data set which was best predicated by a cubic regression line, the R squared would have been lowered, the residual scores consequently would have been higher, and this could have added bias to the final results in terms of what the real gains were. [13]

7.2.3 Tutees lesson materials group analysis

For school A and C, the first lesson materials, the training lesson on equations, also entered the group analysis in order to give a better picture of the intervention. While for school B the training lesson scripts were not collected by the teachers, therefore they could not be included in the analysis. Also, the lesson materials on transformations were not collected by the teachers from school C, therefore they could not enter the analysis. The social interdependent indicators were then applied to analyse the lesson materials of the higher and lower

[13] An example of the steps and SPSS procedures for analysing the data via ANCOVAS and t-test of regression residuals can be found in the Appendix.

performing peer tutoring groups to explore whether there were any differences on the indicators.

For every question in this section an independent t-test was used to analyse the data. Again since SPSS did not provide one tailed t-tests the p value was multiplied by 2. The next section reports the findings.

7.3 Findings

This section reports the findings including the attrition rate. Specifically on the main findings four inter-related findings are reported:

1) The intervention's fidelity extent based on both, peer tutoring lesson materials and classroom observations for each school, as well as a comparison between schools.

2) The intervention's findings on performance.

3) The intervention's findings on attitude variables.

4) And the findings on group comparisons of the higher and lower performing tutees' peer tutoring lesson materials in regards to the social interdependent indicators.

The findings from the first part, analysing the intervention implementation extent via comparing lesson materials and observations across schools, serves three purposes:

- To investigate the extent to which the schools stayed true to programme specifications.
- Also, both peer tutoring lesson scripts and observations were used to further investigate the question *how* ICAT worked or did not work for a particular school by comparing the schools to one another.
- The final aim was to further add to data analysis triangulation and see which indicators seem to overlap between analysing peer tutoring lesson scripts and analysing observations.

The finding from the second analysis, answers the question: ‘Did the ICAT as shaped by social interdependence produce the expected results on performance?’.

The findings from the third investigation, on *affective/attitude* variables such as ‘mathematic self-concept’, ‘mathematics enjoyment’, ‘social self-concept’, ‘relating to others in mathematics classes’, and ‘choice of how mathematics is done in class’, answer questions regarding the impact of the intervention on both *social* and *academic* psychological variables.

The findings from the final analysis, comparison of the higher and lower performing groups of tutees’ lesson materials in regards to social interdependent indicators, serves two aims: Firstly, the comparison findings serve as a form of triangulation, adding more depth to the investigation of the learning process under ICAT. Secondly, concentrating on multiple social interdependent indicators generally paints an enhanced picture of the learning processes used during ICAT and *how* peer tutoring worked or did not work for particular groups.

7.3.1 Attrition rates

Table 26 shows the attrition rate for each group and data analysis type for each school.

In terms of performance data attrition rates, some of the groups with the lowest attrition rates were the Year 7 peer tutoring group in school B, 3%; and the Year 6 peer tutoring and control groups for school A, 17% and 16% respectively. The highest percentage attrition rates were for school B Year 9 students for the control group, with an attrition rate of 60%. The second highest was that of school C Year 8 students, with an attrition rate of 41%. School B Year 7 control group also showed a high attrition rate, 34%.

On attitude questionnaires, the groups with the lowest attrition rates were all those of the Year 6 and 8 students in school A, for both control and peer tutoring, as well as the Year 7 peer tutoring students in school B. The highest attrition rates were for school C Year 8 students, with a 43% attrition rate. The second highest attrition rate on attitude questionnaires was that of school B Year 9 students, with a 40% attrition rate, as well as Year 7 students for the control group in school B, 28% attrition rate. The Year 10 students in school B also showed a high attrition rate, 25%.

Finally in terms of lesson materials the lowest attrition rates were present in school B, 3%, while school C showed an attrition rate of 41% and school A, 57%.

The main reason for the high attrition rates for performance and attitude instruments is that these attrition rates are based on combining both pre and post-tests to represent the gains for each student. Many students were missing either during the pre-test or during the post-test. Secondly, the highest attrition rates were found within the control groups, in other words in groups in which the teacher and the students were not very motivated to use the instruments since they did not carry out peer tutoring.

The high attrition rates for the lesson materials in school A can be explained by a lack of communication between the teachers regarding what was required of them in terms of data collection, as only half of the classes, two out of 4 classes, reminded the students to write down their names on each lesson material. School C's high attrition rate on lesson materials can be partially explained because the school as a whole took a slightly more independent route to implementing and conducting peer tutoring, which deviated from the research design and requirements initially agreed.

Table 26. *Attrition rate % for each data collection type by group and school*

Schools	*Year*	*Groups*	*Performance tests %**	*Attitude questionnaires %**	*Lesson materials %**
School A Year 8 students tutoring Year 6 students	8	Peer Tutoring		6 missing 6/54=11%	31missing 31/54=57%
		Control		3 missing 3/62=5%	
	6	Peer Tutoring	9 missing 9/54=17%	8 missing 8/54=15%	
		Control	9 9/58=16%	10 missing 10/58=17%	
School B Year 9 students tutoring Year 7 students	9	Peer Tutoring	11missing 11/36=31%	8 missing 8/36=22%	1missing 1/36=3%
		Control	25 missing 25/42=60%	17 missing 17/42=40%	
	7	Peer Tutoring	1 missing 1/36=3%	2 missing 2/36=6%	
		Control	13missing 13/39=34%	11missig 11/39=28%	
School C Year 10 students tutoring Year 8 students	10	Peer Tutoring		20 missing 20/80=25%	30 missing 30/74=41%
	8	Peer Tutoring	30 missing 30/74=41%	32missing 32/74=43%	

**In order to arrive at the attrition percentage rates the number of missing students was divided by the total number of students within each group for each data collection type.*

7.3.2 School level implementation analysis

This section reports implementation findings for both the lesson material analysis and observations. This is done in two ways; by evaluating the extent to which the schools' implementation of ICAT reflected the interventions' specifications, and secondly by comparing the schools to each other based on different data collection methods.

Implementation of ICAT according to program specifications:

This section reports the extent to which the schools implemented ICAT according to the programme.

Lesson materials. In terms of implementation of the intervention by looking at the lesson materials, school A came the closest to implementing the ICAT according to the programme specifications at 85%. Schools B and C showed an implementation of 63.5% and 64.88% respectively. Overall the element least implemented was 'lessons with student set goal', 58%; and the element implemented the most was 'lessons in which all exercises in the practice-test section were attempted', 79%.

Table 27 next page presents data from the analysis of the lesson materials.

Overall school/classroom observations experience. For school A the observations of the set up overall corresponded to the planned intervention. The classes were spacious, and there was enough space between the pairs. The teachers directed the pace of the peer tutoring as they were trained to do, guiding the pairs through each peer tutoring part. The lesson scripts indicated that the schools made use of the training lesson. The pairing of the students was conducted as planned in terms of same sex pairing.

Table 27. *Percentage of implemented lessons according to programme specifications*

Lesson Materials	*School A %*	*School B %*	*School C %*	***Average** % per indicator*
Total peer tutoring lessons attended by students with names on materials	149/161=93	152/245=62	206/308=67	**74**
Lessons with student set goals	145/149=97	60/152=39	80/206=39	**58**
Lessons in which all exercises in the practice-test section were attempted	116/149=78	120/152=79	164/206=80	**79**
Lessons with attempted exercises connect section	105/149=70	85/152=56	154/206=75	**67**
Lessons in which all exercises in the turn-taking section were attempted	121/149=81	106/152=70	120/206=58	**70**
Correct answers for practice test	779/957=81	632/986=64	1410/1759= 80	**75**
Specific answers connect section	363/382=95	92/130=71	124/256=48	**71**
Correct answers turn-taking section	785/942=83	651/967=67	819/1132=7 2	**74**
Average per school	**85**	**64**	**65**	

School B seemed to have a few major problems: One of the classrooms was slightly small. Even though there were only nine pairs the students were sitting very close to one another. Although the teacher did try his best to guide the pairs through each peer tutoring part, many of the students did not remain seated and caused noise. The pairing of the students was only partly conducted as planned; specifically the school had not entirely managed to secure same-sex pairing. Two additional drawbacks, both relating to the experimental design, were observed in School B. Firstly, one of the teachers who served in the peer tutoring group disclosed that he was a keen user of technology and that most of his classes incorporated interactive ICT programmes in order to raise students' attainment. Secondly, the school in general conducted mathematics formative tests every two weeks, with the aim of providing students with more feedback and raising their attainments. In other words, the classroom conditions in school B were not normal.

Regarding school C the observations revealed that overall the classroom sizes and pairs' seating were spacious. However, there were a few problems: Firstly, boys and girls were sitting in different rooms. This was not an issue in general. However, for the boys' observation there were three teachers in the room, the appointed teacher, the mathematics head teacher and the school's head teacher, showing extra effort in implementing peer tutoring. Overall, they managed to guide the peers appropriately. This was not the case for the girls' class observed at a later date in which there was only one teacher. However, for the boys' classroom, later on the day it became known that the school was being inspected by Ofsted, who were thought to be keen on peer tutoring interventions. Therefore, the schools' effort to manage peer tutoring effectively for the boys' class was very likely to have been an extra effort in light of the Ofsted inspection rather than a peer-tutoring motivated attempt.

Table 28 next page presents findings on general observations: Overall general observations reveal that only school B was not able to stick to the programme specifications, as one of the classrooms observed was not spacious enough and some of the desks were too close to one another.

The chart on figure 28, shows the extent to which student pairs implemented ICAT according to the programme based on 5 windows of observation. The pair observations illustrate that the elements least implemented from ICAT were goal interdependence, tutor praises correctly and tutee connects/categorises. Considering that there were 5 observation windows per pair, and the average for most variables is less than two points, pair observation would suggest that less than half of the ICAT programme specifications were implemented.

Table 28. *General school observations*

Observations topics	*School A*	*School B*	*School C*
Classrooms space	Enough space	One classroom too small	Enough space
Teachers knowledgeable with ICAT	Yes	Yes	Yes
Teachers aiding students	Yes	Yes	Yes
Desks organized appropriately	Yes	Mostly	Yes
Materials (pencils, rulers, etc.)	Yes	Yes	Yes
Same sex pairing	Yes	Mostly	Yes

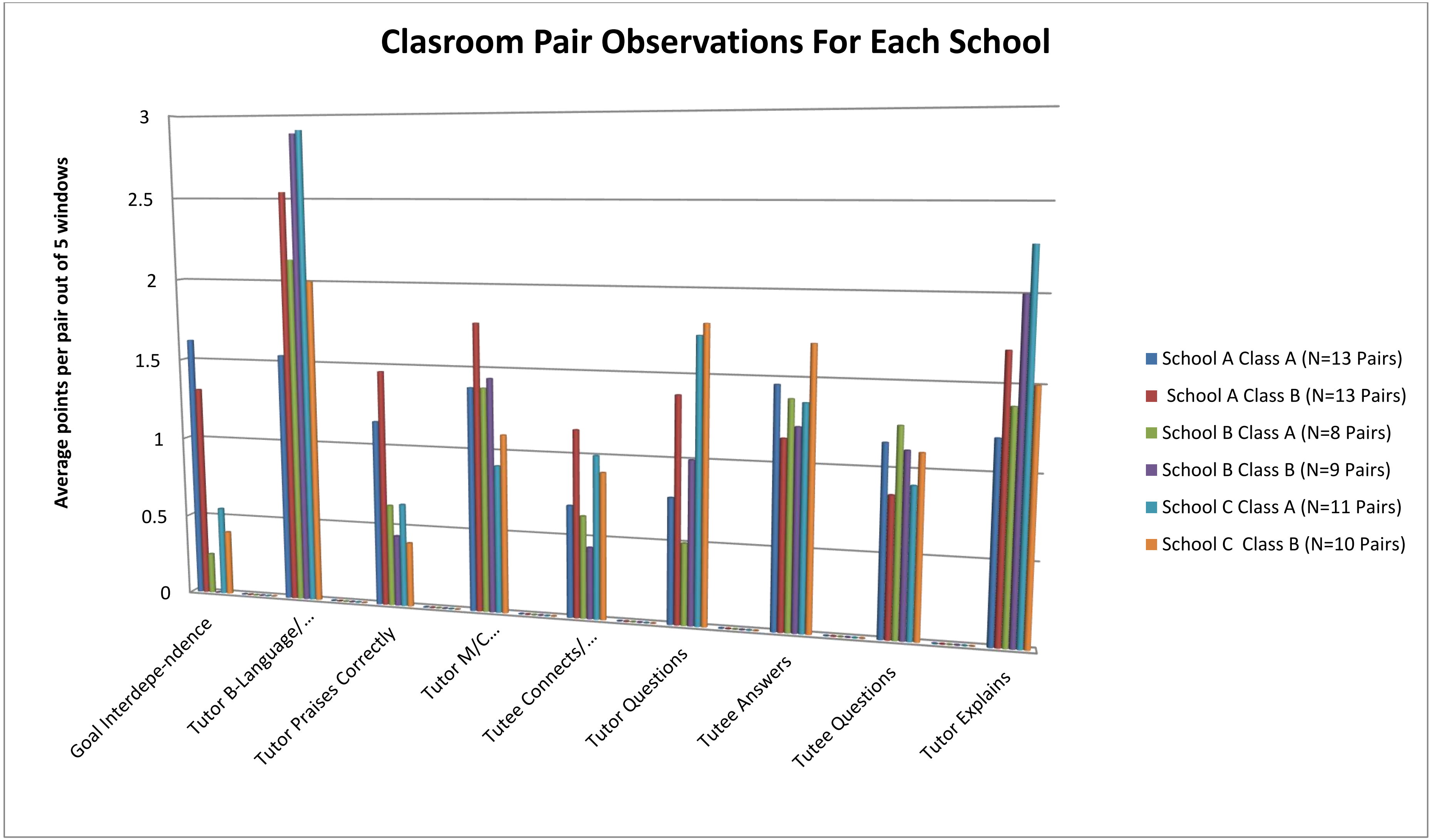


Figure 28. *Classroom pair observation by school*

Comparing implementation between the schools

Lesson materials. The following section presents the findings on the extent to which peer tutoring was implemented appropriately across the three schools. Schools are compared on each implemented indicator/element, and effect sizes for each comparison are provided. The findings are reported in table 29, next page.

The implementation indicators with the highest effect sizes from table 29 were:

A) ‘Lessons with student set goals’; for this element the effect sizes favoured school A. For the comparison of schools A-B, the effect size was 3.10, whereby school A mean was 6.39 (Stud78) and school B mean 1.71 (Std 1.84); for the comparison of schools A-C, the effect size was 3.81, where school C mean was 1.59 (Std 1.44).

B) ‘Quality of answers in turn taking test’; again the effect sizes favoured school A. For the comparison of schools A-B, the effect size was 2.29, in which school A mean was 34.5 (Std 5.79) and school B mean 18.6 (Std 7.58); for the comparison of schools A-C, the effect size was 3.21, in which school C mean was 15.9 (Std 5.80).

C) ‘Attempted exercises in the turn-taking section’; similarly for this area the effect sizes favoured school A. For the comparison of schools A-B effect size 2.33, whereby school A mean was 41.4 (Std. 5.21), and school B mean 27.8 (Std 8.34); for the comparison of schools A-C, the effect size was 1.87, with school C mean 25.6 (Std 7.46).

A similar picture emerged for all the remaining indicators, with the exception of the ‘feedback by ticks and crosses practice section’. For this element school A showed a mean 0, compared to school B mean 2.6 (Std 0.44), the effect size was 7.65. This effect size, however, was not fully reliable since SPSS deemed all school A positive values as outliers, and therefore they needed to be adjusted to coefficient ‘0’.

Table 29. *School lesson comparison*

Variables	*Schools*	*Mean*	*Std. Dev*	*F-value*	*p.*	*Effect Size*
Peer tutoring lessons	A	6.48	.67	32.22	.000	A-B 2.11
	B	4.43	1.12			A-C 1.86
	C	4.77	1.03			B-C 0.32
Lessons with student set goals	A	6.39	.78	91.48	.000	A-B 3.10
	B	1.71	1.84			A-C 3.81
	C	1.59	1.44			B-C 0.07
Attempted exercises practice test section	A	43.8	10.2	25.97	.000	A-B 1.87
	B	26.8	8.17			A-C 0.49
	C	38.8	10.1			B-C 1.29
Attempted exercises connect section	A	11.6	4.99	39.60	.000	A-B 2.27
	B	3.66	2.67			A-C 1.48
	C	6.61	2.63			B-C 1.11
Attempted exercises turn-taking section	A	41.4	5.21	35.95	.000	A-B 1.87
	B	27.8	8.34			A-C 2.33
	C	25.6	7.46			B-C 0.28
Quality practice test	A	34.1	5.00	56.46	.000	A-B 2.57
	B	18.0	6.96			A-C 0.39
	C	31.6	6.94			B-C 1.96
Quality connect section	A negative*	0	0	13.14	.000	A-B 0.78
	B negative	.43	.70			A-C 1.04
	C negative	1.43	1.69			B-C 0.74
	A broad	.52	.67	6.97	.001	A-B 0.05
	B broad	.57	.74			A-C 0.68
	C broad	1.41	1.54			B-C 0.60
	A specific	12.8	8.18	44.5	.000	A-B 1.90
	B specific	2.51	2.16			A-C 1.78
	C specific	3.59	2.47			B-C 0.46

Quality turn-taking	A	34.5	5.79	65.87	.000	A-B	2.29
	B	18.6	7.58			A-C	3.21
	C	15.9	5.80			B-C	0.41
Total types of feedback	A	29.3	17.9	28.39	.000	A-B	2.34
	B	2.26	3.44			A-C	0.70
	C	17.6	16.1			B-C	1.25
Feedback by ticks & crosses practice section	A*	0	0	31.21	.000	A-B	7.65
	B	2.6	.44			A-C	1.31
	C	7.52	7.07			B-C	0.93
Feedback by ticks & crosses turn-taking section	A	18.6	11.8	40.58	.000	A-B	2.51
	B*	0	0			A-C	1.12
	C	7.98	8.07			B-C	1.32
Students checking back if goal is met	A	3.04	2.06	45.44	.000	A-B	2.07
	B	.26	.44			A-C	1.69
	C	.73	.82			B-C	0.69

School A n=23, School B n=35, School C n=44. *After adjusting for the outliers the schools have a mean of '0', hence the standard deviation, skewness, kurtosis and normality are all '0'. In other words, there were few cases with positive values which acted as outliers, since the remaining of the cases were all '0', the outliers needed to be adjusted to the next value in order not to bias the results,

Finally, as illustrated by table 29, school C showed a better implementation than school B in most indicators. School B showed a better implementation than school C only in three areas: First 'quality of answers in the turn-taking section', with an effect size of 0.41, school B mean was 18.6 (Std 7.58) and school C mean 15.9 (Std 5.80). The second area was 'attempted exercises in the turn-taking section', with an effect size of 0.28, school B mean stood at 27.8 (Std 8.34) and school C mean 25.6 (Std 7.46). The third area in which school B did better than school C was 'lessons with student set goals', with an effect size of 0.07, with school B mean standing at 1.71 (Std 1.84), and for school C mean 1.59 (Std 1.44). Also, the indicator ‘quality of answers in connect section’ showed mixed results when comparing school B to school C. For the remaining 8 variables school C showed a better implementation than school B

Overall, school A outperformed schools B and C in terms of implementing peer tutoring, and school C outperformed school B.

School observation findings. This section concentrates on the findings of student pair observations. Table 30 provides an overall representation of pair observations:

Similar to the previous school level fidelity analysis of the lesson materials, the indicators with the highest effect sizes, and with a significant p value, were also those between schools A-B and A-C. With observations showing that school A was able to implement peer tutoring better.

The observed element with the highest effect size was that of ‘goal interdependence’, specifically students setting goals for their lessons in pairs, $F(2.61)=26.76$, significant at ($p<.001$). For this indicator school A ‘goal interdependent’ mean stood at 1.46, (Std 0.76), school B mean 0, and school C mean 0.48 (Std 0.41). For the comparison of schools A-B, the effect size was 2.47, for schools A-C the effect size 1.51. While for the comparison of schools B-C the effect size stood at 1.33; with post-hoc analysis showing a high significance for only the comparisons of schools A-B and A-C ($p<.001$), whereas schools C-B was significant at ($p=.037$).

The second highest effect size was for indicator ‘inaudible’, the noise and distraction level, $F(2,61)=8.01$, ($p=.001$). For this observation school A mean stood at 0.46 (Std 0.58), school B mean 1.35 (Std. 0.93), and school C mean 1.14 (Std 0.85). For the comparison of schools A-B the effect size was 1.39, ($p<.05$), and schools A-C the effect size 1.1 ($p<.05$).

School A also showed higher scores in relation to school B and C on ‘task engagement (alone)’ and ‘tutor-praises correctly’. Regarding the ‘task engagement (alone)’

F(2,61)=10.26, significant at ($p<.001$), with school A mean 3.11 (Std. 1.03), school B mean 3.58 (Std 1.06), and school C mean 4.33 (Std .58). For the comparison of school A-B the effect size was 0.45. However, post-hoc analysis showed this was insignificant. For the comparison of schools A-C the effect size stood at 1.42, post-hoc analysis revealed a significance level of ($p<.001$).

For indicator ‘tutor praises correctly’, F(2,61)=8.40, significant at ($p=.001$), with school A having a mean of 1.31 (Std 1.06), school B mean 0.47 (Std .62), and school C mean 0.48 (Std 0.51). In terms of effect sizes when comparing schools A-B there was an effect size of 0.92, post-hoc analysis showed a significance level at ($p<.05$), when comparing schools A-C the effect size was 0.98, with post-hoc analysis also significant at ($p<.05$).

Another two elements with significant effect sizes were ‘tutor questions’ F(2,61)=13.60, significant at ($p<.001$), and ‘tutee connects/categorises’, F(2,61)=4.03, significant at ($p<.05$). For both elements school C did better than school B: First, for ‘tutor questions’ school C mean stood at 1.7 (Std 0.85), and school B mean 0.71 (Std 0.77), with an effect size of 1.34; post-hoc analysis revealed a significance level at ($p=.001$). Second, for ‘tutee connects/categorises’ school C mean stood at 1.00 (Std 0), school B mean 0.53 (Std .51), effect size 1.38; post-hoc analysis confirmed the significance level at ($p<.05$).

Observation element ‘tutee self-corrects’ was omitted from the analysis since in all six classes from three different schools, there was only one instance where a tutee self-corrected. In short, similar to the lesson materials implementation fidelity, school observation show that school A outperformed schools B and C, and school C outperformed school B.

Table 30. School pair observations

Variables	*Schools*	*Mean*	*Std. Dev*	*F-value*	*p.*	*Effect Size*	
Goal Interdependence	A	1.46	.76	26.76	.000	A-B	2.47
	B*	0	0			A-C	1.51
	C	.48	.81			B-C	1.33
Tutor B-Language T-Voice---	A	2.04	1.15	.959	.389	A-B	0.49
	B	2.59	1.06			A-C	0.27
	C	2.43	1.75			B-C	0.11
Tutor Praises Correctly	A	1.31	1.05	8.395	.001	A-B	0.92
	B	.47	.62			A-C	0.98
	C	.48	.51			B-C	0.02
Tutor Meta-Cognitive Questions	A	1.58	.64	2.745	.072	A-B	0.24
	B	1.41	.80			A-C	0.73
	C	1.09	.70			B-C	0.43
Tutee Connects/ Categorises	A	.96	.77	4.032	.023	A-B	0.63
	B	.53	.51			A-C	0.07
	C*	1	0			B-C	1.38
Tutor Questions	A*	1	0	13.60	.000	A-B	0.60
	B	.71	.77			A-C	1.46
	C	1.7	.85			B-C	1.34
Tutee Answers	A	1.31	.68	.862	.427	A-B	0.03
	B	1.29	.85			A-C	0.36
	C	1.58	.81			B-C	0.35
Tutee Questions	A	1	.75	.000	1.00	A-B	0.00
	B*	1	0			A-C	0.00
	C*	1	0			B-C	0.00
Tutor Explains	A	1.46	.76	1.564	.218	A-B	0.43

	B	1.71	.77			A-C	0.52
	C	1.86	.79			B-C	0.19
Task Engagement (Alone)	A	3.11	1.03	10.26	.000	A-B	0.45
	B	3.58	1.06			A-C	1.42
	C	4.33	.58			B-C	0.90
Non-Audible	A	.46	.58	8.009	.001	A-B	1.39
	B	1.35	.93			A-C	1.1
	C	1.14	.85			B-C	0.24

*After adjusting for the outliers the schools have a median of either ‘0’ or ‘1’, hence the standard deviation, skewness, kurtosis and normality are all ‘0’ . School A n=26, School B n=17, School C n=21.

7.3.3 Experimental findings

This first section reports the following: violated statistics assumptions, effect sizes, and differences and significance levels:

***Statistical assumptions*.** This section reports the statistical assumptions which were required to be achieved in order to conduct the analysis on attitude and mathematics performance variables. Different statistical methods have different statistical assumptions, which are a pre-condition in order for the statistical methods to provide un-biased results. Statistical methods are usually built on assumptions of a particular distribution of errors within groups, the relationship of the score distribution between groups, the absence of outliers, etc. For more information on all the assumptions for ANCOVAs, ANOVAs and t-tests please see Howell (2010), who provides an elaborated discussion on issues such homogeneity, normality, homoscedasticity, the relationship of the pre scores to the post scores, and outliers. In order to deal with outliers the extreme values were winsorised to their next closest score. All the remaining statistical assumptions were tested; the process is illustrated in the appendix for each analysis. It must be emphasised that ANCOVAS are usually considered as robust statistical methods when the sample sizes are around 30 and over (Howell, 2010). Also it was necessary to double check and conduct further analysis such as t-test residual gain analysis, which is less effected by violated assumptions (Delaney & Manheimer, 1985), in order to confirm the findings.

Mathematics performance. There were a total of four statistical assumptions which were not met when conducting ANCOVAS for the mathematics performance investigation. The violated assumptions were found in schools A and B. Firstly, two violated assumptions were present in school A when comparing the mathematics performance of the Year 6 peer tutoring group to the Year 6 control group; they were normality assumption for the peer tutoring group on the pre-test, and the homogeneity of variance between the groups. The remaining of the assumption violations were found in school B. There were violated assumptions when comparing the Year 9 peer tutoring to the control group, namely: the assumptions of linearity of regression and homogeneity of regression slopes were not met.

Attitude variables. Table 31 and 32 on the next page summarise all the violated statistics assumptions for each school on the attitude factors. As illustrated, most of the assumptions have been met.

Table 31. *ANCOVA assumptions by school and attitude variable*

Schools	*Year*	*Variables*	*Assumptions*									
			Outlier after Adjusti-ng	*No group has n double the other*	*Norm-ality of pre-test**	*Norm-ality of post-test**	*Normality residual control*	*Normality residual peer tutoring*	*Homos-Cedasti-city*	*Homoge-neity of varianc-e*	Homogenei-ty of regression slopes	*Positive liner pre-post score relation**
		Mathematics self-concept	√	√	√	√	√	√	√	√	√	√
		Mathematics enjoyment	√	√	√	x	√	√	√	x	√	√
	6 Tutee	Relating to people in Mathematics classes	√	√	√	x	√	√	√		√	√
		Social self-concept	√	√	x	x	√	√	√	√	√	√
		Choice in mathematics	√	√	√	√	√	√	√	√	√	√
		Mathematics self-concept	√	√	√	√	√	√	√	x	√	√
A		Mathematics enjoyment	√	√	√	√	√	x	√		√	√
		Relating to people in Mathematics classes	√	√	x	√	√	√	x	√	√	x
	8	Social self-concept	√	√	√	√	x	√	√	√	√	√

	Tutor	Choice in Mathematics	√	√	x	√	√	√	x	√	√	√
		Mathematics self-concept	√	√	√	√	√	√	√	√	√	√
		Mathematics enjoyment	√	√	√	√	√	√	√	√	√	√
		Relating to people in Mathematics classes	√	√	√	√	x	√	x	√	x	x
	7	Social self-concept	√	√	√	√	√	√	x	√	√	x
	Tutee	Choice in mathematics	√	√	√	√	√	√		√	√	x
B		Mathematics self-concept	√	√	√	√	x	√	x	√	√	√
		Mathematics enjoyment	√	√	√	√	x	√	x	√	√	√
		Relating to people in Mathematics classes	√	√	√	√	x	√	√	√	√	x
	9	Social self-concept	√	√	√	√	x	√	x	√	x	x
	Tutor	Choice in mathematics	√	√	√	√	√	√	x	√	x	x

*The assumptions were analysed for both control and peer tutoring group separately. For most of the variables the violations occurred for only one of the two groups.

Table 32. *T-test assumptions for school C variables*

Schools	*year*	*Variables*	*Outlier after Adjusti-ng*	*Norm-ality of pre-test*	*Norm-ality of post-test*	*Homos cedasti city*	*Homoge neity of variance*	*Positive linear pre-post score relation*
C	***8 Tutee***	Mathematics self-concept	√	√	√	√	√	√
		Mathematics enjoyment	√	√	√	√	√	√
		Relating to people in Mathematics classes	√	√	√	√	√	√
		Social self-concept	√	x	√	√	√	√
		Choice in mathematics	√	√	√	√	√	√
	10 Tutor	Mathematics self-concept	√	√	√	x	√	√
		Mathematics enjoyment	√	x	√	x	√	√
		Relating to people in Mathematics classes	√	√	√	x	√	√
		Social self-concept	√	√	√	√	√	√
		Choice in mathematics	√	√	√	√	√	√

Effect sizes and significance levels. Tables 33 and 34, provide a summary of the effect sizes and their significance levels, for both tutees and tutors, for performance and attitude variables for the three schools. For table 33 the ANCOVA and the independent t-test of regression residuals approaches were used as a way of analysing the data, and for table 34 a dependent t-test was applied.

In terms of *performance*, the highest effect size was that of Year 6 tutees within school A as compared to the control group within that school, with an effect size of 0.92, significant at ($p<.001$). This was followed by the Year 9 tutors within school B as compared to the control group within this school, with an effect size standing at 0.85. The third highest effect size was that of Year 8 tutees within school C, which compared the peer tutoring pre-tests scores to the post-tests, effect size standing at 0.79, significant at ($p<.001$). Finally, the lowest effect size was that of Year 7 tutees within school B as compared to the control group within the school, effect size 0.22. T-test effect sizes of the regression residuals were also very similar to those of the ANCOVA scores.

Investigating the impact of ICAT on broader process of learning factor related to social as well as academic attitudes was another aim of this research, providing a more detailed view of ICAT's potentials.

Overall the tutors gained slightly more than the tutees in *attitude variables* in nearly all schools. The highest effect size was that of Year 8 tutors in 'mathematic self-concept', with an effect size of 0.69, significant at ($p=.001$), followed by Year 9 tutors in 'social self-concept', effect size 0.53, significant at ($p=.048$), then Year 6 tutees' 'choice', with an effect size of 0.52 ($p<.05$), and finally Year 8 tutors' 'social self-concept', effect size 0.48 ($p<.05$).

The smallest effect sizes for attitude were found in school B Year 9 tutors 'relating to people in mathematics classes', effect size -0.33, followed by school C, Year 8 tutees 'social self-concept', with an effect size of -0.24.

Table 33. *School A and B effect sizes*

Student Performance	*Peer Tutoring Est. Mean*	*Std. Error*	*Control Est. Mean*	*Std. Error*	*MSE*	*ANCOVA Effect Sizes*	*t-test Effect Sizes*
Year 6 (Tutee) School A	25.90	.485	22.95	.470	10.32	0.92**	0.81**
Year 7 (Tutee) School B	15.37	.702	14.46	.817	16.93	0.22	0.16
Year 9 (Tutor) School B	18.26	1.069	13.08	1.301	27.35	0.85*	0.80*
Student Attitude							
School A Year 6 (Tutees)							
Mathematics self-concept	19.36	.495	19.70	.49	11.05	-0.10	-0.11
Mathematics enjoyment	13.30	.402	12.3	.39	7.32	0.35	0.33
Relating to people in Mathematics classes	19.77	.578	19.45	.57	15.30	0.08	0.09
Social self-concept	14.20	.42	14.36	.41	8.16	-0.06	-0.00
Choice	12.10	.53	10.07	.52	13.01	0.56*	0.56*
School B Year 7 (Tutees)							
Mathematics self-concept	17.69	.70	16.49	.77	16.51	0.29	0.32
Mathematics enjoyment	11.51	.81	10.46	.89	22.32	0.22	0.09
Relating to people in Mathematics classes	18.94	.85	18.65	.94	24.30	0.06	0.09
Social self-concept	14.22	.53	13.98	.58	9.47	0.08	0.02
Choice	12.05	.70	12.69	.78	16.75	-0.16	-0.02
School A Year 8 (Tutors)							
Mathematics self-concept	19.75	.49	17.47	.44	10.97	0.69*	0.69*
Mathematics enjoyment	13.49	.50	11.69	.45	11.86	0.52*	0.49*
Relating to people in Mathematics classes	18.85	.59	18.14	.53	16.58	0.17	0.20
Social self-concept	14.733	.48	13.20	.43	10.78	0.47*	0.48*
Choice	11.130	.50	10.95	.44	10.99	0.05	0.09

School B Year 9 (Tutors)							
Mathematics self-concept	16.48	.42	16.54	.44	4.62	-0.03	0.00
Mathematics enjoyment	10.89	.68	10.08	.72	12.61	0.23	0.15
Relating to people in Mathematics classes	17.15	.75	18.39	.80	14.40	-0.33	-0.22
Social self-concept	14.21	.49	12.84	.51	6.61	0.53	0.54
Choice	12.93	.69	12.68	.73	12.66	0.07	-0.04

** p<.001, *p<.05. (One-tailed for the t-test)

Table 34. *School C effect sizes*

Performance	Pre-test Mean	Std.	n	Post-test Mean	Std.	n	Effect size
Year 8 Performance	8.50	2.81	44	11.45	4.46	44	0.79**
Attitude							
Year 8 Attitude (tutees)							
Mathematics self-concept	16.31	5.54	42	16.29	5.70	42	-0.00
Enjoyment	10.98	5.29	42	10.69	5.00	42	-0.05
Relating to people in Mathematics classes	18.31	5.10	42	17.21	6.25	42	-0.19
Social self-concept	14.98	4.23	42	13.93	4.42	42	-0.24
Choice	11.31	4.89	42	11.38	5.73	42	0.01
Year 10 Attitude (tutors)							
Mathematics self-concept	15.33	4.96	60	16.23	5.10	60	0.18
Enjoyment	9.33	3.95	60	10.35	4.74	60	0.23
Relating to people in Mathematics classes	15.98	4.74	60	16.08	5.94	60	0.02
Social self-concept	12.42	4.28	60	12.67	3.94	60	0.06
Choice	10.85	4.32	60	11.58	4.54	60	0.16

** p<.001 using dependent t-test (one-tailed).

In order to provide a more detailed picture of the results the following are the findings expressed in terms of mean differences and score gains:

Performance: School A. When testing for class effect there was no significant main effect within the peer tutoring or within the control classes for the year 6 students, i.e. the tutees in school A. Meaning that any difference between the peer tutoring and the control group were probably not influenced by the class context.

For school A, for the Year 8 student tutoring a Year 6 student, there was a significant main effect on Year 6 student $F(1, 94) = 18.26$, ($p < .001$). With the treatment group having gained a higher performance score (mean 25.90) than the control group (mean 22.95), Mse=10.32, score mean difference=2.95. One tailed t-test residual gain analysis also showed that for School A, the Year 8 students tutoring year 6 students, there was a significant main effect on Year 6 students, $t(94) = 4.00$, ($p < .001$). With the treatment group having gained a higher performance score (mean 1.36, Std.=2.65) than the control group (mean -1.27, Std.=3.65). Hence, confirming the ANCOVA finding that the year 6 students of the peer tutoring group did gain a significantly higher performance score than the control group, mean difference=2.63, slightly lower than the mean difference shown by the ANCOVA.

School B. When testing for class effect there was no main significant effect on the performance scores within peer tutoring or control classes, for both the Year 7 and the Year 9 students. Also, the t-test analysis revealed that the tutors pre-test data were significantly higher than the tutees', tutors' mean=16.8 (Std 5.09) n=35, tutees' mean=11.52 (Std. 4.99) n=32, t=4.27, significant at ($p < .001$), effect size difference 1.05. Hence teachers' concerns that the tutors and the tutees were of similar ability did not apply to the topics covered.

For school B, for the year 9 student tutoring a Year 7 student, ANCOVA showed no significant main effect for treatment on year 7. Also, t-test residual gain analysis showed that there was no significant main effect on year 7 performance scores.

There was, however with ANCOVA, a significant main effect for treatment on year 9 students, the tutors, $F(1, 41)=8.95$, ($p=.005$), the treatment group gaining higher performance scores (mean 18.26) than the control group (mean 13.08), Mse=27.35. Hence, Year 9 peer tutoring gained a significantly higher performance score than the control group, mean difference=5.18. This was also confirmed by the t-test residual gain analysis which showed that there was a significant main effect for treatment on Year 9, $t(41)=2.59$, ($p=.026$).

School C. For the year 10 student tutoring a year 8 student the dependent-tests showed that there was a significant main effect on performance score improvement for the year 8 student, the tutees, $t(43)=5.88$, ($p<.001$). Students scored higher on the post-tests (Mean 11.45, Std.=4.45) than on the pre-tests (mean 8.5, Std.=2.80), gaining mean difference=2.95.

*Attitudes***.** This section reports the findings on the attitude variables for each of the three schools individually:

School A - The Year 8 student tutoring a Year 6 student. Attitude findings are reported for both tutee and tutor:

Year 6 (tutees). - There was no significant main effect for the Year 6 students on; a) mathematics self-concept, b) mathematics intrinsic motivation, c) relating to people in mathematics classes, and d) or social self-concept. Analogous findings were confirmed by the t-test residual gain analysis.

There was a significant main effect on the level of choice perceived in mathematics classes for the Year 6 students, $F(1, 93)= 7.37$, ($p=.008$). Year 6 peer tutoring students gained significantly more scores in choice of how to do mathematics in class (Mean 12.09) relative to the control group (mean 10.07), Mse=13.01 with mean difference=2.02. An analogous conclusion was also supported by the t-test residual gain analysis, $t(92)=2.71$, $(p=.016)$, with the treatment group having gained a higher score on choice (mean 1.02, Std.=3.95) than the control group (mean -.98, Std.=3.11).

Year 8 (tutors). - There was a significant main effect on mathematics self-concept for the Year 8 students, $F(1, 106)=11.89$, $(p=.001)$. The treatment group gained higher mathematics self-concept scores (mean 19.75) than the control group (mean 17.48), Mse=10.97, mean difference=2.28. Also, t-test residual analysis showed that there was a significant main effect $t(104.04)=3.65$, $(p<.001)$. With the treatment group gaining significantly higher mathematics self-concept scores (mean 1.19, Std.=2.6) than the control group (mean -.97, Std.=3.50), mean difference=2.16.

There was also a significant main effect for treatment on mathematic enjoyment for the Year 8 students, $F(1, 106)=7.12$, $(p=.009)$, the treatment group gaining significantly higher mathematic enjoyment scores (mean 13.49) than the control group (mean 11.69), Mse=11.86 mean difference=1.80. The t-test residual gain analysis showed the same conclusion, $t(105)=2.55$, $(p=.024)$, with the treatment group gaining significantly higher mathematic enjoyment scores (mean .94, Std.=3.09) than the control group (mean -.76, Std.=3.67), mean difference=1.70.

There was no significant main effect for how Year 8 students related to people in mathematics classes, or choice of how mathematics was conducted in class. Similar conclusions were derived from the t-test of residual gain analysis.

There was a significant main effect for the social self-concept of the year 8 students, $F(1, 106)=5.72$, $p=.05$, with the treatment group having gained a significantly higher social self-concept score (mean 14.73) than the control group (mean 13.20), Mse=10.78, mean difference=1.53. Also, the t-test residual gain analysis showed similar results on social self-concept, t(105)=2.49, (p=.028). The treatment group had a higher social self-concept score (Mean .86, Std.=2.88) than the control group (mean -.70, Std.=3.51), mean difference=1.57.

Finally, there was no significant main effect for the Year 8 students on the level of choice perceived in terms of how mathematics was conducted in class. The t-test residual gain analysis showed similar results.

School B: The Year 9 student tutoring a Year 7 student. None of the attitude variables were statistically significant for the tutors or the tutees on school B. This was also confirmed by the t-test residual gain analysis.

School C: The Year 10 student tutoring a Year 8 student**.** Similarly, none of the attitude variables were statistically significant for the tutors or the tutees of school C.

7.3.4 Comparing the lesson materials between different tutee groups

This section provides the findings for the objective: 'comparing the lesson materials of those students who emerged as highest performing tutees with those of the lower performing tutees in order to explore any differences in how different students learned during ICAT'. Please refer to chapter 4, section 4.4, for the research questions.

As mentioned, most of the questions form five categories/conceptualisations when comparing the lesson materials of the higher performing tutees to those of the lower performing tutees:

1. **Total Peer Tutoring Lessons** (question 1)
2. **Goal Interdependence** (question 2)
3. **Quantity** of attempted questions per each part of the lesson (questions 3, 4, 5)

4. **Quality** of answers for each part of the lesson (questions 6-7-8)

5. **Feedback** (questions 9, 10, 11, 12)

Effect size and significance levels. Table 35 on the next page provides the findings of the lesson material comparisons between the two groups, the highest performing tutees to the lowest performing tutees.

The highest effect size when comparing the lower performing group to the higher performing group on the social interdependent indicators was 'total peer tutoring lessons', t(100)=3,74, significant at ($p<.001$), for the higher performing group mean 5.40 (Std. 1.27), and for the lower performing group mean 4.50, (Std. 1.09), with an effect size of 0.76.

The second highest effect sizes to follow were for indicators 'lessons with student set goals', t(100)=3.47, significant at ($p<.05$), for the higher group mean 3.33 (Std 2.55) and for the lower group mean 1.76 (Std 2.03), with an effect size of 0.67; as well as 'attempted exercises in the turn-taking section', t(100)=3.35, significant ($p<.05$), for the higher performing group mean 32.5 (Std 9.36) , and lower performing group mean 26.4 (Std 8.76), also with an effect size of 0.67.

Also, high effect sizes were found on the following areas: 'quality of answers in the practice-test', t(100)=3.22, significant at ($p<.05$), with higher performing group mean 29.8 (Std 9.12) and the lower performing group mean 23.7 (Std 10.00), with an effect size of 0.64; as well as the 'quality of comments given in the connect section' when compared between the higher performing group and the lower performing group on 'specific' comments t(100)=3.17, significant at ($p<.05$), with the higher performing group mean 5.46 (Std 4.36), and lower performing group mean 3.2, (Std 2.82), with an effect size of 0.59.

Table 35. *Group comparison of lesson materials: Effect sizes*

Variables	*Group*	*Mean*	*Std. Dev*	*t-test*	*One-tailed p-value*	*Difference in Effect Size Term*
Peer tutoring lessons	Higher	5.40	1.27	3.74	.000	0.76
	Lower	4.50	1.09			
Lessons with student set goals	Higher	3.33	2.55	3.47	.002	0.67
	Lower	1.76	2.03			
Attempted exercises practice test section	Higher	38.1	10.85	2.46	.032	0.50
	Lower	32.5	11.93			
Attempted exercises connect section	Higher	7.22	4.25	1.29	.400	0.26
	Lower	6.07	4.62			
Attempted exercises turn-taking section	Higher	32.5	9.36	3.35	.002	0.67
	Lower	26.4	8.76			
Quality practice test	Higher	29.8	9.12	3.22	.004	0.64
	Lower	23.7	10.00			
Quality connect section	Higher negative	.57	.81	-.894	.746	-0.17
	Lower negative	.71	.83			
	Higher broad	.62	.80	-2.18	.062	-0.44
	Lower broad	1.00	.96			

	Higher specific	5.46	4.36	3.17	.004	0.59
	Lower specific	3.2	2.82			
Quality turn-taking	Higher	22.4	10.55	1.97	.104	0.37
	Lower	18.8	8.20			
Total types of feedback	Higher	17.6	18.09	1.71	.182	0.34
	Lower	11.8	14.82			
Feedback tics & crosses practice section	Higher	3.93	6.25	-.17	1.73	-0.03
	Lower	4.14	5.99			
Feedback tics & crosses turn-taking section	Higher	10.2	11.38	2.92	.08	0.55
	Lower	4.81	7.05			
Students checking back if goal was achieved	Higher	1.48	1.95	3.00	.08	0.53
	Lower	.64	.79			

Higher group n=60, Lower group n=42

Attempted exercises in the practice-test section showed a medium and significant effect size of 0.50, ($p<.05$).

Other areas which showed a medium effect size, however a non-significant p value, were ‘feedback by ticks/crosses in the turn-taking section’, with an effect size of 0.55 and ‘students checking back if the goal was achieved’, effect size 0.53.

7.4 Chapter conclusion

This chapter has reported several findings. It has covered: attrition rates, implementation findings, violated assumptions, experimental findings, and findings from comparing the high performing tutees' lesson materials to those of the low performing tutees.

Attrition rate: Attrition rates were the highest for the control groups in terms of both performance tests and attitude questionnaires. The highest attrition rates for the lesson materials were for school A.

Implementation: Regarding ICAT implementation according to programme specifications, different data collection techniques paint different pictures. While the observation of the pair interactions show that the students scored very low on various interactive elements, general classroom observations and the findings from analysing the lesson materials provided a more positive picture. It could be argued that one reason why the pair observations showed low sores on interactive variables was due to the short time spent for each of the five windows the pairs were observed. In other words, had each window lasted longer than 30-40 seconds more scores could have been recorded.

In terms of school comparison implementation findings, it is necessary to first comment on why the variables were significant and why so many of the effect sizes were very high.

There were at least three reasons for this phenomenon: Firstly, for the lesson materials the number of participants for each school was not equal. Especially school C contained nearly double the number of participants to that of school A, which as a result increases the F value and consequently the p-value. Therefore it was of little help to investigate the post-hoc coefficient for significant p-values[14]. Secondly, the student ages in each school were different, hence having implications for the lesson materials and observation implementation

[14] The unequal group phenomenon should not, however, have influenced the magnitude of the effect size.

analysis. And finally the topics covered during the lessons differed. Therefore, although the chapter reported the effect sizes, the reason why this was done was to explore and compare the extent of implementation across the schools on the interest variables, rather than to make any generalizable statistical claim regarding the variables themselves.

In terms of comparing schools in terms of implementation, the highest effect sizes were found when comparing school A to school B or School A to school C, with school A having a higher mean. This illustrated that implementation of peer tutoring had been more successful in school A as compared to the remaining two schools.

Statistical violations. There were ANCOVA statistical violations for most of the attitude variables, few of them were also found in mathematics performance tests, however not as many as in the attitude tests. ANOVA and ANCOVA statistical violations do not have an enormous impact on the final results if the group sizes are around 30 and over (Howell, 2010). Although the group sizes for school B were slightly smaller, the use of t-test residual analysis, a method robust against statistical violations (Delaney & Manheimer, 1985), further safeguards against the concern.

Performance. Overall in terms of performance, both forms of statistical analysis applied here revealed that tutees in school A had the highest effect size, as well as significant p values. Also, school B showed the second highest effect size for the tutors. However, it has to be kept in mind that the control group for the older age in this specific school was not concentrating on similar mathematics topics. The same can be said for school C which although it showed a high and significant effect size, was a single group design which deviated from the original design. Without a control group it is challenging to account for the result.

Attitude variables. There was also strong statistical triangulation for most attitude variables, both in terms of effect size and significance levels between ANCOVA and t-test of residuals analysis. Overall, the tutors have done better in terms of attitude variables than the tutees, especially for school A.

Regarding the tutees in school A, they gained higher effect sizes than tutees from other schools, especially for ‘choice in how mathematics was conducted in class’ ($p<.05$), this supports previous findings (Winter, 1996). Mathematics enjoyment, was another attitude variable in which the tutees from school A differed from the tutees in schools B and C in comparison to their control groups. However, the effect size for this variable was insignificant.

Tutors’ attitude score gains in school A on the other hand were positive with significant p values ($p<.05$). Only two attitude variables, ‘relating to people in mathematics classes’ and ‘choice about how mathematics was done’ showed an insignificant p value. Therefore, in terms of attitude gains, this year group gained more than any other group in this study.

Comparing higher performing tutees to the lower performing tutees on lesson materials. In terms of comparing the lesson materials between the highest and the lowest group of tutees, the highest effect sizes were found for indicators: ‘Total lessons’, ‘goal interdependence’, and ‘quality of answers in the turn taking test’. With the effect sizes favouring the higher performing group of tutees.

The next chapter provides an elaborative discussion on the findings.

PART III

8 Discussion

8.1 Introduction

This chapter discusses the findings derived from three interconnected analyses: a) implementation analysis, or school level analysis by investigating both, lesson materials and observations, b) experimental findings, i.e. the impact of ICAT on students' mathematics performance and attitudes, c) group level comparison of lesson materials between the students who gained the most and the least in terms of 12 social interdependent indicators.

More specifically the following are to be discussed:

1) How can the implementation findings be understood?
2) What explains the findings on performance? How do the findings on performance relate to the literature?
3) What can be learned about the process of learning variables? How do the attitude variables relate to the literature? Why were the results mixed for each school? Specifically, why was it that the highest achievers in terms of attitude variables were the tutors from school A, with three out of 5 attitude variables having significant effect sizes: mathematics self-concept, intrinsic motivation and social self-concept? Also, why did most of the attitude variables have a low effect size and were non-significant? How do the group level analysis findings relate to the literature? The group level analysis of tutees' lesson materials explored that some

of the variables with medium to high effect sizes were: ‘peer tutoring lesson materials’, ‘setting goals’, ‘giving correct answers in practice section’, ‘giving specific answers in the connect section’, ‘attempting a high number of exercises in the practice-section’ and ‘attempting a high number of exercises in the turn-taking section’, ‘feedback by ticks and crosses in the turn-taking section’ and ‘checking if goal was achieved’. What explains these differences?

4) How do the findings relate to one other?
5) What are ICAT’s strengths?

It is safe to suggest that the findings of this project alone should not be generalised to any population, since both internal and external validity were not strong and the sample was only a convenience sample rather than a true random selection of the participants, something which is extremely hard to generate in Social Sciences (Gorard, 2014). However, previous meta-analyses and studies similar to this have already established the importance of cross-age tutoring, and peer tutoring incorporating the elements which were adopted in this study.

Also, the language used throughout this research has been that the ‘paper tests the impact of ICAT on mathematic performance’. It must be pointed out that the findings here cannot support the argument that ICAT has enhanced ‘all’ mathematics performance. However it can, to some extent, support the argument that ICAT has enhanced the mathematics areas covered by ICAT in the intervention. Simply because ICAT worked on the areas of mathematics covered here, does not prove that it has done so for mathematics in general. More research needs to be done for the impact of ICAT on other areas of mathematics.

Prior to discussing the findings of each data collection method in more detail it is necessary to shortly discuss ‘regression to the mean’:

The phenomenon of regression to the mean could be a problem when conducting quasi-experimental designs. It emerges when the projects’ groups, intervention or control, are not equal on pre-test scores, a phenomenon usually eliminated by strict randomisation or performance matching. Hence, it is usually the case that the control or the intervention group result in bias due to chance, specifically a particular group could fall at the extreme end of the entire samples' mean during the pre-test, and then during the post-test the group adjusts itself (Trochim, 2012).

Regression to the mean is very hard to detect, with many experienced researchers failing to do so (Trochim, 2012). Trochim (2012) suggests that one way to establish whether there was a regression to the mean is to look at the relationship of the pre-test and the post-test scores for each group, intervention and control, and see if the students who scored high in the pre-test also scored high in the post-test, hence establishing that the results were not by chance.

The formula for measuring the amount of regression to the mean is 100(1-r), where r is the correlation. The higher the correlation, the lower the percentage of the regression to the mean (Trochim, 2012). Regression to the mean can *overstate* or *understate* the size of the effect depending on which group has significantly lower scores on the pre-test. When the experimental group has a significantly lower mean score than the control group there is a risk of overstating the effect size. When the reverse is the case one can easily understate it (Trochim, 2012).

However, regression to the mean analyses were undertaken for all variables and both schools A and B. All variables showed a 20-30% regression to the mean. However, since this phenomenon appeared for both groups in all variables, control and peer tutoring, the threat to inflating the effect size is extremely small, as regression to the mean for any variable has cancelled itself out, and therefore poses no danger to the results.

8.2 Discussing implementation fidelity

How can implementation be understood? This question can be answered in three ways: a) by looking at how the schools differed in terms of implementing the design agreed on prior to the intervention, b) by comparing schools to each other in terms of which school had the best ICAT implementation overall and c) by investigating the areas with least implementation of the ICAT components according to the programme specifications.

Firstly, we can explore the extent to which the schools stayed true to the initial experimental design arrangement. This is straight forward, out of the three schools one of them dropped the control group, school C, and another incorporated irregular teaching methods, school B. Therefore, it is clear that overall the schools did not adhere to the initial agreements. Consequently there was a difference between the schools in terms of implementing one unified experimental design.

Secondly, in terms of implementation of the ICAT elements, not all of them were implemented equally. Within each school, as the lesson materials illustrated, the mean rate of ‘set goals’, ‘attempted exercises in the connect section’, and ‘checking back if the goal was achieved’, were all very low compared to other sections and indicators/elements investigated here. Could the students have simply engaged more verbally? The answer is ‘no’ for at least two of these elements; since the observations also showed that ‘goal interdependence’ was very low, as was ‘connecting ideas to previous ideas or real life experiences’. Other characteristics which were not implemented to a great extent were ‘praising correctly’, and ‘tutee asking questions’. There was also a difference across the schools in terms of properly implementing peer tutoring. The school level implementation analyses, comparing students’ work on the lesson materials and the pair observations, showed that peer tutoring was not implemented in the same manner across the schools.

In terms of the lesson materials implementation fidelity school A showed the highest effect sizes when compared to school B and C in all areas, with the exception of 'feedback ticks/crosses in the practice section'. School A's implementation of ICAT was also closer to the programme specifications. School C on the other hand showed a better implementation of peer tutoring lesson materials than school B, with school B implementing peer tutoring better *only* in three indicators: 'attempted exercises turn-taking section', 'quality of answers turn-taking section' and 'student set goals'.

Moreover, in terms of the lesson materials, only in one out of twelve areas did school A not implement peer tutoring better than schools B and C, 'feedback by ticks/crosses in the practice section'. One explanation for this could be that the tutors in school A did not see it necessary to give feedback by ticks/crosses at this stage, as it was only a practice section and did not include the final exercises with which the students could asses themselves against their set goal.

During the observations the differences were not as high as when comparing the lesson materials. However, school A also led in terms of the number of indicators with high and statistically significant effect sizes. In four out of 11 indicators, school A showed a mainly high and significant effect size compared to schools B and C, those were: 'goal interdependence', 'tutor praises correctly', 'task engagement alone' and 'inaudible'. School C on the other hand did better than school B in implementing peer tutoring in two areas; 'tutee questions', 'tutee connects/categorises'.

The most unexpected finding of the observations was perhaps the lack of school difference for the second indicator 'synchronising body language and tone of voice', i.e. non-verbal communication, which is thought to be an important interpersonal interdependent element in the peer tutoring context. The expectation was that because school A gained more in terms of performance achievements, they would also show better interpersonal communication skills. On the other hand, most research on non-verbal cues in peer tutoring is conducted via lengthy video recordings, rather than short window snapshots, therefore the unexpected finding could be partly put down to the weakness of the observation method.

8.3 Discussing performance findings: Relating the findings to previous research

The highest effect size on performance was that of tutees at school A, effect size 0.92, followed by school B tutors' performance with an effect size 0.85, then tutees from school C, with an effect size 0.79, ending with tutees in school B, effect size 0.22.

What explains the high effect sizes on performance for schools A? Apart from the fact that these schools showed the highest implementation rate, there are at least three additional explanations:

Firstly, ICAT combines within it elements of the most effective peer tutoring interventions. Hence the first explanation would be the peer tutoring framework adopted here, which was informed by the 'what works' literature.

Secondly, research design issues could have also biased the findings. Specifically, it is very likely that the effect size could have been influenced by the Hawthorne effect, which leads students in the intervention group to make more of an effort during learning as well as in the tests. Also, since the intervention was conducted at the individual level within a school, it is very likely that the control group participants, students and teachers, could have been demoralised, and therefore not giving their full attention during the post-tests. Moreover, another explanation for the high effect size in school A could be that the teachers in the intervention groups could have 'taught to the test' during the six weeks; in other words the teachers could have indirectly provided the students with the answers to the post-tests considering that all the school were keen to see peer tutoring work.

The final reason why the effect size was so high in school A is that the instrument was developed to evaluate the topics covered by the ICAT intervention. If a national test was used, this would not have captured the true impact of the intervention, since such tests cover more than 6 areas of mathematics. This conclusion is consistent with the tables in the systematic review shown under chapter 3, in which most researcher-made instruments provide very high effect sizes as compared to those instruments which do not measure the impact of the intervention on the topics covered during peer tutoring. A national test or an instrument which covers many areas of mathematics are broader in nature and are designed to capture interventions which are intensive and broad in scope. When comparing the size of the effect gained in this study for school A to other researcher-made instruments which test the impact of an intervention by looking at the precise area of where the intervention was applied, the findings are very much in line with previous findings in peer tutoring that incorporate cross-age or interdependent elements.

How else are the findings related to past research? The positive findings on performance for school A are in line with the literature review. Specifically, the findings support one of the earliest claims regarding the effectiveness of peer tutoring in enhancing mathematics performance, the meta-analysis by Cohen, (1982); as well as one of the latest studies with a strong experimental design, which is that of Tymms et al, (2011), on elementary school students within an entire region in Scotland. They have shown that cross-age peer tutoring provides higher effect sizes relative to same-age peer tutoring. Higher effect sizes were also found for mathematics as compared to other subjects.

The findings from this study are also in line with the cross-age study conducted in mathematics by Bar-El (1982), a four-month study measuring students’ performance.

Finally, the findings concur with the notion that interdependent peer tutoring in mathematics provides a high effect size[15],as illustrated by the five-month study by John, Fantuzo, King and Heller (1992), the 18-week study by Fuchs, Fuchs, Hamlet, Philips, Karns and Dudka (1997), the ten-week study by Ginsburg-Block and Fantuzzo (1997), the seven-week study by Ginsburg-Block and Fantuzzo (1998), the three-week study by Fantuzzo Polite and Grayson, and the five-week study by Menses and Frank (2009).[16]

How can the findings be interpreted for school B and C? The only non-significant effect size on performance was that of tutees at school B, which is very likely to have been due to the limited implementation of the ICAT and the identified contextual factors such as making use of other interventions, specifically ICT techniques and frequent formative tests used in most mathematics classes including the control groups. Although school B tutors showed a high effect size, this is very likely to have been due to the fact that, unlike with the tutees, the control group for tutors did not cover those particular exercises covered by the tutors during the peer tutoring intervention.

Design issues could also be said to have influenced the high effect size in school C. Specifically, the magnitude of the effect size could simply be due to the maturation element since the control group was lacking for this school.

[15] Please see the findings of the systematic review on same-age reciprocal peer tutoring interventions, table 6. Observations of tables 6, 7 and 8 also reveal that researcher made instruments often have higher effect sizes.

[16] For more information on the findings of previous mathematics peer tutoring studies please see tables 6, 7 & 8, section 3.6.3.

8.4 Discussing the peer tutoring learning process

8.4.1 Attitude variables

Why were the effect sizes low for attitude variables compared to performance? The measurement debate can also be expanded to explain why the effect sizes for the attitude variables were lower. Specifically, academic attitude questions did not ask what the students thought of 'equations', or of 'functions' and 'data interpretation', rather the questions were related to mathematics in general. Similarly, social self-concept questions did not necessarily measure whether their one-to-one interaction skills changed; rather the questions were broader in nature. Had the attitude questions been designed for the specific areas of mathematics covered here, or were they to be more specific in terms of social interactions then the effect sizes could have been higher. Again, another way to solve the academic attitude instrument issue would have been to cover more topics in mathematics and expand the length of the intervention. In other words, whereas for performance the instrument measured precisely the mathematics areas covered during the ICAT, attitude variables by contrast were all too broad.

Why were most of the attitude variables with the highest effect sizes in school A? Again the explanation for this can be put down to the implementation success of ICAT, which was supported by the analysis of the lesson materials as well as the observations.

How did the findings differ in terms of tutor and tutee? Overall, most of the effect sizes were small and non-significant. The tutors gained more than the tutees on affective/attitude factors when compared to the control group, both in terms of effect size and significance levels, especially for school A.

Moderate effect size for the tutees, which were, however, non-significant, were mathematics intrinsic motivation for the Year 6 tutees from school A, with an effect size of 0.35 and mathematics self-concept for the Year 7 school B, effect size 0.29.

How do the findings on attitudes relate to the previous literature? For school A, the Year 8 students tutoring Year 6 students, tutors' social self-concept was medium-high effect size 0.47, significant at ($p<.05$). This is in line with previous meta-analyses that demonstrate that peer tutoring enhances social self-concept (Ginsburg-Block, et al., 2006). Other studies which have shown this to be the case for mathematics specifically are those by Topping, Campbell, Walter, and Smith (2003), a cross-age peer tutoring in mathematics study, Fantuzzo, et al., (1992), and Ginsburg-Block and Fantuzzo (1997, 1998) same-age reciprocal in mathematics studies with positive interdependence.

The attitude findings, specifically those on mathematic intrinsic motivation with an effect size 0.52 and mathematics self-concept with an effect size 0.69 both ($p<.05$), on the Year 8 tutors, are also in line with previous findings:

In terms of the findings on mathematics intrinsic motivation this corroborated the finding from Ginsburg-Block and Fantuzzo (1998), which established that peer tutoring students gained more than the control group in mathematics in terms of mathematics intrinsic motivation. This is also supported by the idea that informational structures, which help students with the sequence and directions of their interaction, rather than controlling structures directed by teachers with detailed scripts, help the improvement of intrinsic motivation (Deci & Ryan, 1985; Fantuzzo, et al., 1992).

The mathematics self-concept findings are also in line with the findings by Tymms, et al, 2011, which showed that the tutors gained more than the tutees in a cross-age tutoring settings in mathematics. Or the findings by Topping, (2003) on cross-age mathematics peer tutoring, or the meta-analysis by Cohen, et al., (1982). The mathematics attitude findings found here also concur with mathematics peer tutoring interventions which have employed same age, positively interdependent structures, such as those by John, et al., (1992), Ginsburg-Block and Fantuzzo (1998).

In terms of affective factors for the tutees, only ‘choice in how mathematics was done in class’ for school A showed a high and significant effect size 0.56, ($p<.05$) for the tutees. These findings are in line with previous findings on peer tutoring looking at the issue of choice (Winter, 1996).

8.4.2 Lesson materials group comparison discussion

The lesson material comparisons were conducted in order to explore how students of different abilities learned during ICAT.

Out of twelve indicators used to compare the lesson materials between the higher performing group of students and the lower group, four indicators showed high and statistical significant effect sizes: ‘peer tutoring lessons’, ‘lessons with student set goals’, ‘attempted exercises in the turn-taking section’ and ‘quality in the practice test’. For the indicator ‘quality in the connect section’ two of its indicators did not have a significant effect size, ‘negative’ and ‘broad’ comments, while ‘specific’ comments did have a high and significant effect size.

Indicators with medium effect sizes were: ‘attempted exercises in the practice-test’ being significant. Two other indicators showed a moderate however non-significant effect size: ‘feedback by ticks/crosses in the turn-taking section’ and ‘checking back if the aim was achieved’.

The remaining indicators showed a low or negative effect size and they were non-significant, those being: ‘Attempted exercises in the connect section’, ‘total-feedback’, ‘quality turn-taking’, ‘feedback ticks/crosses practice section’.

How do the findings relate to the research literature? In terms of indicator ‘goal interdependence’ as set by the students, the finding supports the general notion by Social Interdependence Theory that groups who set goals, as a result perform better (Rohrbeck, et al., 2003). A causation claim, however, cannot be made for this particular case. The ‘goal interdependence indicator’ measured the extent to which students were engaged via a set goal.

Regarding the finding on the category ‘quality of answers’, especially on indicators ‘quality in the practice test’ and sub indicator ‘specific comments’, these are in line with previous research. Specifically, Fuchs, et al., (1997), and a review by Roscoe & Chi, (2007) have illustrated that students perform better overall when they provide clear, specific and elaborated answers while interacting with one another. While quality of answers in the practice test measured the level of cognitive engagement and cross-ability within a pair, the quality of answer in the connect section measured meta-cognitive ability and the level of cross-ability.

Also, the findings on the indicator ‘turn-taking section’ sit well with both Social Constructivism and Social Interdependence, i.e. indirectly showing that the higher performing group has been more cognitively engaged.

Group comparison findings on ‘feedback ticks/crosses in ‘turn-taking section’, supports the social interdependent and social constructivist notion that feedback is a key element in peer tutoring. As mentioned in the instrument development chapter, feedback can be seen as a way of measuring the level of interpersonal interdependence, cognitive and meta-cognitive engagement and the amount of cross-ability within a pair.

Therefore, exploration analyses of the lesson materials show that the higher gaining tutees were more interdependent by goal, engaged more cognitively and meta-cognitively, and could have had more cross-ability-pairings, which makes giving feedback, and providing better quality of answers possible.

Finally, although these explanations seem to go hand in hand with a social interdependent perspective, it must be pointed out that in terms of significance levels a proportion can be explained due to the unequal number of students between the groups, the higher group 60 students and the lower performing group 42. Consequently, a substantial portion of the between group sum of squares was accounted for by the imbalance rather than the actual difference of characteristics, leading to a higher between group sum of squares, higher F ratio, and then consequently to being significant.

Another issue that could have influenced the findings is that the higher group tutees could have very well been the upper ability pairs, as the higher ability older students were paired with the higher ability younger students, and so on down the line[17]. Since the pairing was conducted by the schools' head-teachers, it was not possible to conduct further investigation if and to what extent that was the case. However even if such clustering was present, the issue is not of great concern since the findings still provides insights into how different groups approached ICAT and how they differed in certain indicators.

8.5 How do the findings aid one another?

In order to gain a better picture of why school A performed better than school B and C one has to look at the interaction of different variables made possible by different data collection and data analysis techniques. The findings aid one another at different levels:

[17] Please see section 6.5.1 for the threshold used to categorise higher group of tutees to lower group of tutees. Not to be confused with higher ability or lower ability tutees concept used for the pairing of the students.

To begin with, the first indicator for the observations, goal interdependence, can also directly be aided by the completed section of 'goal interdependence' at the lesson material level, as this is straightforward. One can clearly see that for goal-interdependence the triangulation is strong. Both, observation analysis and lesson material analysis indicate that 'student goal setting' in school A showed a higher effect size than any other indicator when compared to the remaining two schools. Therefore one can rule out that goal interdependence took place in schools B and C to the extent that it did in school A, since both, the observations and lesson materials supported this argument.

Hence, what is interesting in terms of findings aiding one another, is that indicator 'goal-interdependence', which showed one of the highest effect sizes on the group level comparison (higher group students having a higher mean of set goals), also showed the highest effect size on both implementation analysis types when comparing school A (the higher performing school on peer tutoring) to schools B and C.

Similarly, the indicator 'attempted exercises in the turn-taking section', showed the highest effect size in the tutees' lesson materials group level analysis, with the better performing students having a higher mean of attempted exercises in the turn-taking section. What is interesting is that the school level implementation analysis of lesson materials also showed the better performing school, school A, to have attempted more exercises then the other schools.

Moreover, the findings on goal interdependence are also interesting since 'goal interdependence' findings can also be aided by the past cooperative peer learning research findings. Slavin, Sheard, Hanley, Elliott, and Chambers (2013), through a high scale clustered randomised control trial of cooperative learning, found that the method did not produce the expected results based on USA experience. However, on closer inspection of the actual implementation they revealed that important aspects such as goal interdependence and individual activity were not implemented to the required degree by the cooperative groups' schools. The importance of goal interdependence as set by students is already discussed in the review of meta-analyses section, showing that peer tutoring interventions incorporating goal interdependence, with goals set by students themselves and other autonomous structures enhancing student autonomy, show the highest effect sizes (Rohrbeck, et al., 2003; Ginsburg-Block, et al., 2006).

Also, the school with the highest effect size on mathematics self-concept or mathematics intrinsic motivation was school A, the same school which also showed the highest frequency of set goals. Apart from the social interdependent perspective which would suggest that positive interdependence makes the students engage more with the materials and one another, another explanation would be that ICAT contained two very important elements which were applied to impact academic self-concept, specifically the test-like context of ICAT combined with enabling the students to easily achieve their set goals. Hence, a student who sets an achievable goal in a test-like format every week, a component of ICAT, would be more likely to improve their performance than a student who attends such test-like context but does not set a goal. This specifically applies when we consider the reciprocal effect of subject specific self-concept with performance tests (Marsh & O'Mara, 2008). In other words, if a student observed every week that he/she was good at achieving the set goal in a test, this would improve how they perceived themselves in mathematics. Especially when considering that one of Marsh's items specifically asks students how they see themselves in mathematics tests, an important item measuring subject specific attitude. Again, it would have been harder to explore these kinds of themes, had the implementation investigations of the lesson materials not taken place.

Another way the findings aid one another is when we investigate 'correct praising between the schools'. School A recorded nearly double the frequency compared with schools

B and C, as well as being less in-audible (noisy), compared to the other schools. Praising is an important factor for Social Interdependence Theory, and it is a type of feedback strongly endorsed by both current dominant theories, Social Interdependence and Social Constructivism, and has been shown to provide good results (Chalk & Lewis, 2004; Goyen & McClelland, 1994; Harrison & Cohen, 1971). As the chart in figure 28 in the previous chapter illustrated, compared to other indicators, correct praising was the variable least implemented, with the lowest mean. This is not strange, since even adult teachers find correct praising hard to master. Similar findings on the low frequency of praising were also revealed by Topping, et al., (2011).

Also, the collected performance data of the Year 9 tutors from school B, illustrated the ability difference between the tutors and the tutees, an issue raised by the teachers of school B prior to the intervention. Collecting information on the tutors as well helped to ensure that this variable could not have counted to explain why school B tutees did not perform as well as school A's tutees later on in the analysis. The simple analysis conducted did show that the tutors were of higher ability than the tutees on the topics covered during peer tutoring.

On the issue of comparing tutees' performance of different schools, general observations and school visits provide further aid in understanding why school B and C did not perform as well as school A. For example, with school C one could argue that the school only participated in the project since they were expecting a crucial visit from Ofsted, a knowledge which the researcher did not hold in the beginning, however was made aware of this during the observation and visiting sessions. For school B on the other hand, the school was conducting tests very regularly trying to increase students' performance, as well as one of the peer tutoring teachers using interactive ICT (information communication technologies) in mathematics, both of which could impact the final results. Lemons, et al., (2014) report on the effectiveness of PALS, a same-age reciprocal peer tutoring format, investigated 8 years of

PALS with students of ages 4-5 and revealed that suddenly during the last few years PALS was not effective any longer. On closer inspection however, they revealed that the schools were committed to governmental reforms which were based on the 'what works' literature. In other words the comparison was not being made between PALS and a normal classroom but between PALS+reform versus reform. The same could be said regarding school B’s extensive use of the formative tests, which is an effective learning tool.

Another reason why school B did not have a significant p value for its tutees’ score gains could have been that school B did not have access to the number of participants needed for the effect size to be significant at the .05 level.

8.6 Discussing ICAT Strengths

By simply choosing a cross-age tutoring intervention two important interdependent elements were automatically achieved, role interdependence and cross-ability interdependence in which the young needed the older student. Most of the elements and parts in the ICAT framework have already been used by previous same-age or cross-age methods. However, they have not been applied in the way shown here. The following were some interdependent elements, which set the ICAT framework apart from other cross-age peer tutoring interventions:

Academic goal interdependence. Unlike most *cross-age* peer tutoring interventions, which rely only on the idea of ‘role’ as positive interdependence, the framework here also incorporated academic-goal interdependence.

Social skills and academic training. Most cross-age peer tutoring interventions concentrate on academic training rather than social skills training. The intervention here concentrated on both simultaneously.

Interpersonal interdependence through praising. Apart from being a good feedback mechanism, praising would also contribute to improving pair bonding if provided correctly, i.e. in a kind, synchronised manner and only when the tutee responded correctly (Johnson, 1990). Praising was included in the training package as well as written in the ICAT framework to remind students.

Positive autonomous & informative structural interdependence. This was done by combining scripts, exercises and praising cards all in one document, providing students with more autonomy from the teacher. This is unique to the ICAT framework. The students were given a choice in setting their goals, a choice in the turn-taking test, and in marking the tests. Also, in order to improve structural interdependence a numbers system was used to provide students with information on the steps within each part. An evident problem in peer tutoring is that tutors give the answer to the tutee too early (Harrison & Cohen, 1971). The timing of when the answer is given, an academic skill, is very crucial (Topping, 2001). Similarly the timing of praising, a social as well as academic skill, is also important (Johnson, 1990).

Individual task assignment interdependence. The turn-taking test section was an individual task, in which the combined result would determine the level of success for the pair. Because the final stage also required the tutor's individual input the tutor had to take the learning in the previous sections more seriously, as he/she too was going to be tested.

Task interdependence. By ranking the exercises from very easy to very difficult, the task took the form of a game in which the tutors' role was to push the tutee up the knowledge ladder. Had the "ladder" not been there, the tutor and the tutee would not be able to see their level of progress and hence could have been discouraged to interact.

Advanced cognitive and metacognitive engagement. As already mentioned, ICAT explicitly engaged students in meta-cognitive discussion by providing a section entitled "Connect", in which the students were asked to engage into two different, yet similar, ways of thinking: First to connect what they just covered to previous mathematics classes or subjects in order to provide some kind of *categorisation* (Kramarski & Mevarech 2003; 2004), and second to relate the topic to real life events, as in the case of 'Shared Maths' peer tutoring (Tymms, et al., 2011). Also, the ranking of the exercises from very easy to very difficult was aimed at enabling the students to monitor themselves and each other, as would the final reflection of investigating whether the goal was achieved. Self-monitoring is a crucial meta-cognitive task (Roscoe and Chi, 2007).

ICAT also had the following two additional characteristics which were inserted to the method in order to further improve students' academic self-concept:

Familiarising students with tests by providing a test-like peer tutoring environment. This was achieved by naming two out of four peer tutoring sections as; "Practice-Test" and "Turn-Taking-Test". Again this is similar to Fantuzzo, Polite and Grayson's (1992), Class-wide Peer Tutoring, with three major differences. Firstly, the exercises, for both parts, range hugely in the level of complexity, hence aiding the students to work on their true ZPD. Considering that in peer tutoring the teachers very often do not know what level the tutors or tutees are at, since the classes are mixed between the older and the younger ages, teachers find it hard to identify or provide the correct exercise levels. Secondly, the final test is a Turn-Taking test; hence the students carry on working together, rather than completely alone. They do so, however, in such a format that the tutee and the tutor both see how their partner solves the exercises. Finally, and most importantly, the ICAT does not rely on flash cards for the tutor to give feedback to the tutee, the tutor has to search his/her own thoughts for the answers, formulate it and present it, a slightly harder task than just formulating and presenting the information.

The main reason for titling the second part and the final parts as ‘tests’ was that the most famous reciprocal relationship in education is that between subject-specific academic self-concept and previous performance as measured by tests (Marsh, 2008). However, in order to make sure that such context was made enjoyable and the test experience positively impacted mathematics self-concept one further element needed to be applied, specifically providing more chances for the pairs to reach their goals.

Providing more chances for the pairs to reach their goals. Regardless of the level of challenge students set for themselves, students were provided with more chances to achieve their goal. In the goal setting section the students were asked to set a grade they would like to achieve, ranging from 5-15. However, there were 7 exercises for each turn-taking test with 4 points each, totalling up to 28 points. This further aided students’ improvement of their subject-specific self-concept when they realised that they had bypassed the highest point set. Goal setting by the student is extensively used by Fantuzzo, Polite and Grayson (1990) in the same-age context, however with tangible incentives.

8.7 Chapter conclusion

This chapter has provided a discussion on the findings. A discussion was provided on the level of implementations, performance findings, attitude variable findings, and on the lesson material findings.

In terms of implementation, it was argued that this can be analysed by looking at the extent to which the schools stayed true to the design and the extent to which they applied

ICAT according to the instructions. Regarding the first issue schools A and B were the schools which stayed true to the design. In terms of ICAT application, school A implemented peer tutoring better than both school B and school C as well as stayed closest to ICAT programme specifications. Further, school C implemented peer tutoring better than school B. This was the case when analysing both the lesson materials and the observation data.

In terms of the effect sizes on performance it was argued that the highest impact was for the tutees in school A, which was also significant, with school B tutees having a small effect size and non-significant p value. Although there were moderate to high effect sizes for the tutees in school C and those of the tutors in schools B, it was recommended that due to very problematic research design issues such impacts were harder to justify.

When explaining why school A had the highest effect size, four issues were identified. Firstly, this can be due to the application of the ICAT, an intervention incorporating effective characteristics. Secondly, it can also be due to the fact that school A implemented ICAT better than schools B and C. Thirdly, it can be due to bias resulting from the research design. Finally, it can be due to the instruments used, since the instruments were not broad in scope, rather they concentrated on measuring the impact of the mathematics areas covered during ICAT.

The findings were then linked to past research evidence which has concentrated on the impact of peer tutoring towards performance. Caution needs to be applied with generalising the findings to all mathematics areas, or generalising in general, since the interventions in this research only concentrated on limited topics and the design was not as strong.

A discussion regarding the impact of ICAT on process of learning attitude variables was also provided, together with exploring interdependent variables by comparing lesson materials of different tutee groups.

In terms of the attitude variables two points were also established. Firstly, the highest attitude effect sizes were for school A, which can also be linked to the fact that this school achieved the highest ICAT implementation. Secondly, the fact that the effect sizes for the attitude variables were smaller than the performance effect sizes, can be explained by the sub-tests for measuring each attitude type having been broad in nature.

The findings on attitude were then linked to past peer tutoring literature.

When exploring which social interdependent indicators gained the highest effect sizes, favouring the highest performing group of tutees when investigating lesson materials, four indicators were pointed out: 'peer tutoring lessons', 'lessons with student set goals', 'attempted exercises in the turn-taking section' and 'quality in the practice test'. Two of the sub-indicators for 'quality of answers in the connect section' did not have a significant effect size, 'negative' and 'broad' comments, while 'specific' comments did have a high and significant effect size.

Three areas showed a moderate effect size, favouring the higher performing group of tutees: 'attempted exercises in the practice-test', 'feedback by ticks/crosses in the turn-taking section', and 'checking back if the aim was achieved', a form of self-feedback or self-monitoring mechanism.

The findings on the indicators were then linked to previous research findings.

Finally, a section exploring how the findings aided one another was also provided. This section showed how the analysis of the lesson materials and the observations, while investigating implementation, showed similar conclusions in areas such as goal interdependence and feedback. Also, a theme was discovered between the higher performing tutees (group level findings) and the higher performing schools (implementation findings) in terms of the 'quantity of exercises attempted in the turn-taking section' and 'goal interdependence’, specifically, in both these indicators the higher performing side (group or school) contained favourable effect sizes.

The next chapter extends the discussion chapter by reflecting mainly on the limitations.

9 Discussing the Limitations

9.1 Introduction

This chapter concentrates on three areas: theoretical, methodological and other limitations.

At the theoretical level the paper covers three related issues: The issue of who benefited more, the tutor or the tutee. Secondly, the lack of additional cooperative elements in the ICAT framework. Finally, the lack of additional theoretical elements in the ICAT intervention.

In terms of providing an adequate methodological evaluation it is also necessary to cover the following: a) experimental limitations, b) limitations on group analysis of tutees' lesson materials, c) limitations on implementation analysis and d) other methodical limitations.

In terms of the last area, 'other methodical limitations', this will include issues such as: the lack of qualitative investigations, the absence of additional support for the schools, the absence of a second marker, the lack of comparing ICAT to a normal cross age peer tutoring condition, collecting additional data and the lack of investigating the control group after the end of the project are all discussed.

Some of the limitations in this research are the result of the researcher having been included and later excluded from two initial research projects in the form of clustered (at the school level) randomised controlled trials, one in mathematics and later one in reading. The aim in those projects was similar to the paper here, to understand better how Social Interdependence Theory can be applied to cross-age peer tutoring. The research would have used Path Analysis. The exclusion from the projects, the details of which are not relevant to this thesis, had implications for the choice of research questions and the methodology of the paper as it placed a constraint on the time factor.

9.2 Theoretical Limitations:

9.2.1 Who benefited more from ICAT, tutors or tutees?

To some extent the question of who benefits the most in peer tutoring has been discussed in the literature long ago, and different theories either point to the tutor, tutee or both (Goodland & Hirst, 1989). In peer tutoring there is a belief that different peer tutoring interventions informed by different theories, help different students in different areas (Yarrow & Topping, 2011). One limitation in terms of understanding who benefited more from ICAT is that the research was unable to test tutors' performance[18], since it was thought that the comparison would lack legitimacy, considering that the control group of the older students concentrated on different mathematics topics. Therefore, currently it is not possible to say whether a cross-age peer tutoring intervention informed by social interdependent theory would benefit more the tutee, the tutor or equally both in terms of performance.

Also, one has to consider that the concept of 'academic benefit' is further subdivided into *process of learning* attainments and *performance attainments* as another form of educational benefit, as is the concept of *social benefits* gained from peer tutoring, which can be linguistic, affective or behavioural. Again the findings from this study cannot fully answer these questions considering the methodological limitations.

9.2.2 Lack of additional cooperative elements in the ICAT framework

A limitation faced in this paper was that the ICAT framework did not incorporate equally well all the main theoretical elements of Social Interdependence Theory. Specifically, the method developed here did not accommodate the concept of 'reward interdependence'. Also,

[18] Only for school B were such analysis conducted. This was conducted in order to investigate whether the tutors were of higher ability to the tutees in general as required by ICAT, a concern raised by the school teachers.

the notion of pair or ‘group reflection’ (Jolliffe, 2005) could have been further extended and applied at the end when the students reflected back to their original goal, allowing students to discuss what went well and what needed improvement. Therefore, broadly speaking certain aspects, such as cross-ability, were more prominent than other social interdependent characteristics.

However, even the implementation of ‘cross-ability’ as a characteristic of social interdependence could be questioned if we look at school B; in school B teachers were of the opinion that some Year 9 students were not distant enough in terms of overall mathematics ability to those of Year 7 students. Even though the pre-test analysis of tutors’ and tutees’ mathematics knowledge of the actual peer tutoring topics did show a significant difference, and high effect size, in other mathematics areas the tutors could have had similar ability. This could have impacted their explanation abilities, and hence the final results of tutees ’achievements (Fuchs, et al., 1996).

9.2.3 Lack of incorporating other theories

Another clear theoretical limitation of this research was the inability to incorporate theories of personality and personal beliefs. A study conducted on university students by Beckmann, Wood, Minbashian and Tabernero (2013), illustrated the importance that effort attributions and efficacy plays on peer group dynamics and outcomes. Concluding that groups who consider ability highly incremental set higher goals for themselves, put more effort into the intervention and performed better than the students whose personal views on ability were less incremental.

Therefore it would have been interesting to see if there was a way to pair students with different personalities and views on ability and how they would have impacted their partners.

Although it would have been hard to place students into personality categories according to their learning strategies, especially when considering that many students use diverse learning techniques (Jolliffe, 2007).

9.3 Methodological Limitations:

This section looks at methodological limitations by looking at the limitations associated with the experimental design, the issues around the implementation analysis as well as problems with the group level analysis of the lesson materials.

9.3.1 Experimental design limitations

In order to effectively evaluate the limitations of this research it is necessary to first outline the evaluative benchmarks for research trials:

Experimental design benchmarks: In terms of measuring the effectiveness of a particular intervention on educational performance outcomes, or even process of learning outcomes, the golden standard is a Randomised Controlled Trial (RCT). However, even with RCTs there are many associated problems (Slavin, 2008b: Lachin, Matts & Wei, 1988). Therefore the following benchmarks, taken by Consolidated Standards of Reporting Trials (CONSORT, 2010) (medicine field) and the ‘evidence-based’ literature, concentrate on issues of reliability, internal and construct validity, all of which, if effected, pose problems for external validity (Hair, et al., 1998):

1. *Scientific background.* Research needs to provide scientific background and the rationale for the research.
2. *Objectives.* Research objectives need to be clearly defined.
3. *RCT or a matched design with description of sequence.* In Education RCTs are considered to have more credibility than simple surveys, ex-post facto or quasi-experimental designs when evaluating the impact of a particular intervention (Holland, 1986, Torgerson, C.J., & Torgerson, D.J., 2011; Trochim, 2012). This issue relates mostly to internal validity.
4. *The groups need to be strictly matched or strictly randomised, such as blocked randomisation.* This is an issue of internal validity. Randomisation does not necessarily guarantee group equalisation, especially when the sample size is small (Slavin, 2008b: Lachin, Matts & Wei, 1988). The solution is either blocked randomisation on performance data (Lachin, et al., 1988) or matching the groups according to important characteristics (Slavin, 2008b).
5. *Changes to trial design.* The research has to report changes to the design.
6. *Participants.* The study needs to identify the eligibility criteria for the participants.
7. *Setting.* The study needs to describe participants' background.
8. *Intervention.* Researchers need to describe the intervention in detail.
9. *Outcomes.* The study needs to pre-define what outcomes are being measured.
10. *Conducting statistical power analyses for the purpose of sample size determination.* Statistical power analyses are useful in order to aid researchers in deciding whether to reject or accept the null hypothesis after they have decided on the p value, protecting them from conducting either type I errors (wrongly rejecting the null hypothesis, concluding the intervention was successful) or type II errors (wrongly accepting the null hypothesis, concluding the intervention was

not successful) (Hair, et al., 1998). Hence, power analyses can to some extend influence external validity claims.

11. *Implementation fidelity.* Without implementation fidelity data it is hard to measure the true effect of a particular intervention (Arkoosh, et al., 2007; Bradshaw, et al., 2008; Gresham, 1989), thereby reducing internal validity.
12. *Standardised instruments.* In peer tutoring, as in other fields, meta-analyses have shown that researcher-made instruments yield a higher effect size than standardised instruments (Cohen, et al., 1981). This issue relates mainly to 'instrument reliability' and 'construct validity', two terms which are used to refer to what is being measured. The Best Evidence Encyclopaedia even goes as far as to suggest that studies using newly modified instruments or instruments which are inherent to a treatment should be excluded from systematic analyses or be reported separately (Slavin & Madden, 2008; 2011). However, a better argument would be whether the instrument used is the one appropriate for the task at hand rather than just using an instrument simply due to its high reliability and construct validity. Although a reliability and construct validity analysis would need to be undertaken in order to establish the strength of the instrument.
13. *Blinding.* Randomisation, implementation and evaluation to be blinded, so that each stage is managed by different researchers.
14. *Recruitment.* Identifying how the recruitment of the participants was conducted.
15. *Report the reliability of the instrument if previously used, if new, the instruments' reliability needs to be assessed by an independent third party.* In order to avoid bias, the reliability of the instruments needs to be tested (CONSORT).

16. *The length of the intervention needs to be at least 12 weeks.* According to the Best Evidence Encyclopaedia, studies need to be more than 12 weeks long in order to compensate for the Hawthorne effect. (Slavin & Madden, 2008).
17. *Ethical considerations are met.* No mental or physical harm should take place during the experiment, consent should be acquired from students, parents and teachers, and everyone participating in the experiment must be given the opportunity to participate at a later date (Morrison, 2001)
18. *Ensuring that the content of the topic in which participants are to be tested is covered in all experimental groups.* This is also an issue of internal validity. A major problem in educational research is that very often when a new teaching method is implemented the students in the experimental group cover completely different topics to the control group, which explains the high effect sizes for some educational interventions (Slavin, 2008a).
19. *Identifying the proposed statistical methods, and their appropriateness in answering the question at hand.* This is required in order to prevent researchers from 'fishing' for positive findings. (Moher, et al, 2010). This would impact internal and external validity.
20. *Making sure that the required statistical assumptions for a particular statistical method are met and reported.* Again, the findings would be biased if the appropriate statistical assumptions are not met (Hair, 1998), thereby effecting internal validity and external validity.
21. *Identifying the attrition rate for each group.* This is necessary in order to determine the level of external validity of a particular intervention (Moher, et al., 2010).

22. *Baseline data.* Providing information regarding the nature of the groups prior to the intervention.
23. *Outcomes and estimation.* The outcome data to be reported in terms of effect size and significance levels.
24. *Ensure that the trial does not become contaminated, i.e. social threats such as participants from the experimental group consulting with the control group.* Another very common problem with experimental interventions is that of contamination, with the best remedy being to conduct a Clustered RCT, as opposed to a simple RCT. (Trochim, 2012). This would impact internal validity.
25. *At what level are the results generalizable? I.e. the issue of external validity, with stratified random selection being one of the highest external validity points*. The biggest problem associated with experiments in general is that they lack external validity, due to a lack of stratified random population selection (Trochim, 2012).
26. *Harms*. Reporting if there was harm to any of the participants.
27. *Limitations*. A section providing the limitations of the research.
28. *Interpretation*. Findings to be interpreted.
29. *Funding*. Sources of funding bodies to be identified.

Evaluating the ICAT experimental design. Based on the above CONSORT checklist, the experimental design applied in this research falls short in the areas of the following elements:

Dealing with design issues and internal validity. The main limitation with the experimental design adopted here was that it was not a RCT. A RCT was not possible due to the schools' inability to cope with the radical changes to their classes. Specifically, the main concern was that teachers were anxious about having classrooms with students they were not familiar with; an anxiety further strengthened by the fact that the intervention already involved splitting classes in half and pairing the older students with younger students, as a mathematics head teacher from school C made clear. Therefore, the lack of strong RCT introduced bias at many levels, especially resulting in unequal groups in terms of performance as the pre-tests indicate. The lack of clustering at the school level could have also demotivated the students in the control groups and their teachers (Trochim, 2010).

Lack of standardised instrument. Another limitation of the study was the lack of standardised tests. The lack of standardised tests with strong reliability and construct validity has negative implications for generalisation and external validity. The choice of using researcher-made instrument was made in order to capture the impact of the intervention with more precision, since the students only covered 6 mathematics topics, and a broad standardised test would have been unable to capture the impact of such a short intervention with precision. Consequently since there was a genuine limitation with construct validity, specifically that the instruments only measured limited mathematic aspects, it was not possible to make any statements or discuss the impact of ICAT on mathematics as a whole.

Power analysis. This was conducted prior to the experiment to investigate the size of the sample needed. A limitation in terms of power analysis was that the necessary sample size was not achieved for school B, rather than having 42 participants in each group, the groups ranged from 17-35. The reason why this is important is that we could have wrongly accepted the null hypothesis, hence committing a Type 1 Error as a result of the p value being non-significant due to the low sample size. In other words, if school B had a larger sample size their group mean difference would have shown a higher probability to be significant.

Implementation fidelity limitation. Implementation fidelity was measured in two ways, via structured observations and the collection of lesson materials for every pair for each week.

The pair observations, however, were relatively short. It would have been more useful to have spent more time observing the pairs directly or through a visual recording mechanism.[19]

Ensuring that the content of the topic in which participants were to be tested was covered in all experimental groups. In every school the younger students in the control groups covered the same mathematics topics as the intervention groups. However, this was not the case for the tutors, which led to the reason for not measuring their mathematics performance.

Making sure that the required statistical assumptions for a particular statistical method were met and reported. All the necessary statistical assumptions were conducted to ensure the validity for using ANCOVA. However, there were many violated assumptions for some of the variables.

Identifying the attrition rate for each group. For schools B and C the attrition rates were slightly high for some of the groups, both in terms of mathematics performance as well as attitude variables. Specifically school B showed an attrition rate of 60% in terms of performance for the Year 9 (tutors) control group, and 40% for the attitude variables; as well as 34% attrition rate for the Year 7 (tutee) control group. Also, school C showed a high attrition rate for Year 8 (tutees) performance, 41%, and on attitude variables 43%. These high attrition rates are concerning, since the higher the attrition rates for each group the farther the mean for each group in the analysis from the actual school's population group mean. Therefore even though the study identified the attrition rate, the actual results pose concern.

Ensuring that the trial did not become contaminated, specifically, social threats such as participants from the experimental group advising the control group. The intervention was extremely complex for contamination to take place. However, there was a main issue regarding contamination which took place in school C, in which the management decided from the beginning to print more materials and roll out the intervention to the entire age

[19] The next section covers the limitations on implementation assessments.

group; in the analysis this was not treated as a trial contamination, it was treated as design alteration due to its scale.

Blinding. The study would have also benefited from blinding procedures related to data coding or data analysis, which among others would have aided the credibility, analysis and interpretation of the findings. Again, the ‘time limit’ was the main issue regarding this aspect.

The length. Another major limitation of the project was that the intervention lasted only six weeks. It would have been useful if the intervention was at least 12 or 16 weeks long to provide further assurance.

Generalisation and external validity. Finally, in terms of the experimental limitations, another issue was the extent to which the findings are generalizable to other populations of the same or a similar background. The lack of a proper stratified random selection, the small sample size, and issues of internal validity have all negatively impacted the extent to which the findings can be generalised.

9.3.2 Implementation limitations

Limitations regarding the implementation investigation can be divided into two areas: limitations for the school comparisons of the lesson materials and limitations for the school comparison of the actual observations.

Limitations for the school comparisons of the lesson materials: The following are some of the limitations to bear in mind for the school comparison of the lesson materials when investigating implementation:

Firstly, statistical differences and effect sizes between the schools in terms of lesson materials could have been derived simply from school differences, especially when considering that students' ages and peer tutoring materials differed across the schools.

Secondly, there was a high difference in the number of participants from each school on lesson materials, especially between school A, n=23 and school C, n=44. This partly explains why all the F values were high and significant when comparing tutees' lesson materials on various social interdependent indicators.

Finally, certain topics such as 'data interpretation' and 'representation', as in the case of school B, or 'enlargement', as in the case of school C, required students to be creative and productive, activities which according to Bloom's learning taxonomy require more mental power and consequently more time. Hence partially explaining why school A showed a higher rate of attempted questions, or more correct answers.

Observations limitations: The main limitation of the observations was that there were not enough participants, thereby making the data problematic. There were at least two inter-related factors why not more observations were conducted: Firstly, the three schools were located in different counties of England, and secondly the intervention only lasted for six weeks.

Also, the actual observations were short, 30-40 seconds per window per pair; hence there was not enough time for the students to feel fully comfortable around the researcher. Although this was the case with only the first window of observation, during the remaining four windows the students were more relaxed. Such short window observations can explain why for example indicator 'tutee self-correct' was not picked up, and was observed only once in one of the pairs for all the schools.

It is also not clear to what extent the presence of a researcher impacted the students' interactive behaviour, one of the limitations regarding observation research in general.

Moreover, the observed lesson materials for each school differed in nature. Certain mathematic concepts could have required or motivated the students to interact, or not interact.

Finally, the study would have also benefited from an additional observer to cross check the implementation level.

9.3.3 Group level analysis of tutees' lesson materials

This section looks at the limitations of the group level analysis of tutee's lesson materials. Specifically, it looks at two areas, limitations derived from social interdependent indicators used to conduct the comparison, and limitations of the actual analysis of the data itself.

Limitations derived from social interdependent indicators: A substantive limitation of the group level analysis of tutees' lesson materials was that some of the indicators were not previously established and well tested since most of the previous research on peer tutoring does not concentrate on evaluating students' lesson materials. Some of the indicators in mind were those found under the category of '*quantity of exercises'* and '*feedback'*. Specifically, those are:

- Quantity of exercises attempted in the practice section
- Quantity of exercises attempted in the connect section
- Quantity of exercises attempted in the turn-taking section
- Feedback by ticks/crosses (in various sections).

It could be argued that another important limitation of the indicators used here is that most of them did not measure specific elements, in other words they did not concentrate on measuring either cognition, or level of task interdependence. Although theoretically it makes sense to place these two areas under one indicator at the practical and interpretation level, such indicators were problematic in the sense that no specific statements could be made.

Limitation deriving from the actual analysis: Another limitation of the group comparison of tutees' lesson materials relates to the actual analysis.

Firstly, an average of 37.8% of all tutees did not have their ICAT lessons enter the analysis due to a lack of clarity in terms of who the lesson materials belonged to, since some students did not write their names on the ICAT lesson materials. Especially for school A, in which up to 50% of the students did not do so due to a breakdown in communication.

Secondly, the actual analysis contained unequal groups. Specifically, the higher performing group of tutees included 60 participants and the lower group 42. As already mentioned above, the reason why this matters is that a high difference in the number of participants between the groups can increase the t-test or F values, and as a consequence the p values.

9.3.4 Other Limitations

At least six further additional limitations of the paper can be identified, those are; the lack of a qualitative investigation, the lack of possibility for the teachers from different schools to get together, the lack of a second marker, the lack of a comparative regular cross age peer tutoring group, collecting additional data and the absence of rolling out the intervention to the control groups. The first five issues are the result of the time constraint element, with the last limitation being due to the schools' decisions and contexts.

The need for qualitative investigations. The study would have greatly benefited from qualitative investigation such as open-ended interviews, open-ended questionnaires, group discussions or action research initiatives, which could have been applied to students, teachers, or head teachers. Especially when considering that the intervention is new in nature. Such insights would have been extremely helpful in identifying not only how everyone perceived the intervention but also to give clues, hints and recommendations for future directions for the development and the implementation of ICAT. Such additional approach would have improved research triangulation. Triangulation is helpful in order to ensure a detailed picture (Altrichter, Feldman, Posch, Somekh, 2008; Cohen & Manion, 2000) and to cross check the findings (O'Donoghue & Punch, 2003).

Extra support for the school. The study would have also benefited from some form of space in which the teachers or head-teachers could have exchanged ideas prior to, during and after the intervention. Such networking spaces are used in peer learning (Jolliffe, 2015), or more specifically in cross-age peer tutoring (Shared Maths project managed by Durham University), and could have provided more confidence and additional help to everyone. Also, the teachers could have benefited from a handbook of ICAT in order to provide them with additional guidelines and serve as a map. Handbooks are recommended to be effective in aiding the implementation of cooperative interventions (Jolliffe, 2011). Finally, another element which would have enhanced the implementation of cooperative learning and benefited both the schools and the students would have been the assessment of group work skills and learning skills (Jolliffe, 2011). This could have been carried out either by the teachers, the facilitators or the researcher, who can then provide feedback in real time.

Lack of second marker. Another limitation of the study was that there was no second marker to cross check the tests, to conduct the observations (as mentioned earlier), or cross check the lesson materials. A second researcher would have strengthened any reliability issues deriving from the work load.

The lack of a comparative regular cross age peer tutoring group. The study would have also greatly benefited from a three way factorial design, in which one group adopted ICAT, another adopted a regular cross age tutoring intervention and another group served as a control. This would have provided more insight in terms of where ICAT is positioned within the peer tutoring and cross age tutoring literature. The main reason for not proceeding with such design is the apprehension that the schools would have dropped out from such a complex intervention.

Collecting additional data. The study would have also greatly benefited from systematic collection of data relating to how ICAT was perceived by the students and the teachers. This is a great limitation considering that ICAT promises to introduce something slightly new to normal cross age peer tutoring. Although overall the head-teachers regularly commented that the teachers and the students enjoyed ICAT, information coming directly from students and teachers would have been greatly helpful for future ICAT developments.

Further investigations with the control group and follow ups. Finally the study would have also benefited from further investigation with the control group, which was the initial project design presented to the schools. Since school C dropped the control group on the first day of ICAT, this option was open for the two remaining schools A and B. While school B decided not to continue with the project any longer, perhaps due to the small effect size found for this school, most of the Year 6 control group students at school A had moved schools. School A did show an interest in continuing ICAT with other classes, as USB sticks with all the intervention details and plans were provided to all three schools for their future professional development.

9.4 Chapter conclusion

This chapter has covered the limitations.

In terms of theoretical limitations, the intervention would have benefited from incorporating additional social interdependent elements such as 'reward interdependence' and the appropriate pairing on cross-ability for school B.

Most of the limitations of this paper are located at the methodological level and some derive due to the inability to have continued to work on the original projects. Within the methodological level, most limitations derive from the experimental design and instruments used to test the intervention. Specifically, the study would have greatly benefited from a clustered (at the school level) randomised blocked and controlled trial, with blocked randomisation at the stages of participant selection as well as allocation. The design would have been stronger in a two stage design, groups changing places after a certain period, across two years via incorporating standardised tests and blind procedures.

Also, qualitative data collection, open-ended interviews or questionnaires at multi-levels (student, teachers, and head-teachers) would have provided further triangulation and insight into the findings, as would have other directions such as double markings, extra support for the schools, three way factorial designs and fellow ups or collecting additional attitude data from students and teachers regarding ICAT.

By pointing out the limitations the chapter has simultaneously, albeit indirectly, also pointed out the strengths of the paper. Overall, one of the main strengths of the paper has been its methodology as a whole, specifically the effort to provide what is called “multi-level triangulation” (Denzin, 1978); theoretical triangulation, with social interdependence being the broadest cooperative learning theory, data collection triangulation, data analysis triangulation, and data interpretation triangulation.

The final chapter provides an overall conclusion as well as making various recommendations for future cooperative learning and ICAT implementation and development.

10 Conclusion and Recommendations

10.1 Introduction

This thesis started by outlining a theoretical review of peer tutoring, concentrating on both traditional and current theories. It concluded that the theories in peer tutoring share many similarities. Specifically, the theory shown to have accommodated most other theories was identified to be Social Interdependence Theory. The findings then influenced the remaining empirical reviews.

The empirical review chapter started by outlining common peer tutoring frameworks. After reviewing and comparing the elements of peer tutoring interventions, such as cross-age peer tutoring, RPT, CWPT, and PALS, the next step was to conduct a detailed review of meta-analyses and other reviews in peer tutoring. The aim was to understand the impact of various small group learning or peer tutoring characteristics on students' social and academic gains. The examination of meta-analyses and systematic reviews then identified peer tutoring elements which provided the highest effect sizes, those being: goal setting by students, reward setting by students, and autonomous or informative peer tutoring structures, elements associated with Social Interdependence Theory.

The investigation of meta-analyses then led to a methodical evaluation of the extent to which past peer tutoring interventions applied social interdependent components and a systematic review of peer tutoring effects in mathematics in terms of improving performance, social and academic self-concept.

In order to evaluate the extent to which past peer tutoring interventions applied social interdependent elements or a broad/pragmatic perspective, 11 benchmarks were used and over 127 articles were evaluated. Concluding that a) most peer tutoring interventions were conducted in literacy, followed by mathematics and then science, b) many peer tutoring interventions appeared not to apply broad theoretical components, and c) that the theoretical elements which have shown to provide the highest effect sizes in peer tutoring (according to the meta-analyses) have been the least implemented to date, those being: goal interdependence, reward interdependence, and resource interdependence which aids students' autonomy. Although not evaluated here, the number of papers in which students set their own goals and rewards, characteristics for which the highest effect sizes are found, is even smaller considering that only RPT offers such options.

Also, in order to explore what the picture looked like in peer tutoring mathematics a systematic review was conducted. The systematic review in mathematics investigated cross-age peer tutoring, same-age normal and reciprocal peer tutoring, in terms of their mean average effect size in performance, academic attitude and social attitudes. Again, it was revealed that the highest effect sizes were found in those peer tutoring interventions which incorporated many social interdependent elements, such as cross-age peer tutoring, and same-age reciprocal peer tutoring (with all the reciprocal peer tutoring interventions in the systematics review making use of goal and reward interdependence).

The theoretical and the empirical reviews influenced the research aim, objectives and the research questions. Specifically, the aim was to incorporate the peer tutoring elements which provided the highest effect sizes and trial the new method in mathematics areas to investigate its impact on mathematics attainment and attitude variables, as well as to conduct implementation analyses and explore how tutees who gained the most from ICAT differed from those who gained the least.

Due to the time and financial limits a six week pre post-test quasi experimental design study was planned with three different schools, with students of different ages and in different areas of England; with two of the schools, B and C, deviating from the research design. The teachers were allowed to choose the topics, so that the peer tutoring and the control groups concentrated on the same mathematics subjects as planned during their curriculum. New instruments were devised to measure attainment, and for attitude measurement previous sub-scales were put together and tested and re-developed with trial schools in Durham County. The investigation of ICAT's impact on broader aspects of learning, such as social as well as academic attitudes, was one of the main objectives of this research, allowing for richer analysis of ICAT's potential. Consequently data were collected on performance, attitude, lesson materials of ICAT and observations.

All instruments were evaluated for reliability and coding procedures were outlined for various tasks. The findings were reported for implementation and the evaluation of the trial as well as for tutee group comparison on the lesson materials. Finally, two forms of discussions were provided, one which addressed the findings and one which addressed the theoretical and methodical limitations. Although kept under separate chapters for clarity and emphasis, sometimes the discussion on the findings and the discussion on the limitations showed similarities due to the complexity of the investigation.

It must be made clear that the findings of this research cannot be generalised due to methodical limitations. And although the effect sizes for various instruments were high, favouring ICAT conditions, it is very likely that other aspects connected to methodology limitations can also count for their explanation.

Although the original plan was to conduct the research within clustered RCT projects, there were also benefits of not being part of those projects: a) the researcher experienced more control or perhaps shared control with the teachers, b) there was more room for the literature to influence the development of ICAT, which led to multiple data collection types, c) experience was gained in terms of developing, implementing, managing and evaluating short trials.

Overall, the attainment findings of this research are consistent with previous meta-analyses in cooperative peer learning (Roseth, et al., 2008; Othman 1996), meta-analyses in peer tutoring (Cohen, et al., 1982; Ginsburg-Block, et al., 2006; Leung 2014; Rohrbeck, et al., 2003), the mathematics systematic review provided here, cross-age studies in mathematics (Bar-El 1982; Fitz-Gibbon 1990 ; Topping, et al., 2003 ; Tymms, et al., 2011) and same-age reciprocal studies in mathematics which have been influenced by Social Interdependence Theory (Fantuzzo 1990; Fuchs, et al, 1997; Ginsburg-Block & Fantuzzo 1997; 1998; John, et al., 1992; Menesses & Frank 2009).

ICAT was developed to foster student engagement in as many diverse interconnected areas as possible in a joint and fun way. Specifically, what is crucial about ICAT is that all parts required students to engage in different processes of thinking (cognitive, meta-cognitive), exercises with different complexity levels to approach the ZPD (in each of three remaining parts), different interaction styles (in each of the parts), as well as having to think back to the past (in the connect and turn-taking sections), and forward (goal-setting, practice-test, connect sections). Participating in such a learning structure, while making sure to answer the questions correctly, is a socially and academically demanding and rewarding task.

Prior to providing recommendations on how to implement and further develop ICAT it is necessary to first outline the recommendations provided by other researchers involved in cooperative group learning and cross-age peer tutoring. This is necessary for two purposes: firstly, in order to illustrate how ICAT is similar or different from other peer learning interventions, secondly, to place the ICAT recommendations within the broader peer learning framework:

10.2 Recommendations for implementing co-operative group learning

A detailed recommendation of how to implement cooperative learning is presented by Abrami, Chambers, Poulsen, DeSimon, D'Apollonia and Howden (1995). They pay attention to seven broad and interrelated aspects of cooperative learning: 1) class building and team building, 2) how to group students, 3) how to foster positive interdependence, 4) how to encourage positive accountability, 5) how to develop interpersonal and cognitive skills, 6) how to evaluate and reflect, 7) and how to use cooperative learning for various classrooms.

One of the most important aspects in successfully implementing cooperative learning is to make the teachers believe why and how the intervention is successful. Abrami, Poulsen, Chambers (2004) collected cooperative learning implementation questionnaire data from a sample of 933 teachers. They measured three broad categories; First, how the teachers perceived the value of the intervention, whether cooperative learning was consistent with their philosophy, career advancement or whether they though the students would benefit in terms of attainment, attitude and interpersonal skills. Secondly, teachers' expectancy, such as whether teachers expected the cooperative group learning to be successfully implemented; specifically, if the teachers acquired the necessary self-efficacy and skill, as well as the right classroom environment or school support. And finally, if they thought that cooperative learning was worth the cost, in particular physical and psychological burdens in terms of time and effort. The authors report that the best category to predict whether the teachers would make use of a cooperative technique was its perceived expectancy of success, with an effect size of 0.53, then cost with an effect size of 0.42, and value with an effect size of 0.36. They conclude that teachers need to be trained to have the necessary skills and knowledge in cooperative learning so that they develop strong self-efficacy and perceive the method as being easy to implement.

Other researchers have also gone to the root of understanding what the successful implementation of a cooperative learning strategy involves. Recent studies concentrate explicitly on how to succeed at implementing a cooperative learning intervention, taking both qualitative and quantitative directions in doing so (Jolliffe, 2005, 2007, 2011, 2015; Jolliffe & Hutchinson, 2007).

The following 10 points are given as necessary practice to succeed at implementing cooperative learning interventions (Jolliffe, 2011):

1. Commitment from the whole school and linking the intervention to other key priorities within the school.
2. Training the teacher on the theoretical aspects of cooperative learning.
3. Making sure that there is a transition from informal group work to a formal setting.
4. Having an inside facilitator to help and work closely with the teachers.
5. Ensuring that key elements of cooperative learning are in place, such as positive interdependence and promoting interactions.
6. A programme to continually teach small group skills and helping students in need.
7. Supporting teachers with planning tasks and the implementation of cooperative ideas.
8. Helping teachers with group composition.
9. The need to assess both group work and learning skills.
10. Being part of a network of schools in a similar position.

10.3 Recommendations for implementing cross-age peer tutoring

Some of the interdependent elements which are thought to be necessary are already evident in cross-age peer tutoring. Specifically, the older student has a clear role of being the tutor, this clarity is emphasised by the age difference and the fact that the tutor comes from a different classroom. Also, individual accountability is higher in a cross-age setting as compared to same-age one to one or same-age higher group settings. The reason for this is that the older students know that they are accountable for providing help when asked for it, this goes back to the idea of responsibility and role theory (Sarbin, 1976), and the younger student is accountable for understanding and concentrating on the topic, since they are constantly being monitored by an older student.

Overall monitoring is also important when implementing peer tutoring. The following are some additional recommendations from Fitz-Gibbon's (1992) on what a monitoring system should include in order to improve tutoring:

a) Information feedback at every school management level.

b) Clarity in terms of measuring performance and giving advice.

c) Agreement of the monitoring type by everyone within the school.

d) Positive behavioural changes by providing more attention.

e) Not to place a heavy load on the system it monitors, i.e. by taking too much time and space.

Fitz-Gibbon (1992) realized that monitoring systems can place great obstacles in the way of enhancing intrinsic motivation, as reported by (Buttler, 1988), however improve other forms of motivation, as suggested by Parson (1974). Fitz-Gibbon's argument against Butler's conclusion is that it is the type of feedback that is given which is the main issue in determining the impact of a monitoring system on *intrinsic motivation*, rather than feedback from monitoring in general.

Fitz-Gibbon (1985, 2000a, 2000b), Tymms and Merrell (2015), and Thurston (2014) also share many of the recommendations outlined below on how to improve the implementation and management of a cross-age peer tutoring intervention:

1) Identifying the tutors and the tasks. Teachers should take time to identify which is going to be the tutor class. Also, teachers need to specify their tasks according to what socio-psychological factors they need to improve, i.e. if cognition is what is strived for then teachers should ensure that the task is defined and testable.
2) Ensuring the tutors are suitable. The tutees should be two years younger than the tutors. This is necessary if the tutor is socially at risk. Hence, the tutor needs to feel comfortable in his/her role.
3) Locating a venue. The best rooms for tutoring are those with much space in which booths are provided, enabling the pairs to work without being disturbed. If booths are not available or cannot be created with display boards, then ensuring that the students sit at their tables while facing outwards from the centre of the room so that disturbance is minimised.
4) Pre-test. It is recommended that both the tutors and the tutees have had some kind of pre-test in order to identify their ability levels. Then pair the top tutor with the top tutee, the second best tutor with the second best tutee and so on down the line.
5) Train the tutors. The tutors need to be pre-trained in order to reduce their anxiety of the unknown.
6) Prepare materials. Tutors could help to prepare the materials by producing flash cards or cards which have a problem on one side and the answer on the other.

7) Manage the situation without heavy interference. Teachers need to observe how the students are teaching one another. If the teacher is not happy with the way the tutor is tutoring it is better to provide feedback to the tutor after the session rather than during.
8) Test the tutees, then discuss the outcome with their tutors. By testing the tutees this will convey to the tutors the seriousness of the activity. Certain diagnoses and actions can then be taken.
9) End the project and prepare for the next one. Rather than using cross-age tutoring throughout the year without a stop, it is better everyone takes three week breaks in order to emphasise different and important parts of the curriculum.

Further crucial elements in peer tutoring interventions are the implementers, i.e. teachers and helpers (Lippit, 1976). Lippit, suggests that there should be at least four important helpers in any peer tutoring intervention; a) a volunteer teacher from the tutor class, b) a volunteer teacher from the tutee class, c) a supportive administrator for the teachers, d) and someone to train the tutors. The crucial helper for the tutors is the tutee's teacher, who to some extend is the person asking for their help (Lippit, 1976). Although Lippit's suggestions are important, one has to bear in mind that one of the main benefits of peer tutoring is its high academic impact with low economic costs, and that peer tutoring is desirable precisely due to a lack of extra teachers. Lippit's views are consistent with the more recent Abrami, (2004) study, in the sense that it too acknowledges the teachers as important when it comes to successfully implementing a peer learning intervention.

10.4 Recommendations for implementing ICAT

Most of the steps to successfully implement ICAT are similar to cross-age peer tutoring in general as outlined by Carol Fitz-Gibbon or those outlined in the broader cooperative learning literature by names such as Robert Slavin, Bette Chambers and Wendy Jolliffe. The following additional 10 steps are important when a researcher or mathematics head teacher intends to implement ICAT in a school setting:

1) Start planning for ICAT well in advance. The success of ICAT, just like any other educational intervention, is dependent on the amount of preparation carried out beforehand. The general rule is that the longer the intervention is going to be the more preparation time is needed.
2) Getting the teachers involved in organisational issues. Rather than the researcher or the head teacher identifying how to best implement ICAT in a school, attention should be given to teachers' preferences and ideas. This is for two reasons: firstly, to make the teachers comfortable and knowledgeable in terms of the method, and secondly, in order to make the teachers feel part of the main decision which in return will make them associate with the method better, especially when the teachers see that the ICAT framework consists of questions they created or chose. In other words, the teachers should be given some choice in terms of timetabling, setting, number of students to be taught, etc. This process increases the teacher's interdependence to the project.
3) Training the teachers. Teachers should be trained in both how to train students in interpersonal skills and in the actual academic ICAT procedures. A one-to-one role play with the head teacher, or the researcher, would be the most efficient way, and would give the impression that the intervention is important.

4) Getting the teachers involved in producing the exercises and materials. In order to further make the teachers engaged in the project the teachers can be given the ability to choose some of the exercises and materials that they would like the students to concentrate on. This way ICAT is personalised to the teachers, and gives them a sense of ownership of the method. This is necessary especially when considering that most teachers concentrate on ensuring they cover the correct curriculum material, for the correct student level. Having all the exercises coming from an outsider or from the head teacher would go against teachers' overall aim, and the exercises are more likely to be at an inappropriate level for the student. Also, most of the exercises should be produced prior to the intervention so that when the intervention starts the teachers do not have to worry about finding what exercises to choose. This is important since ICAT, like most cross-age tutoring procedures, is more complex to manage, as students move classes and there needs to be communication between teachers. Hence, there would not be enough time to start preparing the exercises just before the session.
5) Ensure the materials are ranked from very easy to very difficult. There are three main reasons for this: Firstly, this is necessary for parts two and four of the ICAT, and will help students to monitor themselves and each other within each part. Secondly, such ranking would also improve task interdependence since it will give the impression of a game, and make it more enjoyable to the students. Finally and most importantly, since many teachers will not know what level the tutors or the tutees are, as they will come from different classes, they will find it very hard to give the correct level of the exercises to a particular pair. However, if the exercises are ranked from very easy to very hard there will be exercises for every level.
6) Students need to be trained in both social and educational skills. Although one to one tutoring interventions do not need as much training in interpersonal communication

skills, since the dynamics are not as complex as they would be in a group of 3, 4, 5, the students still need to be taught the basic skills of how to ask questions politely, praise, and be nice to one another. Training for both interpersonal skills and the ICAT procedures would be best in a one-to-one role play.

7) Initial sitting. During the training session and the next few sessions the pairs have to sit on the correct sides of the desk, in order to better see the instructions written on the ICAT materials. This is necessary for the students to internalise the mechanics of ICAT in a fast and accurate manner.

8) Ensure that the students are setting the goals. At the heart of ICAT sits goal interdependence. Therefore, in order to get the students even more engaged and enjoy their tasks, make sure that they have all set a goal, which they will work to achieve. This is necessary because the goal setting system in ICAT is designed in such a way that the chances for the students to win are maximised, so that every ICAT session positively influences students' academic and social skill attitudes.

9) Monitor students' interactions. Teachers need to walk around the classroom and monitor both, students' interpersonal and academic skills. Most of the feedback should be given to the students in good time, and preferably after an interaction has finished at the end of the lesson, to make the students feel independent. Monitoring is also important in order to make the tutors feel comfortable and confident, specifically teacher monitoring would serve as a back-up for those tutors who do not know how to answer tutees' questions.

10) Help from the organisers. The head teacher or the researcher should consult with the teacher on a weekly basis in order to provide further support and encouragement. Teachers could also arrange specific times during which they can meet and exchange more information regarding the ICAT.

10.5 Recommendations for future research and development of ICAT

The impact of ICAT needs to be further tested in other mathematics areas with better research methods. ICAT also needs to be further improved to incorporate other social interdependent elements or applied in different settings.

Methodological development. ICAT would benefit from future research which takes into account the methodical limitations identified in the limitation chapter.

Firstly, ICAT needs to be tested for a longer period than six weeks and in more mathematics topics. Longer timing will reduce the chances of the Hawthorne effect (McCartney, Warner, Iliffe, Haselen, Griffin & Fisher, 2007).

Secondly, in order to increase internal validity, ICAT would benefit from a Clustered (at the school level) Randomised Blocked/Minimised Trial. Although this would be expensive and require many organisers, the framework has already been established in this paper and there is a huge body of literature, which indicates that both cross-age peer tutoring and goal interdependent peer tutoring interventions, usually same-age reciprocal, provide the highest effect sizes in peer tutoring. Attention should however be given to identifying whether particular schools are implementing national or local policy interventions, which have been informed by research. This is important since the aim is to compare ICAT to normal classroom rather than to other interventions. As mentioned, Lemon, et al., (2014) reported that when they tested PALS with young children the effects were high in the first few years but then faded away, even though the schools still carried out the cross age peer tutoring intervention. However, on closer inspection they identified that the control schools were implementing governmental policies which were heavily influenced by the 'what works' literature. Hence the comparison was taking place between effective interventions rather than normal classroom settings.

Thirdly, in order to have a strong external validity, i.e. in order to generalise the findings to the population, schools should be chosen randomly. Also, in order to further strengthen external validity, the randomisation could be stratified on the basis of school performance, age and socio-economic background, as well as geographical areas.

Also, in order to further improve the construct validity of an ICAT intervention, future research should pay more attention to the instruments. Future ICAT interventions need to ensure that the instruments used are established and have high reliability and construct validity.

The difficulty of these recommendations is that many of these scientific ideas cannot be carried out by research students or academics operating on their own. The above methodological recommendations to test ICAT can only be applied by few institutions, such as the CEM centre at Durham University, the Institution for Effective Education at York University or Centre for Effective Education at Queens University. These institutions all have the experience and expertise of strong methodologies and peer tutoring, and the advantage to carry out blind allocation of schools to group or blind assessment.

Theoretical developments. In order to make ICAT even more interdependent, it would be useful if future research tested two further interdependent elements:

Firstly, after the students have checked if they have achieved their goals they could engage in a short 5 minutes conversation to reflect how their social and academic interaction went, and where it needs improvement. This would further improve their meta-cognitive skills.

Secondly, ICAT would also benefit from a choice between tangible or non-tangible extrinsic reward systems similar to that of reciprocal peer tutoring in which the students choose themselves what reward they want to achieve. This could be easily implemented since the point and goal system is already in place. However, this could place an extra burden on the teacher in terms of managing the intervention, especially when considering that cross-age tutoring is slightly harder to manage than same-age tutoring.

ICAT, however, does have a strong reward system in place, specifically the intrinsic reward system. Abrami, et al., (1995) and Jolliffe (2007) elaborate on how cooperative learning ticks many boxes of Maslow's hierarchy of needs triangle, such as educational needs, belonging needs and actualisation needs:

In terms of educational needs, ICAT contributes to both tutor and tutee educational needs more than normal same-age tutoring, since in a normal same-age tutoring situation the tutor could find himself being unable to provide explanations to the tutee (Paulo, et al., 1989). This in return would hurt both tutors' and tutees' academic needs since the tutee does not gain any new information and the tutor would not be able to engage in explanation and meta-cognitive activities.

Also, in a cross-age tutoring context the tutors are given the chance to improve their self-actualisation needs. Enabling the tutors to help others will in return make tutors feel good about themselves by becoming a better person. This could increase in a situation in which the tutee is younger and more vulnerable.

For the tutee, a cross-age system can improve belonging needs more than the tutor, since the younger student would have an older friend to look up to at school. This also improves tutees' safety needs as having an older friend will make tutees feel safer at school. All of these needs would be harder to achieve if the positive interdependence, or cooperative ideas, are taken out of the cross-age tutoring intervention.

Applying ICAT in different settings. ICAT can also easily be used in science education. Science exercises could be ranked from easy to hard and inserted into the model. The connect section of the model is also appropriate for students to link science concepts to what they have learned previously or to real life situations.

Also, on top of applying a choice of rewards interdependence to ICAT, it would also be interesting to see what the benefits of a ‘Reciprocal Interdependent Cross-age Tutoring’ (RICAT) are. ICAT already provides most of the foundations. In order to make it reciprocal, during the first 12 weeks the older student can be the tutor and the next 12 weeks the younger student can take the tutor role. To make this achievable the younger tutor could be aided by being provided with an answer sheet to the questions in order to assist the older student, as is the case in the same-age reciprocal peer tutoring (RPT). During the first 12 weeks the exercises can be adjusted for the younger students’ level, while during the second 12 weeks the exercises can be adjusted for the older students’ level. Such strategy ensures that both, the younger and the older students, take the role of the teacher, and having set roles for up to 12 weeks ensures role interdependence. Apart from improving performance, RICAT would also be a powerful instrument in the hands of the younger students and be expected to positively enhance their social and academic confidence, considering they would be helping an older peer.

Applying reciprocity to ICAT would make the intervention fully cooperative, and yield maximum social and academic benefits for both older and younger students.

References

Abrami, P.C., Chambers, B., Poulsen, C., DeSimone, C., d'Apollonia, S., & Howden, J., (1995). *Classroom connections - Understanding and using cooperative learning,* Toronto: Harcourt Brace.

Abrami, P.C., Poulsen, C., Chambers, B., (2004). "Teacher motivation to implement an educational innovation: factors differentiating users and non-users of cooperative learning", *Educational Psychology: An International Journal of Experimental Educational Psychology*, 24(2), 201-216.

Al-Hasan. S. (2003). Reciprocal Peer Tutoring effect on high frequency sight word learning, retention, and generalisation of first- and second-grade urban elementary school students. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database.* (UMI No: 3148152).

Allan, R. (2008). *MathsLinks 1; Year 7*. Oxford, Oxford University Press.

Allen, V. L. (1976). The helping relationship and socialization of students: Some perspectives on tutoring. In Allen, V. L. (ed.) *Students as Teachers: Theory and Research on Tutoring* (pp.9-27). London: Academic Press, INC.

Allen, V. L. (1976). Theoretical considerations. In Allen, V. L. (ed.) *Students as Teachers: Theory and Research on Tutoring* (1-9). Academic Press, INC: London.

Allen, V.L. & Feldman, C. (1973). Learning through tutoring: Low-achieving children as tutors. *The journal Of Experimental Education*, *42*, 1-5.

Allen, V. L., & Feldman, R. S. (1975). Decoding of students non-verbal responses. Technical Report, No. 365. In *Wisconsin Research and Development Centre for Cognitive Learning,* Madison, Wisconsin.

Allen, V. L., & Feldman, R. S., & Devin-Sheehan. L. (1976). Nonverbal indicators of Behaviour inconsistency by tutors: Nonverbal leakage. Working Paper No. 143 *Wisconsin Research and Development Centre for Cognitive Learning*, Madison, Wisconsin.

Allen, V. L. & Feldman, R. S. (1976). Studies on the role of tutor. In Allen, V. L. (ed.) *Students as Teachers: Theory and Research on Tutoring* (pp. 113-128). Academic Press, INC: London.

Allsopp, D.H. (1995). Classwide Peer Tutoring and the acquisition of algebra problem-solving skills for middle school students at risk of math failure in mainstream general math classes. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database.* (UMI No: 9607485).

Altrichter, H., Feldman, A., Posch, P., & Somekh, B. (2008). *Teachers investigate their work; An introduction to action research across the professions.* (2nd Eds.). Oxon; Routledge.

Argyle, M. (1976). Social Skills Theory. In Vernon L. Allen (eds.) *Students as Teachers: Theory and Research on Tutoring*. Academic Press, INC: London.

Argyle , M., & Ingham, R. G.. (1972). Mutual gaze and proximity. *Semiotica*. 1, pp. 32-49.

Argyle, M. (1976). Social Skills Theory. In Allen, V. L. (ed.) *Students as Teachers: Theory and Research on Tutoring* (pp. 57-71). London: Academic Press, INC.

Arkoosh, M.K., Derby, K.M., Wacker, D.P., Berg, W., McLaughlin, T.F., & Barretto, A. (2007). A descriptive evaluation of long-term treatment integrity. *Behaviour Modification*, 31, 880–895.

Arreaga-Mayer, Carmen., Terry, B.J, & Greenwood, C.R. (1998). Classwide Peer Tutoring. In Topping, K.J., & Ehly (Eds.). *Peer Assisted Learning* (pp.105-121). London UK: Lawrence Erlbaum.

Atherley, C. A. (1989). ‘Shared Reading’: An experiment in peer tutoring in the primary classroom. *Educational Studies*, *15* (2), .145-153.

Baker, H.R. (2005). Teacher Directed Instruction plus Classwide Peer Tutoring and the reading growth of first grade students. (Master's Thesis). Available from *ProQuest Dissertation and Thesis Database.* (UMI No: 1428630).

Barron. A-M., & Foot. H. (1991). Peer tutoring and tutor training. *Educational Research*, 33(3), 174-185.

Beckmann, N., Wood, R.E., Minbashian, A. & Tabernero, C. (2012). Small group learning: Do group members' implicit theories of ability make a difference? Learning and Individual Differences 22(5), 624-631.

Bland, M., & Haqrris, G. (1989). Same-age Peer Tutoring in the secondary school Science Curriculum. In *Paired Learning: Tutoring by Non-Teachers: Incorporating the Paired Reading Bulleting, ERIC Document Reproduction Services, 5.*

Bonarius, J. (1965). Research on the Personal Construct Theory of George A. Kelly: Role construct repertory test and basic theory. In Maher, B. A. (Ed.). *Progress in experimental personality research* (pp.2-46). New York: Academic Press.

Bradshaw, C.P., Reinke, W.M., Brown, L.D., Bevans, K.B., & Leaf, P.J. (2008). Implementation of school-wide positive behavioural interventions and supports (PBIS) in elementary schools: Observations from a randomized trial. *Education and Treatment of Students,* 31(1), 1–26.

Brawn, Tymothy A. (2006). Confirmatory Factor Analysis for Applied Research (Methodology in Social Sciences). The Guilford Press, London.

Breece, M. (2012). A Case Study of Peer-Assisted Learning Strategies among Eleventh Grade Students. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database.* (UMI No: 3536306).

Brei-Crawley. M.J (2002). Gender differences in partial interaction during after-school science peer tutoring program. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database.* (UMI No:3055262).

Browning, T.L.H. (1994). The effects of cross-age tutoring on the mathematics anxiety of fifth grade tutors. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database.* (UMI No: 1358426).

Bruner J S. (1978). The role of dialogue in language acquisition. In *The Child's Conception of Language* Sinclair A, Jarvella R & Levelt W J M (Eds) New York: Springer-Verlag.

Bruno, M.A. (2004). The effects of participation in a cross-age tutoring program on the writing behaviours and attitudes of the students in the role of the tutor. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database.* (UMI No: 3120725).

Burns, E. (2006). Pause, prompt and praise-peer tutored reading for pupils with learning difficulties. *British Journal OF Special Education*, *33* (2), 62-67.

Burns, M. (1985). The role of questioning. *Arithmetic Teacher*, February, 14–16.

Burns, M.K. (2004). Middle School Students' Perceptions: The Experience of Cross-Age Tutoring. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database.* (UMI No: 3144251).

Buttler, R. (1988). Enhancing and understanding Intrinsic Motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, 58, 1-14.

Byrne, D. (1998). *Complexity Theory and the Social Sciences.* London: Routledge.

Byrne, D. (2002). *Interpreting quantitative data.* California: Thousand Oaks

Byrne. D. (2008). Complexity, Cities and Social Action – a consideration of the implications of the complexity frame of reference for praxis. In Jan Bogg & Robert Gayer (Eds), *Complexity Science and Society*. Oxford: Radcliffe Publishing.

Cameron, F.L.J. (1982). A comparative study of cross-age and parent tutoring in two ESEA title I reading interventions: Academic and non-cognitive changes in primary-grade students. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 48106).

Campbell, D. (1969). Reforms as experiments. *American Psychologist*. 24, 409-29

Carberry. D.J. (2003). The effect of cross-age tutoring in reading on tutees, tutors and Meta-cognitively-trained tutors. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 3083266).

Caroll, B.P. (1995). An experimental study of the effect of cross-age on the mathematical achievement of ninth grade students enrolled in Algebra I. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database.* (UMI No: 9806649).

Cazden, C. (2001). *Classroom discourse: The language of teaching and learning*, (2nd Ed.). Westport, CT: Heinemann.

Chalk, K., & Lewis, A. B. (2004). Specific praise improves on-task behaviour and numeracy enjoyment: A study of year four pupils engaged in numeracy hour. *Educational Psychology in Practice*. 20, 335-351.

Chambers, B. (2011). Little ones working together. *Journal for Co-operative Studies. 44*(3), pp. 43-48.

Cilliers, P. (1998). *Postmodernism and Complexity.* London: Routledge.

Cilliers, P. (2001). Boundaries, hierarchies and networks in complex systems. *International Journal of Innovation Management*. *5*(2), 135-147.

Cilliers, P. (2005). Knowledge, limits and boundaries. *Futures*. *37*(10), 605-61.

Coats, L.B. (2007). Cross-age tutoring: Effects on reading achievement of tutors and tutees in an after-school program. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 3272963).

Coe, Robert. (2002). It's The Effect Size Stupid: What Effect Size is and why it is important. Paper presented at the British educational research association annual conference, Exeter, 12-14, September. Retrieved on the 8th of February 2012 From: https://duo.dur.ac.uk/bbcswebdav/pid-1799612-dt-content-rid-10944500_2/courses/EDUC59215_2011/EDUC59215_2011_ImportedContent_20110728055830/EDUC59215_2010_ImportedContent_20110722104611/Course%20Documents/Coe%202002%20effect%20size/Coe%202002%20effect%20size.pdfCohen,

Cohen, P.A., Kulik, J.A., & Kulik, C-L.C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, 19, 237–48.

Cohen, L., Manion, L. Morrison, K (2000). *Research methods in education (5th Eds.).* London: Routledge.

Conrad, E.E. (1975). The effects of tutor achievement level, reinforcement training, and expectancy on peer tutoring. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database.* (UMI No: 76-1407).

Cooper, K.J. (1973). A comparative study of five selected cross-age tutoring program. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database.* (UMI No: 73-31,711).

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of Processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour*, 11, 671-684.

Dancer, A.P. (2007). The effect of a required Character Education and Class-Wide Peer Tutoring program on 5th-grade students' reading and writing performance. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 3291628).

Daniels, L.J. (2004). The effects of a cross-age reading tutorial program: Hispanic students who are not proficient in reading tutored by non-Hispanic students who are proficient in reading. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database.* (UMI No: 3154193).

Deci, E. L., Eghrari, H., Patrick, B. C., & Leone, D. (1994). Facilitating Internalization: The Self-Determination Theory Perspective. *Journal of Personality, 62,* 119-142.

Deci, E, L., & Ryan, R, M. (1985). *Intrinsic motivation and self-determination in human behaviour.* New York: Plenum Press.

DePaulo. B.M., Tang. J., Webb. W., Hoover. C., Marsh. K. and Litowitz. C., (1989). Age Differences in Reactions to Help in a Peer Tutoring Context. *Child Development*, 60, (2), 423-439.

De Lisi, R., & Golbeck, S. L. (1999). The Implications of Piagetian theory for peer learning. In O'Donnell, A., & King, A. (Eds), *Cognitive perspectives on peer learning*. (pp. 3-37). Hillsdale, NJ: Erlbaum.

Dennis, L.M.G. (2013). Multiplication achievement and self-efficacy in third-and fifth-grade students: Effects of cross-age peer tutoring and skill training. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 3561581).

Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods.* New York: McGraw-Hill.

Deutch, M. (1949). A Theory of Cooperation and Competition, *Human Relations.* 2 (1), 129 152.

Dineen, J.P., Clark, H.B. & Risley, R.R. (1997). Peer tutoring among elementary students: Educational benefits to the tutor. *Journal of Applied Behaviour Analysis, 10* (2), 231-238.

Dion, E., Roux, C., Landry, D., Fuchs, D., Wehby, J., & Dupere, V. (2011). Improving attention and preventing reading difficulties among low-income first-graders: A randomised study. *Prevention Science*. 12 (1), 70-79.

Dufrene, B.A., Reisener, C.D., Olmi, J., Zoder-Martell, K., McNutt, M.R., & Horn, D.R. (2010). Peer tutoring for reading fluency as a feasible and effective alternative in response to intervention systems. *Journal of Behavioural Education*. *19* (3), 239-256.

Duran, D., & Monereo, C. (2005). Style and sequences of cooperative interaction in fixed and reciprocal peer tutoring. *Learning and Instruction, 15* (3), 179-199.

Dures, A. (1971). *Schools*. London: Batsford.

Dutch, M. (1962). Cooperation and Trust: Some theoretical notes. In Jones, M. R. (Ed.) *Nebraska symposium on motivation (pp. 275-320).* Lincoln, NE: University of Nebraska Press.

Eason, K.E. (2000). Cross-age peers as partners: Students' perspectives on the reciprocal effects of paired reading using cross-age peer tutors. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 9984328).

Early, J.W. (1998). The impact of peer tutoring on Self-Esteem and Texas Assessment of academic skills mathematics performance of tenth-grade students. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 9829500).

Ehly, S.W., & Vasquez, E.G. (1998). Peer Counselling. In Topping, K.J., & Ehly (eds.). *Peer Assisted Learning* (pp.219-236). London UK: Lawrence Erlbaum.

Ekstrand, C. (2011). The effects of peer tutoring on reading fluency. (Master's Thesis). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 1492146).

Elder-Vass, D. (2008). Searching for realism, structure, and agency in actor network theory. *British Journal of Sociology*. *59*(3), 455-73.

Ellson, D. G., Barber, L., Engle, T. L., & Kampwerth, L. (1965). Programmed tutoring: a teaching aid and a research tool. *Reading Research Quarterly,* 1, 77-127.

Ellson, D. G., Harris, PL., & Barber, L. (1968). A field test of programed and directed tutoring. *Reading Research Quarterly*, 3, 307 – 367.

Fantuzzo, J., Davis, G., & Ginsburg, M. (1995). Effects of collaborative learning and parent involvement on mathematics achievement and perceived competencies. *Journal of Educational Psychology*, 87*,* 272-281.

Fantuzzo, J.W., Davis, G.Y., & Ginsburg, M.D. (1995). Effects of parent involvement in isolation or in combination with peer tutoring on student self-concept and mathematics achievement. *Journal of Educational Psychology, 87*(2), 272-281. Fantuzzo, J., & Ginsburg-

Fantuzzo, J.W., King, J.A., & Heller, L.R. (1992). Effects of reciprocal peer tutoring on mathematics and school adjustment: A component analysis. *Journal of Educational Psychology*, *84* (3), 331-339.

Fantuzzo. J.W., Polite, K., Grayson, N. (1990). An evaluation of reciprocal peer tutoring across elementary school settings. *Journal of School Psychology, 28*(4), 309-326.

Felman. R.S., Shelagh, M.T. & Allen. V.L. (1976). Two field studies on cross-age tutoring in the school. *Wisconsin Research and Development Centre for Cognitive Learning the University Of Wisconsin Madison,* Technical Report No.361.

Field, Andy (2003*), Discovering Statistics using SPSS for Windows*. Sage Publications, London.

Fitz-Gibbon, C.T. (1985). Peer tutoring projects, social education improves achievement. *Journal of the National Organisation for Initiatives in Social Education*, *4*(2), 5-10.

Fitz-Gibbon, C.T. (1990). Success and failure in peer tutoring experiments. In Goodland, S., & Hirst, B., (Eds.). *Explorations in Peer Tutoring* (pp. 26-57), Oxford: Basil Blackwell.

Fitz-Gibbon, C.T. (1992). Empower and Monitor: The Em Algorithm for the creation of Effective schools. In Bashi, J., & Sass, Z. (Eds.). *School Effectiveness and Improvement* (pp 257-266.). Jerusalem: The Magnes Press.

Fitz-Gibbon, C.T. (2000a). Cross-age Tutoring: Practical Advice. In Walraven, G., Parsons C., Veen, D., & Day, C. (eds.) *Combating Social Exclusion through Education: Laissez-faire, Authoritarianism or Third Way?.* Leuven: Garant.

Fitz-Gibbon, C. (2000b). Cross-age Tutoring: Should it be required in order to reduce social exclusion? In Walraven, G., Parsons, C., Veen, D., & Day, C. (eds.) *Combating Social Exclusion through Education: Laissez-faire, Authoritarianism or Third Way?* (pp. 307-315). Leuven: Garant.

Fogarty, J.L., & Wang, M. (1982). An investigation of the cross-age peer tutoring process: Some implications for instructional design and motivation. *The Elementary School Journal, 82*(5), 450-469.

Foreman, E. (1989). The role of peer interaction in the social construction of mathematical knowledge. *International Journal of Educational Research. 3*(1), 55-70.

Fuchs, L. S., Fuchs, D., Bents, J., Phillips, N. B., & Hamlet, C. L. (1994). The nature of student interactions during peer tutoring with and without prior training and experience. *American Educational Research Journal*, 31, 75–103.

Fuchs, L.S., Fuchs, D., Hamlet, C.L. Phillips, N.B., Karns, K., & Dutka, S. (1997). Enhancing students' helping behaviour during peer-mediated instruction with conceptual mathematical explanations. *The Elementary School Journal, 97*(3), 223-249.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., & Karns, K. (1998). High-achieving students' interactions and performance on complex mathematical tasks as a function of homogeneous and heterogeneous pairings. *American Educational Research Journal, 35* (2), 227-267.

Fuchs, S.L., Fuchs, D., Karns, K., Hamlett, C.L., Dutka, S., & Katzaroff, M. (1996). The relation between student ability and the quality and effectiveness of explanations. *American Educational Research Journal, 33* (3), 631-664.

Fuchs, S.L., Fuchs, D., Karns, K., Hamlett, C.L., Katzaroff, M., & Dutka, S. (1998). Comparisons among individual and cooperative performance assessments and other measures of mathematics competence. *The Elementary School Journal, 99*(1), 23-51.

Fuchs, D., Fuchs, L.S., Mathes, P.G., & Simmons. D. (1997). Peer-assisted learning strategies: Makings classrooms more responsive to diversity. *American Educational Research Journal, 34* (1), 174-206.

Fuchs, S.L., Fuchs, D., Kazdan, S., & Allen, S. (1999). Effects of peer-assisted learning strategies in reading with and without training in elaborated help giving. *The Elementary School Journal*, *99* (3), 201-219.

Fuchs, D., Fuchs, L.S., Kazden, S., Karns, K. Calhoon, M.B., Hamlot, C.L. & Hewlett, S. (2000). Effects of workgroup structure and size on student productivity during collaborative work on complex tasks. *The Elementary School Journal, 100* (3), 183-212.

Gabbert, B., Johnson, D.W., & Johnson, R.T. (1986). Cooperative learning, group-to-individual transfer, process gain, and the acquisition of cognitive reasoning strategies. *Journal of Psychology*, 12, 265–278.

Gansle, K.A., & McMahon, C.M. (1997). Component integrity of teacher intervention management behaviour using a student self-monitoring treatment: An experimental analysis. *Journal of Behavioural Education*, 7, 405–419.

Gardner, W.E. (1973). The effects of intergrade tutoring with group guidance activities on the reading achievement, self-concept, attitudes toward school and behaviour of third and fourth grade tutors and on the reading achievement and behaviour of first and second grade tutees. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 73-31,723).

Garton, A.F., & Pratt, C. (2001). Peer assistance in children's problem solving. *British Journal of Developmental Psychology*. *19* (2), 307-318.

Garner, A., Kohler, M.C., & Reisman, F. (1971). *Students Teach Students: Learning by Teaching.* New York: Harper.

Garofalo, J., & Lester, F. (1985). Metacognition, Cognitive Monitoring, and Mathematical Performance. *Journal for Research in Mathematics Education*. Vol. 16, 163-176.

George, C.N. (1980). Differential effects of peer teaching and independent study on vocabulary learning of sixth-grade students. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 8015915).

Gilles, R. M. (2000). The maintenance of cooperative and helping behaviours in cooperative groups. *British Journal of Educational Psychology*, 70, 97–111.

Gill, B.K. (2001). Teacher initiated reform in schools: Cross-age tutoring as a support strategy to address reading achievement. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 3038235).

Ginsburg-Block, M. (1998). Reciprocal Peer Tutoring: Developing and testing effective peer collaborations for elementary school students. In K.J. Topping & S. Ehly (Eds.), *Peer Assisted Learning* (pp. 121-144). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ginsburg-Block, M., & Fantuzzo, J. (1997). Reciprocal peer tutoring: An analysis of

"Teacher" and "Student" interactions as a function of training and experience. *School Psychology Quarterly, 12*(2), 134-149.

Ginsburg-Block, M.D., & Fantuzzo, J.W. (1998). An evaluation of the relative effectiveness of NCTM standards-based interventions for low-achieving urban elementary students. *Journal of Education Psychology, 90*(3), 560-569.

Ginsburg-Block, M. (2005). Peer Tutoring. In S. W. Lee (Ed.), *Encyclopaedia of School Psychology*. Thousand Oaks, CA: Sage Publications, Inc.

Ginsburg-Block, M., Rohrbeck, C., Fantuzzo, J. W., & Lavigne, N. C. (2006). Peer assisted learning strategies. In Bear, G., & Minke, K., (Eds). *Children's Needs III: Understanding and Addressing the Developmental Needs of Children* (pp. 631-645). Bethesda, MD: National Association of School Psychologists.

Ginsburg-Block, M., Rohrbeck, B., Cynthia, A., Fantuzzo, J. W. (2006). A meta-Analytic review of social, self-concept, and behavioral outcomes of peer-assisted learning. *Journal of Educational Psychology*. 98(4), Nov 732-749.

Gmitter, J.W. (1989). Effects of microcomputer-assisted instruction and Classwide Peer Tutoring on computational skill achievement of third-grade students. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 9002866).

Gnadinger, C.M. (2008) Peer-mediated instruction: assisted performance in the primary classroom. *Teachers and Teaching: Theory and Practice, 14* (2), 129-142.

Goodland, S. & Hirst, B. (1989). *Peer tutoring: A guide to learning by teaching*. London: Kogan Page.

Gorard, S. (2014). The widespread abuse of statistics by researchers: what is the problem and what is the ethical way forward?. *Psychology of Education Review*, 38(1), 3-10.

Goyen, J. D., & McClelland, D. J. (1994). Pause, prompt and praise: The need for more research. *Journal of Research in Reading*, 17(2), 108–119.

Graesser, A., Person, N., & Magliano, J. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology*, 9, 495–522.

Greener, I. (2011). *Designing Social Research*. London: Sage

Greenwood, C.R., Dinwiddie, G., Bailey, V., Carta. J.J., Dorsey, D., Kohler, F.W., Nelson, C., Rotholz, D., & Schulte, D. (1987). Field replication of Classwide Peer Tutoring. *Journal of Applied Behaviour Analysis, 20* (2), 151-160.

Greenwood, C.R., Delquari, J.C., & Vance, H.R. (1989). Longitudinal effects of Classwide Peer Tutoring. *Journal of Educational Psychology, 81*(3), 371-383.

Greenwood, C.R., Dinwiddie, G., Terry, B., Wade, L., Stanley, S.O., Thibadeau, S., & Delquadri, J., (1984). Teacher versus peer-mediated instruction: An eco-behavioural analysis of achievement outcomes. *Journal of Applied Behaviour Analysis, 17*(4), 521-538.

Greenwood, C.R., Terry, B., Arreaga-Mayer, C., & Finney, R. (1992). The Classwide Peer Tutoring program: Implementation factors moderating students' achievement. *Journal of Applied Behaviour Analysis, 25* (1), 101-116.

Gregg, G.P. (1994). The use of storybook reading in cross-age tutoring program to enhance the reading skills of low ability high school students. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 9432620).

Gresham, F.M. (1989). Assessment of treatment integrity in school consultation and pre-referral intervention. *School Psychology Review*, 18, 37–50.

Gudbrandsen, B.H. (2005). The effect of tutoring and metacognition on 5th- and 6th-Grade students' reading strategies, reading comprehension, and attitude toward reading. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 3200943).

Hair, J. F., Anderson, R. E., Tatham, R, L., Black, W. C. (1998). *Multivariate Data Analysis*, Prentice-Hall International, London.

Hannah, D.C. (2008). Attitudinal Study: The interaction of student taking Calculus and Pre-Requisite courses while participating in peer tutorials. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 3323218).

Harris, V.W., & Sherman, J.A. (1973). Effects of peer tutoring and consequences on the math performance of elementary classroom students. *Journal of Applied Behaviour Analysis*, *6*(4), 587-597.

Harrison, GV. (1971). Structured Tutoring, Brigham Young University, Department of Instructional Research and Development, *Eric* no ED 053 080.

Harrison, G. V. (1976). Structured Tutoring: Antidote for low achievement. In Allen, V. L. (eds.) *Students as Teachers: Theory and Research on Tutoring*. London: Academic Press, INC.

Harrison, G. Von & Cohen, A. M. (1971). Empirical Validation of Tutor-Training Procedures. *Annual Meeting of the California Education Association*, San Diego, California, April 1971. Retrieved November, 10, 2012, from http://www.eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_&ERICExtSearch_SearchValue_0=ED058283&ERICExtSearch_SearchType_0=no&accno=ED058283.

Harrison, G.V. (1971). Structured Tutoring. Working Paper 24, ERIC, May 6, No ED 053 080. Retrieved May, 6th, 2013, from http://files.eric.ed.gov/fulltext/ED053080.pdf

Harrison, G.V., & Brimley, V. (1971). The use of Structured Tutoring Techniques in teaching low-achieving six-year-olds to read. *American Educational Research Association*, February, ERIC No ED 047 898.

Hartley, S.S. (1977). Meta-Analysis of effects of Individually Paced Instruction in

Mathematics. (Doctoral dissertation, University Of Colorado 1977). *Dissertation Abstract International*, 1977, 38 (7-A), 403. (University Microfilms No. 77-29,926)

Harvey, David. L. (1992). The new Science and the Old: Complexity and Realism in the Social Sciences. *Journal for the Theory of Social Behaviour, 22*(4), 356-379.

Harvey, David L. and Reed. Michael. (1996). Social Science as the study of complex sys tems. Kiel, L.D. and Elliott, E. (Eds.). *Chaos Theory in the Social Sciences.* Ann Arbor: University of Michigan.

Harvey. David. L. (2002). Agency and community: A critical realist paradigm. *Journal for the Theory of Social Behaviour*. *32*(2), 163-194.

Henington. C. & Skinner, C.H. (1998). Peer Monitoring. In Topping, K.J., & Ehly (Eds.). *Peer Assisted Learning* (pp.237-255). London UK: Lawrence Erlbaum.

Hilger, L.H. (2000). Cross-Age tutoring in reading: Academic and attitudinal effects for high-school tutors and third-grade tutees. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 9991418).

Higgins, S., Kokotsaki, D. & Coe, R.J. (2011). Toolkit of Strategies to Improve Learning: Summary for Schools Spending the Pupil Premium: Technical Appendices. Sutton Trust.

Higgins, S., Katsipataki, M., Coleman, R., Henderson, P., Major, L.E., & Coe, R. (2014). The Sutton Trust-Education Endowment. Sutton Trust.

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association, 81*, 945-960.

Howell, K.W. (1978). Using peers in drill-type instruction. *The Journal of Experimental Education, 46*(3), 52-56.

Hoy, A.W. & Tschannen-Moran, M. (1999). Implications of cognitive approaches to peer learning for teacher education. In O'Donnell, A., & King, A. (Eds), *Cognitive Perspectives on Peer Learning* (pp.257-285). Hillsdale, NJ: Erlbaum.

Inglis, A. (2002). Enhancing self-direction in low-achieving elementary school students: A cross-age tutoring study. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 0-612-69160-8).

Intrinsic Motivation Inventory. (2012). Intrinsic Motivation Scale in Self Determination Theory Website, visited on the 12-11-2012 at: http://selfdeterminationtheory.org/questionnaires/10-questionnaires/50#getScale.

Ismail, H.N. (1999). Learning within scripted and non-scripted peer tutoring sessions: The Malaysian context. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database.* (UMI No: 9950775).

Jennings, C. (2004). The reading together cross-age tutoring program and its effects on the English language proficiency and reading achievement of English language learners. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 3157485).

Joffin, B. (1988). A qualitative study of the mutual benefits for tutors and tutees in a cross-age tutoring programme in reading. (Master's Thesis). Available from *ProQuest Dissertation and Thesis Database*. (ISBN: 0-315-45897-6).

Johnson, D. W. (1990). *Reaching Out: Interpersonal effectiveness and self-actualisation.* Prentice Hall: New Jersey.

Johnson, D.W., Johnson, R.T., Holubec, E.J., & Roy. P. (1984). Circle of Learning: Cooperation in the classroom. Alexandria, VA: Associations for supervision and Curriculum Development.

Johnson, D.W., & Johnson, R.T. (1987). *Learning together & alone: Cooperative, competitive & individualistic learning.* New Jersey: Prentice-Hall, Inc.

Johnson, D.W., & Johnson R, T. (1989). *Cooperation and Competition: Theory and Research*. Edina, MN: Interaction Book Company.

Johnson, D. W., Johnson, R. T., & Holubec, E. J. (1990). *Circles of learning*. (3rd. Ed.). Edina, MN: Interaction Book Company.

Johnson, D.W., & Johnson R.T. (2005). New developments in Social Interdependence Theory. *Social Genetic Psychological Monograph*. 131, 285-358.

Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). The effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin. 89,* 47–62.

Jolliffe, W. (2005). The effective implementation of Cooperative Learning in the Primary Classroom. *British Education Research Association Annual Conference,* University of Glamorgan, 14-17, September.

Jolliffe, W. (2007). *Cooperative Learning in the Classroom: Putting it into Practice*. London; Paul Chapman.

Jolliffe, W. (2011). Cooperative Learning: making it work in the classroom. *Journal of Cooperative Studies*, 44 (3), 31-42.

Jolliffe, W. (2015). Bridging the gap: teachers cooperating together to implement cooperative learning. *Education 3-13: International Journal of Primary, Elementary and Early Years Education*, 43 (1), 56-69.

Jolliffe, W., & Hutchinson, Hazel. (2007). Implementing Cooperative Learning in a Networked Learning Community. *Education 3-13: International Journal of Primary, Elementary and Early Years Education*. 34 (1), 37-48.

Jones, C.P. (1981). A descriptive study of cross-age peer tutoring as a strategy for improvement in two selected middle schools of the Highland Park, Michigan school system from 1970 to 1978. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 8117071).

Jones, D.L. (2004). The effects of a cross-age reading tutorial program: Hispanic students who are not proficient in reading tutored by non-Hispanic students who are proficient in reading. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 3154193).

Joscelyne, T. (1989). Changes in reading style following peer tutored paired reading and listening to reading. In *Paired Learning: Tutoring by Non-Teachers: Incorporating the Paired Reading Bulleting*, *ERIC Document Reproduction Services, 5.*

Kamps, D.M., Greenwood , C., Arreaga-Mayer, C., Verkamp, M.B., Utley, C., Tapia, Y., Bowman-Perrot, L., Bannister, H. (2008).The efficacy of Classwide Peer Tutoring in middle schools. *Education and Treatment of Children, 31*(2), 119-152.

Kang, H. (1999). Assessing literacy growth in a cross-age elementary grade peer tutoring program. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 9941737).

Keer, H.V. (2004). Fostering reading comprehension in fifth grade by explicit instruction in reading strategies and peer tutoring. *British Journal of Education Psychology*, 74 (1), 37-70.

Kenedy, L.J. (2000). The effect of cross-age tutoring routines on the tutor's perception of self as a reader. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 9966643).

Kiarie, M.W. (2003). The differential effects of peer tutoring and peer tutoring with group contingency on the spelling performance and disruptive behaviour of fourth-grade students in a general education classroom. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 3093664).

King, A. (1993). Effects of guided cooperative questioning on children's knowledge construction. *Journal of Experimental Education*, 61(2), 127-148.

King, A. (1994). Building knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal, 32*(2), 338-368.

King, A., Staffieri, A., & Adelgais, A. (1998). Mutual peer tutoring: Effects of structuring tutorial interaction to scaffold peer learning. *Journal of Educational Psychology, 90*(1), 134-152.

Klosterman, R. (1970). The effectiveness of a diagnostically structured reading programme. *The Reading Teacher.* 24. 159-62.

Koffka, K. (1935). *Principles of Gestalt Theory*. New York: Harcourt. Brace.

Kohler, F.W. & Greenwood, C.H. (1990). Effects of collateral peer supportive behaviours within the Classwide Peer Tutoring program. *Journal of Applied Behaviour Analysis*, *23* (3), 307-322.

Kramarski, B., & Mevarech, Z.R. (2003). Enhancing Mathematical Reasoning in the Classroom: The Effects of Cooperative Learning and Metacognitive Training. American Educational Research Journal, 40(1), 281-310.

Kramarski, B. & Mevarech, Z. R. (2004). Ways for new models in mathematics education: Effects of meta-cognitive training and cooperative learning on mathematics communication. In J. Wang and B. Xu (Eds.). Trends and challenges in mathematics education. (pp. 15-25.) Shanghai: East China Normal University Press.

Lachin, J.M., Matts, J.P., & Wei, L.J. (1988). Randomization in Clinical Trials:

Conclusions and Recommendations. *Control Clinical Trials. 9*, 365–74.

LeBlanc, G. & Bearison, D. (2004). Teaching and learning as a socially constructed activity. *Cognitive Development*, 19, 499-515

Lemons, C.J., Fuchs, D., Gilbert, J., Fuchs, L.S (2014). Evidence-Based Practice in a Changing World: Reconsidering the Counterfactual in Education Research. *Educational Researcher*. 43(5), 242-252.

Lepper, M. R., & Greene, D. (1975). Turning Play into Work: Effects of adult surveillance and extrinsic rewards on students' Intrinsic Motivation. *Journal of Personality and Social Psychology*, 31, 479-486.

Lepper, M. R., & Henderlong, J. (2000). Turning "Play" into "Work" and "Work" into "Play": 25 Years of Research on Intrinsic versus Extrinsic Motivation. In Sansone, C. & Harackiewicz, J. (Eds.), *Intrinsic motivation: Controversies and new directions* (pp. 129-137). San Diego: Academic Press.

Leung, K. C., Marsh, H. W., Craven, R. G., & Yeung, A. S. (2005). *Are Peer Tutoring Programs Effective in Promoting Academic Achievement and Self-Concept in Educational Settings: A Meta-Analytical Review.* Presented in symposium in 2005 international conference of the Australian Association for Research in Education. In http://publications.aare.edu.au/05pap/leu05421.pdf.

Leung, K. C., Marsh, H. W., Craven, R. G., Yeung, A. S., & Abduljabbar, A. S. (2013). Domain specificity between peer support and self-concept. *Journal of Early Adolescence*, *33*(2), 227-244.

Leung, K, C. (2014). Preliminary Empirical Model of Crucial Determinant of Best Practice for Peer Tutoring on Academic Achievement. *Journal of Educational Psychology*. 107(2), 558-579.

Lewin, K, (1948). *Resolving Social Conflict.* New York, Harper.

Li, L. (1999). The effects of varying amounts of practice during Classwide Peer Tutoring on spelling performance of third graders. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 9951376).

Limbrick, E., McNaughton & Glynn, T. (1985). Reading gains for underachieving tutors and tutees in a cross-age tutoring programme. *Child Psychology and Psychiatry, 26* (6), 939-953.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. Beverly Hills, CA: Sage Publications.

Lippitt, P. (1976). Learning through cross-age helping: Why and how. In Allen, V. L. (Ed.) *Students as Teachers: Theory and Research on Tutoring* (pp.157-168). London: Academic Press, INC.

Lovett, M. (2008). Teaching Meta-Cognition. *Education Learning Initiative 2008 Annual Meeting.* Retrieved August, 2nd, 2013, from: http://net.educause.edu/upload/presentations/eli081/fs03/metacognition-eli.pdf.

Madrid, L.D., Canes, M., & Ortega-Medina, M. (2007). Effects of team competition versus team cooperation in Classwide Peer Tutoring. *The Journal of Educational Research, 100*(3), 155-160, DOI: 10.3200/JOER.100.3.155-160.

Maheady, L. (1998). Advantages and disadvantages of peer-assisted learning strategies. In Topping, K.J., & Ehly (Eds.). *Peer Assisted Learning* (pp.45-67) London UK: Lawrence Erlbaum.

Marsh, H. W. (1990). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology, 82*, 623-636.

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, 1, 133-163

Marsh, H. W., & O'Mara, A. (2008). Reciprocal effects between Academic Self-concept, Self-esteem, Achievement and Attainment over seven adolescent years: Unidimensional and Multidimensional perspectives of Self-concept. *Personality and Social Psychology Bulletin*, 34, 542-552.

Marsh, H. W., & McDonald-Holmes, I. W. (1990). Multidimensional self-concepts: Construct validation of responses by students. *American Education Research Journal*, 27, 89-117.

Massey, A., Green, S., Dexter, T. & Hammet, L. (2003). Comparability of national tests scores over time: KS1, KS2, and KS3 standards between 1996 and 2011. Final Report to QCA of the Comparability Over Time Project. (Research and Evaluation division of the University of Cambridge Local Examination Syndicate).

Mathes, P.G., & Babyak, A.E. (2001). The effects of peer-assisted literacy strategies for first-grade readers with and without additional mini-skills lessons. *Learning Disabilities Research & Practice, 16*(1), 28-44.

Mathes, G.P., Howards, J.K., Allen, S.H., & Fuchs, D. (1998). Peer-assisted learning strategies for first-grade readers: Responding to the needs of diverse learners. *Reading Research Quarterly, 33*(1), 62-94.

Mathie, E., & Ford, N. (1998). Peer education for health. In Topping, K.J., & Ehly (Eds.). *Peer Assisted Learning* (pp.203-219). London UK: Lawrence Erlbaum.

McCartney. R., Warner. J., Iliffe. S., Haselen. R. Griffin. M. & Fisher, P. (2007). The Hawthorne Effect: A Randomised, Controlled Trial, BMC. *Medical Research Methodology*, *7* (30), 1-8.

McKinstery, J. & Topping, K.J. (2003). Cross-age peer tutoring of thinking skills in the high school. *Educational Psychology in Practice: theory, research and practice in educational psychology*, *19* (3), 199-217, DOI: 10.1080/0266736032000109465.

Medway, F.J. & Baron, R.M. (1977). Locus of control and tutor's instructional style as determinants of cross-age tutoring effectiveness. *Contemporary Educational Psychology, 2* (3), 298-310.

Medway, F.J. & Lowe. C.A. (1980). Causal attribution for performance by cross-age tutors and tutees. *American Educational Research Journal*, *17*(3), 377-387, DOI: 10.3102/00028312017003377.

Menses, K.F. & Gresham, F.M. (2009). Relative efficacy of reciprocal and nonreciprocal peer tutoring: For students at-risk for academic failure. *School Psychology Quarterly, 24* (4), 266-275.

Merrett, F., & Thorpe, S. (1996). How important is the praise element in the pause, prompt and praise tutoring procedures for older low-progress readers?, *Educational Psychology: An International Journal of Experimental Educational Psychology*, *16*(2), 193-206, DOI: 10.1080/0144341960160208.

Merrill, M.E. (2002). The complicated effects of cross-age tutoring biology students. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No:3044881).

Mevarech, Z., & Kramarski, B. (1997). IMPROVE: A multidimensional method for teaching mathematics in heterogeneous classrooms. American Educational Research Journal, 34, 365-394.

Miller, K. (1989). Cross-Age Tutoring in Science. In *Paired Learning: Tutoring by Non-Teachers: Incorporating the Paired Reading Bulleting*, *ERIC Document Reproduction Services, 5.*

Miller, D., Topping, K. J., & Thurston, A. (2010). Peer Tutoring in reading: The effects of role and organization on two dimensions of self-esteem. *British Journal of Educational Psychology*, 80, 417-433.

Moher, D., Hopewell, S., Schulz K.F., Montori, V., Gøtzsche, P,C., Devereaux, P.J., Elbourne, D., Egger, M., & Altman, D.G. (2010). Explanation and Elaboration: updated guidelines for reporting parallel group randomised trial. *BMJ*.340:c869.

Murcia, J., Roman, M., Galindo, C., Alonso, N., & Gonzalez-Cutre, D. (2008). Peers' influence on exercise enjoyment: A self-determination theory approach. *Journal of Sports Science and Medicine*, *7*, 23–31.

Nardiello, S.L. (2009). Classwide Peer Tutoring. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 1463674).

Nath, L.R. & Ross, S.M. (2001). The influence of a peer-tutoring training model for implementing cooperative groupings with elementary students. *Educational Technology Research and Development*, *49*(2), 41-56.

Nattiv, A. (1994). Helping behaviour and math achievement gain of students using cooperative learning. *Elementary School Journal*, 94, 285-297.

Niedermeyer, F. C. (1970). Effects of Training on the instructional behaviours of student tutors, *The Journal of Educational Research*, 64, 119-123.

Niedermeyer. F. C., & Ellis, P. (1971). Remedial Reading Instruction by Trained Pupil Tutors. *The Elementary School Journal*, 71, 400-405.

Nixon, J. G., & Topping, K. J. (2001). Emergent writing: The impact of structured peer interaction. *Educational Psychology: An International Journal of Experimental Educational Psychology, 21*(1), 41-58.

Noell, G.H., Witt, J.C., Slider, N.J., Connell, J.E., Gatti, S.L., Williams, K.L., Resetar, J.L. (2005). Treatment implantation following behavioural consultation in schools: A comparison of three follow-up strategies. *School Psychology Review*, 34, 87–106.

Novak, C.G. (1980). Differential effects of peer teaching and independent study on vocabulary learning of sixth-grade students. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 8015915).

Novotni, S.M. (1985). Peer tutoring: A study of its effect on mathematics achievement and attitude of ninth grade Math I students of Harrisburg high school. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 8525665).

Othman, N. (1996). The effects of cooperative learning and traditional mathematics instruction in grades K12: A meta-analysis of findings. Available from *ProQuest Dissertation and Thesis Database.* (UMI No: 9716375).

O'Donnel, A.M. (1998). Structuring dyadic interaction through scripted cooperation. The implications of Piagetian theory for peer learning. In O'Donnell, A., & King, A. (Eds), *Cognitive perspectives on peer learning*. (pp. 179-197). Hillsdale, NJ: Erlbaum.

O'Donnell, A.M. & Topping, K. (1998). Peer assessing peers: Possibilities and problems. In Topping, K.J., & Ehly (eds.) *Peer Assisted Learning* (pp.255-279). London UK: Lawrence Erlbaum.

Olson, P.C. (2009). Weaker readers as experts: The effects of preferential instruction on the fluency improvement of lower-performing student tutors. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 3364085).

Paquettte, K.R. (2003). Cross-age tutoring writing program: Investigation of academic achievements and attitudes among elementary students. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 3071345).

Parham, J.W. (1993). An analysis of the effects of tutoring on seventh grade students engaged in the mastery of pre-algebra concepts. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 9407677).

Perdomo-Rivera, C. (2002). The effects of Classwide Peer Tutoring on the literacy achievement and language production of English language learners in an elementary school setting. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 3067091).

Person, N.K., & Graesser, A.G. (1999). Evolution of discourse during cross-age tutoring. In O'Donnell, A.M. & King, A. (Eds.), *Cognitive perspectives on peer learning* (pp. 69–86). Mahwah, NJ: Lawrence Erlbaum

Pheleps, E. & Damon, W. (1989). Problem solving with equals: Peer collaboration as a context for learning mathematics and spatial concepts. *Journal of Educational Psychology, 81*(4), 639-646.

Piaget, J. (1978). *The development of thought: Equilibration of cognitive structures*. Oxford, UK: Basil Blackwell.

PISA (2009). PISA 2009 ranking. *OECD, PISA 2009 Database,* retrieved May, 6th, 2012 from: http://www.oecd.org/pisa/46643496.pdf.

PISA. (2012). PISA 2012 results in focus: What 15-year-olds know and what they can do with what they know. *OECD, PISA 2012 Database*, retrieved March, 15th, 2013 from: http://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf.

Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., & Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology*. 95 (1), 240–257.

Roscoe, R.D., and M.T.H. Chi. (2007). Understanding Tutor Learning: Knowledge-Building and Knowledge-Telling in Peer Tutors' Explanations and Questions, *Review of Educational Research*. 77 (4) 534-574.

Roseth, C. J., Johnson, D.W., & Johnson, R.T. (2008). Promoting early adolescents, achievement and peer relationships: The effects of cooperative, competitive, and individualistic goal structures. *Psychology Bulletin*, 134 (2), 223-246

Rosen, S., Powell, E.R., Schubot, D.B., & Rollins, P. (1978). Competence and tutorial role as status variables affecting peer-tutoring outcomes in public school settings. *Journal of Educational Psychology*, *7*0(4), 602-618.

Ryan, R.M., & Deci, E.L. (2000). Self-Determination Theory and the facilitation of intrinsic motivation, social development and well-being. *American Psychologist, 55,* 68-78.

Sadler, B., & Graham, S. (2005). The effects of peer-assisted sentence-combining instruction on the writing performance of more and less skilled young writers. *Journal of Educational Psychology, 97*(1), 43-54, DOI: 10.1037/0022-0663.97.1.43

Sanderson, P., Carrington, B., & Fitz-Gibbon, C.T. (1992). Cross-age peer tutoring in science. *School Science Review, 74*(266), 49-56.

Sarbin. T. R., & Allen V. L. (1968). Role Theory. In Lindzey, G., & Arnseon, E. (Eds.). T*he Handbook of Social Psychology* (pp. 488-567). Massachusetts: Addison-Wesley.

Sarbin, T. R. (1976). Cross-Age Tutoring and Social Identity. In Allen, V. L. (Ed.) *Students as Teachers: Theory and Research on Tutoring* (pp. 27-40). Academic Press, London: INC.

Schunk, D.H. (1998). Peer Modelling. In Topping, K.J., & Ehly (Eds.). *Peer Assisted Learning* (pp.185-203). London UK: Lawrence Erlbaum.

Sharpley, A.M., & C.F. Sharpley. (1981). Peer tutoring: A review of the literature. *Collected Original Resources in Education,* 5(3), C-11.

Sharpley, A.M., Irvine, J., Sharpley, C.F. (1983). An examination of the effectiveness of a cross-age tutoring program in mathematics for elementary school children. *American Educational Research Journal, 2*(1), 103-111.

Siemens, B.K. (2001). The effects of cross-age tutoring on the cognitive achievement of tutors during activity-based science lessons. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 3013123).

Siler, R.S. (1994). Peer tutoring: Assessing the generalized effects of reading improvements made by second grade students. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 9527901).

Skinner, S.H. (1989). A comparison of a peer tutoring and a cover, copy, and compare intervention on multiplication fact acquisition and fluency. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 9008093).

Slavin, R. E. (2008a). What works? Issues in synthesizing educational program evaluations. *Educational Researcher*, 37, 5-14.

Slavin, R.E. (2008b). Evidence-based Reform in Education: Which Evidence Counts? *Educational Researcher,* 37, 47-50.

Slavin, R.E., Lake, C., & Groof, C. (2009). Effective Programs in Middle and High school Mathematics: A Best-Evidence Synthesis. *Review of Educational Research.* 79 (2), 939-911.

Slavin, R. E. and Madden. Nancy A. (2008). Understanding Bias Due to Measures Inherent to Treatments in Systematic Reviews in Education. Retrieved July, 12th, 2012, From; http://www.bestevidence.org/methods/understand_bias_Mar_2008.pdf

Slavin, R. (2010). Cooperative learning: What makes group-work Work, in Dumont, H., D. Istance and F. Benavides (Eds.) in *The Nature of Learning: Using Research to Inspire Practice*. Educational Research and Innovation, OECD Publishing.

Slope, M.A. (2000). Buddy reading program-cross-age tutoring in the elementary school. (Master's Thesis). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 1401307).

Smith, C.L. (2010). Examining the effectiveness of peer-tutoring and computer-aided instruction for mastery of multiplication facts. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 3405891).

Spencer, T.D. (2001). Sociometric change as a function of Classwide Peer Tutoring. (Master's Thesis). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 1407260).

Sporer, N., & Brunstein, J.C. (2009). Fostering the reading comprehension of secondary school students through peer-assisted learning: Effects on strategy knowledge, strategy use, and task performance. *Contemporary Educational Journal, 34* (4) 289-297.

Soper, D. (2013). Statistics Calculators. Retrieved, 15th May, 2013, from http://www.danielsoper.com/statcalc3/calc.aspx?id=47.

Stephens, H.L. (2002). Cross-age tutoring program coordinated with classroom instruction: Effects on elementary students' reading and writing achievement. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 3110311).

Strauss, A., & Corbin, J. (1990). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Thousand Oaks: Saga Publications.

Sutherland, J.A. & Topping, K. (1999). Collaborative creative writing in eight-year-olds: Comparing cross-ability fixed Role and same-ability reciprocal role pairing. *Journal Of Research In Reading, 22*(2), 154-179, ISSN 0141-0423.

Swing, S. R., & Peterson, P . L., (1982). The relationship of student ability and small group interaction to student achievement. *American Educational Research Journal*, 19, 259-274.

Taylor, L.A.k. (2002). The effects of Classwide Peer Tutoring on spelling achievement, reading fluency, and reading comprehension. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 3071088).

Taylor, B.M., Hanson, B.E., Justice-Swanson, K., & Watts, S.M. (1997). Helping struggling readers: Linking small-group intervention with cross-age tutoring. *The Reading Teacher*, *51*(3), 196-209.

Tierney, P. (2005). A critical investigation of cross-age peer tutoring program: Silver bullet, social threat, or sound policy. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 3175773).

Topping, K. (1987). Peer tutored paired reading: Outcome data from ten projects. *Educational Psychology: An International Journal of Experimental Educational Psychology*, *7*(2), 133-145, DOI: 10.1080/0144341870070206.

Topping, K. (1998). Paired learning in literacy. In Topping, K., & Ehly (Eds.). *Peer Assisted Learning* (pp.87-105). London UK: Lawrence Erlbaum.

Topping, K. J. (1995). *Paired reading, spelling and writing: The handbook for teachers and parents*. London: Cassell.

Topping, K. J. (2001a). *Thinking, reading, writing: A practical guide to paired learning with peers, parents and volunteers*. London: Continuum International.

Topping, K. J. (2001b). *Peer assisted learning: A practical guide for teachers*. Cambridge: Brookline Books.

Topping, K. J. (2010). Methodological quandaries in studying process and outcomes in peer assessment: Commentary. *Learning and Instruction*, 20, 339-343.

Topping, K., & Bryce, A. (2004). Cross-age peer tutoring of reading and thinking: Influence on thinking skills. *Educational Psychology: An International Journal of Experimental Educational Psychology*, *24* (5), 595-621.

Topping, K.J., & Ehly, S. (1998). Introduction to peer learning. In Topping, K.J., & Ehly

(Eds.). *Peer Assisted Learning*. London UK: Lawrence Erlbaum.

Topping, K. J., & Ehly, S. W. (2001). Peer assisted learning: A framework for consultation. *Journal of Educational and Psychological Consultation*, 12, 113– 132.

Topping, K. J., Campbell, J., Douglas, W. & Smith, A. J. (2003). Cross-age peer tutoring in mathematics with 7 & 11 year olds: Influence on mathematical vocabulary, strategic dialogue and self-concept. *Educational Research*, 45, 287-308.

Topping, K., Campbell, J., Douglas W & Smith, A. (2003). Cross-age peer tutoring in mathematics with seven- and 11-year-olds: influence on mathematical vocabulary, strategic dialogue and self-concept, *Educational Research*, *45*(3), 287-308.

Topping, K., & Ehly, S. (1998). Introduction to peer learning. In Topping, K.J., & Ehly (Eds.). *Peer Assisted Learning* (pp.1-25). London UK: Lawrence Erlbaum.

Topping, K. J., & Ehly, S. W. (2001). Peer assisted learning: A framework for consultation. *Journal of Educational and Psychological Consultation*. 12 (1), 113– 132.

Topping, K., Nixon, J., Sutherland, J., & Yarrow, F. (2000). Paired Writing: A framework for effective collaboration. *Reading*, *34*(2), 79-89.

Topping, K.J., Peter, C., Stephen, P., & Whale, M. (2004). Cross-age peer tutoring of science in the primary school: influence on scientific language and thinking. *Educational Psychology: An International Journal of Experimental Educational Psychology*, 24(1), 57-75, DOI: 10.1080/0144341032000146449

Topping, K. J., Miller, D., Murray, P., & Conlin, N. (2011). Implementation process in peer tutoring of mathematics. *Educational Psychology*, 31, 575-593.

Topping, K. J., Thurston, A., Tolmie, A., Christie, D., Murray, P., & Karagiannidou, E. (2011). Cooperative learning in science: Intervention in the secondary school. *Research in Science and Technological Education*, 29, 91-106.

Torgerson, C. J. & Torgerson, D. J. (2001). The need for randomised controlled trials in educational research. *British Journal of Educational Studies*, *49*, 316-329.

Thurston, A. (2014). Manual to help develop peer tutoring in the school classroom. Retrieved January 15th , 2015 from: http://pure.qub.ac.uk/portal/files/12775117/LT_Mathematics_PackVQUB.pdf.

Thurston, A., Burns, V., & Topping, K. (2011). Using peer tutor mediated metacognitive strategy to enhance mathematics attainment in elementary schools. ESRC Working paper. Retrieved October, 15th, 2013 from http://www.esrc.ac.uk/my-esrc/grants/RES-000-22-3201-A/outputs/read/242a6ac7-6128-4803-9eb2-eca3858ef692

Thurston, A., Van de Keere, K., Topping, K. J., Kosack, W., Gatt, S., Marchal, J., Mestdagh, N., Schmeinck, D., Sidor, W., & Donnert, K. (2007). Peer learning in primary school science: Theoretical perspectives and implications for classroom practice. *Electronic Journal of Research in Educational Psychology*, 5, 477-496

Trochim, W. (2012), Design, the research method knowledge base. Retrieved January, 14th, 2012 from: http://www.socialresearchmethods.net/kb/design.php.

Tymms, P. (1989). Peer Tutoring with A Level Chemistry students. In *Paired Learning: Tutoring by Non-Teachers: Incorporating the Paired Reading Bulleting, ERIC Document Reproduction Services, 5.*

Tymms, P. (2011). Evidence? The impact of large-scale reform in England. Zeitschrift für Erziehungswissenschaft, 13, 105-115

Tymms, P. (2004). Are standards rising in English primary schools? *British Educational Research Journal*. 30, 477-494.

Tymms, P., Coe, R. and Merrell, C., (2005). Standards in English Schools Changes since 1997 and the impact of government policies and initiatives: a report for the Sunday Times. Durham CEM Centre, Durham University.

Tymms, P., & Fitz-Gibbon C. T. (2002). Theories, hypotheses, hunches and ignorance. *Building Research Capacity: Journal of the ESRC Teaching and Learning Research Program Research Capacity Building Network,* April (2), 10-11.

Tymms, P. B. & Merrell, C. (2015). Cross-age Peer Learning. *Better Evidence-based Education,* 7(1), 18-19.

Tymms, P., Merrell, C., Thurston, A., Andor, J., Topping, K. & Miller, D. (2011). Improving attainment across a whole district: School reform through peer tutoring, a randomized controlled trial. *School Effectiveness and School Improvement, 22*, 265-289.

Valentine, J.C., DuBois, D.L., & Cooper, H. (2004). The relation between Self-beliefs and Academic Achievement: A Meta-analytic review. *Educational Psychologist*, 39, 111-133.

Veerkamp, M.B. (2001). The effects of Classwide Peer Tutoring on the reading achievement of urban middle school students. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 3049533).

Vygotsky, L.S. (1968). *Though and Language*. Cambridge: MA, MIT Press.

Vygotsky, L.S. (1986). *Mind in society: The development of higher psychological processes.* Cambridge: MA: Harvard University Press.

Watt, J. M., & Topping, K. J. (1993). Cued Spelling: A comparative study of parent and peer tutoring. *Educational Psychology in Practice: theory, research and practice in educational psychology, 9*(2), 95-103.

Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research,* 13, 21-40.

Webb, N. M., Troper, J., & Fall, J. R. (1992). Effective helping behaviour in cooperatives mall groups. Paper presented at the Annual Meeting of the American Educational Research Association, April, San Francisco.

Webb, N. M. (1991). Task-related verbal interaction and mathematics learning in small groups, *Journal for Research in Mathematics Education*. 22 (5), 366–89.

Weinstein, C. E., Husman, J., & Dierking, D. R. (2000). Self-regulation interventions with a focus on learning strategies. In Boekaerts, P.R. Pintrich & M. Zeidner (Eds.), *The Handbook of Self-Regulation* (pp. 727-747.). London: M. Academic Press.

White, P.M. (2000). Promoting mathematics achievement, academic efficacy, and cognitive development of at-risk adolescents through deliberate psychological education. (Doctoral Dissertation). Available from *ProQuest Dissertation and Thesis Database*. (UMI No: 9965210).

Willis, P,, Bland, R, Manka, L., & Craft, Cec. (2012). The ABC of peer mentoring – What secondary students have to say about cross-age peer mentoring in a regional Australian school. *Educational Research and Evaluation: An International Journal on Theory and Practice, 18*(2), 173-185.

Winter, S. (1988). Paired reading: A study of process and outcome. *Educational Psychology: An International Journal of Experimental Educational Psychology, 8*(3), 135-151.

Winter, S (1996). Peer Tutored Instruction in Reading. *Mentoring & Tutoring: Partnership in Learning, 3*(3), 3-11.

Wright, J., & Cleary. K.S. (2006). Kids in the tutor seat: Building school's capacity to help struggling readers through a cross-age peer-tutoring program. *Psychology In The Schools, 43*, 99-107.

Yackel, E., Cobb, P., & Wood, T. (1991). Small group interactions as a source of learning opportunities in second-grade mathematics. *Journal for Research in Mathematics Education*, 22, 390–408.

Yarrow, F., & Topping, K. J. (2001). Collaborative writing: The effects of metacognitive prompting and structured peer interaction. *British Journal of Educational Psychology*, 71, 261-282.

Appendix:

Empirical Literature Review Tables

(Number “1” indicates presence of theoretical element)

Table 36. *Quantitative and qualitative peer tutoring studies in literacy by socio interdependent elements*

	Social Interdependence, Main Elements										
Studies	***Rew-ard***	***Goal Per-forma-nce***	***Fixed Roles***	***Interpersonal Communica-tion Skills Training***	***Script Social Inter-dependence***	***Praise***	***Cross Ability***	***Pedagogical Skills Training***	***Elaborative/ Deep - Cognition***	***Meta Cognition***	***Academic Script***
Allen, V.L. & Feldman, C. (1973).			1				1	1			1
Al-Hasan. S. (2003).	1	1			1	1		1			1
Atherley, C. A. (1989).			1			1	1	1			
Baker, H.R. (2005).	1	1					1	1			
Breece, M. (2012).	1	1				1	1	1	1	1	
Bruno, M.A. (2004).			1				1	1			1
Burns, .M.K. (2004).			1				1				
Burns, E. (2006).			1			1	1	1			
Cameron, F.L.J. (1982).			1	1		1	1	1			1
Carberry. D.J. (2003).			1			1	1	1		1	1
			1	1		1	1	1			1

Coats, L.B. (2007).			1			1	1	1			
Conrad, E.E. (1975).	1						1	1			
	1						1	1			
Cooper, K.J. (1973).			1				1				
			1	1			1	1			
			1				1	1			
			1	1			1	1			
Dancer, A.P. (2007).	1	1		1				1	1	1	1
Daniels, L.J. (2004).			1				1	1			
Dineen, J.P., & et al. (1997).	1	1			1	1		1			
Dufrene, B.A., et al. (2010)	1		1				1	1			
Dion, E., & et al. (2011)	1	1				1	1	1	1	1	1
Duran, D., & Monereo,C. (2005).			1				1	1			1
							1				1
Eason, K.E. (2000).			1				1	1			1
Ekstrand, C. (2011).			1		1						1
Felman. R.S., & et al. (1976).	1		1				1				
Fuchs, S.L., & et al. (1997).	1	1					1	1	1	1	1
Fuchs, S.L., & et al. (1999).	1	1					1	1	1	1	1
	1	1					1	1	1	1	1
Gardner, W.E. (1973).		1					1	1			1
		1					1	1			1
Gill, B.K. (2001).			1	1		1	1	1			
Greenwood, C.R., & et al. (1984).	1	1			1			1			
Greenwood, D.R. & et al. (1989).	1	1			1			1			
Greenwood, C.R.,& et al. (1992)	1	1		1		1		1			
Greenwood, C.R., & et al.	1	1		1		1		1			

(1987).											
Gregg, G.P. (1994).			1				1	1			1
			1				1	1		1	1
Gnadinger, C.M. (2008).							1				1
Gudbrandsen, B.H. (2005)			1				1	1			
Hilger, L.H. (2000).			1	1		1	1	1			
Jennings, C. (2004).			1				1	1	1		1
Joffin, B. (1988).			1				1	1			
Jones, C.P. (1981).			1				1				
Josceline, T. (1989).			1				1	1			1
Kamps, D.M., & et al. (2008).	1	1		1		1		1	1		1
	1	1		1		1		1	1		1
Kang, H. (1999).			1				1				
Keer, H.V. (2004).			1	1			1	1	1	1	1
				1			1	1	1	1	1
Kiarie, M.W. (2003).	1	1					1		1		
						1		1			1
						1		1			1
Kohler, F.W. & Greenwood, C.H (1990).	1	1			1	1		1			
	1	1			1			1			
Li, L. (1999).	1	1			1	1		1			1
	1	1			1	1		1			1
	1	1			1	1		1			1
Limbrick, E., & et al. (1985).			1				1				
			1				1				
			1				1	1	1	1	
Madrid, L.D., & et al. (2007)		1	1					1			

		1	1					1			
Mathes, P.G., & Babyak, A.E. (2001)	1	1	1			1	1	1	1	1	
	1	1	1			1	1	1	1	1	
Mathes, P.G., & et al. (1998).	1	1				1	1	1	1	1	1
McKinstery, J. & Topping, K.J. (2003).			1			1	1	1	1	1	1
			1			1	1	1			1
Medway, F.J. & Baron, R.M. (1977).			1				1				
			1				1				
Medway, F.J. & Lowe. C.A.(1980).			1				1				
Merrett, F., & Thorpe, S. (1996).			1			1	1				1
			1				1				1
Nardiello, S.L. (2009).	1	1			1		1	1			
Nath, L.R. & Ross, S.M. (2001)				1		1	1	1			
							1				
Niedermeyer, F.C. (1970)			1			1	1	1			1
											1
Nixon, J. G., & Topping, K. J. (2001)			1		1	1	1	1	1	1	1
Novak, C.G. (1980).	1	1	1					1			1
	1	1	1					1			1
	1	1	1					1			
Olson, P.C. (2009).			1				1	1			
			1				1	1			
Paquettte, K.R. (2003).			1				1				
Perdomo-Rivera, C. (2002).	1	1			1		1	1		1	1

	1	1			1		1	1		1	1
	1	1			1		1	1		1	1
Sadler, B., & Graham, S. (2005).						1	1	1	1	1	1
						1	1	1	1	1	1
Siler, R.S. (1994).	1	1	1		1	1	1	1			
Slope, M.A. (2000)			1	1		1	1	1	1		
Sporer, N., & Brunstein, J.C. (2009).	1	1					1	1	1	1	1
Spencer, T.D. (2001).	1	1			1	1		1			1
Stephens, H.L. (2002).			1	1			1	1			1
Sutherland, J.A. & Topping, K. (1999).					1			1	1	1	1
			1		1		1	1	1	1	1
Taylor, L.A.k. (2002).	1	1			1			1	1		1
Taylor, B.M., & et al. (1997).			1				1		1	1	1
Tierney, P. (2005)			1	1		1	1	1		1	1
Topping, K. (1987).			1		1	1	1	1			1
			1		1	1	1	1			1
			1		1	1	1	1			1
			1		1	1	1	1			1
			1		1	1	1	1			1
			1		1	1	1	1			1
			1		1	1	1	1			1
			1		1	1	1	1			1
			1		1	1	1	1			1
			1		1	1	1	1			1
Topping, K., & Bryce, A. (2004).			1			1	1	1	1	1	1
			1			1	1	1			1
Topping, K., & et al. (2000).			1		1	1	1	1	1	1	1

			1		1	1	1	1	1	1	1
					1	1	1	1	1	1	1
					1	1	1	1	1	1	1
Topping, K., & et al. (2004).			1		1	1	1	1	1		1
Tymms, P., & et al. (2011).			1			1	1	1			1
			1			1	1	1			1
			1			1	1	1			1
			1			1	1	1			1
Veerkamp, M.B. (2001).	1	1			1	1		1	1		1
Watt, J. M., & Topping, K. J. (1993).			1			1	1	1			1
Willis, P,& et al. (2012)			1	1			1	1	1		
Winter, S. (1988).			1		1	1		1			1
			1		1	1	1	1			1
Winter, S (1996).			1		1		1	1			1
			1		1		1	1			1
Wright, J., & Cleary. K.S. (2006).			1	1		1	1	1			1
Yarrow, F., & Topping, K. (2001)			1		1	1	1	1	1	1	1
								1			1

Table 37. *Quantitative and qualitative peer tutoring studies in mathematics by socio interdependent elements*

	Social Interdependence, Main Elements										
Studies	***Rew-ard***	***Goal Perfor-mance***	***Fixed Roles***	***Interpersonal Communication Skills Training***	***Script Social Interdependence***	***Praise***	***Cross Ability***	***Pedagogical Skills Training***	***Elaborative Deep – Cognition***	***Meta Cogniti-on***	***Academic Script***
Allsopp, D.H. (1995).	1	1		1				1			
Browning, T.L.H. (1994).			1				1	1	1		
Caroll, B.P. (1995).			1				1	1	1		
Garton, A.F., & Pratt, C. (2001).							1				
							1				
Cooper, K.J. (1973).		1					1	1			
		1					1				
		1	1				1	1			
Dennis, L.M.G. (2013).			1				1				1
							1	1			1
Early, J.W. (1998).							1	1			
Fantuzzo, J.W., & et al. (1995).	1	1			1	1		1			1
Fantuzzo, J.W., & et al. (1990).		1						1			1
	1	1			1	1		1			1
Fantuzzo, J.W., & et al. (1992).	1	1		1	1	1		1			1

	1	1		1		1					
				1		1		1			1
				1				1			
Felman. R.S., & et al. (1976).	1		1				1				
Fitz-Gibbon, C.T. (1990).			1				1	1			
			1				1	1			
			1					1			
			1					1			
Fogarty, J.L., & Wang, M. (1982).			1				1	1			
Foreman, E. (1989).											
Fuchs, S.L., & et al. (1996).						1	1	1	1		
Fuchs, S.L., & et al. (1998).						1	1	1	1		
Fuchs, S.L., & et al. (1994).			1			1	1	1			
			1			1	1				
Fuchs, D., & et al. (2000)				1		1	1	1	1		
	1	1		1		1	1	1	1		
Fuchs, D., & et al. (1997)						1	1	1			
						1	1	1	1	1	
Fuchs, S.L., & et al. (1998).						1	1	1	1	1	
						1		1	1	1	
Ginzburg-Block, M., & Fantuzzo, J. (1997).	1	1			1	1		1			1
Ginsburg-Block, M.D., & Fantuzzo, J.W. (1998).	1	1			1	1		1	1	1	1
	1	1			1	1		1			
Gmitter, J.W. (1989).								1			
Greenwood, C.R., & et al. (1984).	1	1			1			1			
	1	1			1			1			

Greenwood, C.R., & et al. (1989).	1	1			1			1			
Hannah, D.C. (2008).		1					1	1			
Harris, V.W., & Sherman, J.A. (1973).	1										
	1										
Howell, K.W. (1978).			1								1
Inglis, A. (2002).			1				1	1			
			1				1	1			
			1				1	1			
			1				1	1			
Kenedy, L.J. (2000).			1				1		1		
Menses, K.F. & Gresham, F.M. (2009).	1	1	1		1	1		1		1	1
	1	1	1		1	1		1		1	1
	1	1	1		1	1		1		1	1
	1	1			1	1		1		1	1
	1	1			1	1		1		1	1
Novotni, S.M. (1985).			1	1			1	1			
Parham, J.W. (1993).			1				1	1			
			1				1				
			1								
			1								
Pheleps, E. & Damon, W. (1989).								1			
								1	1		
								1	1		
								1	1		
Rosse, S., & et al. (1978).			1				1				
			1								
Sharpley, A.M. & et al. (1983).			1				1	1			

							1	1			
Skinner, S.H. (1989).								1			
Smith, C.L. (2010).								1			
Tierney, P. (2005).			1	1		1	1	1		1	
Thurston, A., & et al. (2011).			1			1		1	1	1	1
Topping, K., & et al. (2003).			1		1	1	1	1	1		1
Tymms, P.,& et al. (2011).			1			1	1	1	1	1	1
			1			1	1	1	1	1	1
			1			1	1	1	1	1	1
			1			1	1	1	1	1	1
White, P.M. (2000).			1	1			1	1	1		1

Table 38. *Quantitative and qualitative peer tutoring studies in science by socio interdependent elements*

	Social Interdependence, Main Elements										
Studies	***Reward***	***Goal Performance***	***Fixed Roles***	***Interpersonal Communication Skills Training***	***Script Social Interdependence***	***Praise***	***Cross Ability***	***Pedagogical Skills Training***	***Elaborative Deep - Cognition***	***Meta Cogniti-on***	***Acade-mic Script***
Allen, V.L. & Feldman, C. (1973).			1				1	1			1
Bland, M., & Graham, H. (1989)							1	1		1	1
Brei-Crawley. M.J (2002).			1				1	1	1		
			1				1	1	1		
Ismail, H.N. (1999).				1				1	1	1	1
				1				1	1	1	
				1				1			
Kamps, D.M., & et al. (2008).	1	1		1		1		1	1	1	1
	1	1		1		1		1	1	1	1

	1	1		1		1		1	1	1	1
King, A. (1993).								1	1	1	1
								1			
King, A. (1994).								1	1		
								1	1	1	
								1	1	1	
								1	1		
King, A., & et al. (1998).				1		1		1	1	1	1
				1		1		1	1	1	
				1		1			1	1	
Merrill, M.E. (2002).			1	1			1	1			
Miller, K. (1989)			1				1	1			
Sanderson, P., & et al. (1992).			1	1			1	1			1
			1	1			1	1			1
			1	1			1	1			1

			1	1			1	1			1
			1	1			1	1			1
Siemens, B.K. (2001).			1	1			1	1			
Topping, K., & Bryce, A. (2004).			1			1	1	1			1
Tymms, P. (1989).			1				1				

Consent Forms:

Consent Form: Administered by the Teachers.

These questions are to see what you think about Mathematics, Mathematic context, and how you see yourself in general, and also see how good you are in Mathematics exercises. The aim of these questions is to see how successful group learning is. *The individual answers given here will not be shown to your teacher or anyone else.* Also, you are not required to complete this questionnaire, and can stop at any time. The information provided will be stored in a secured computer and only the researcher will have access to it. If the data is published, your name, surname will not be included on the paper; the information in this side page is required only for organizational purposes and consent. If you need to know the general findings of this questionnaire you can contact me on: mirjan.zeneli@durham.ac.uk.

Name and Surname: **School Name:**

Year group: **Teacher's name:**

Please circle one: Male /Female

If you agree to complete this questionnaire please sign below and complete the questions in the back.

Signature: ______________________ **Date:**

Example:

	Not at all True			somewhat True		Very True	
I like rain.	1	2	3	4	5	6	7

Figure 29. *Student consent form*

Student/Parent/guardian Information Sheet for the project called:

“Cooperation among Peer Learning theories of Cooperation?! A Social Interdependent/Constructivist Position in Peer Tutoring

The school has agreed to work with researchers from Durham University for 6 weeks to improve students’ performance in Mathematics and attitude towards mathematics and social skills in general. Older ages work together with younger ages for 25-30 minutes a week for 6 weeks. The main aim of the research is to improve our understanding of how group learning helps students in all these different ways. If we can find out *how* and *where* group learning helps students the most, then we can share this information with other schools and help them as well.

In order to find out *how* and *where* group learning helps students the most, some information would need to be collected. Those will be:

- *Information collected from mini-tests.* For example doing mathematics exercises individually.
- *Information collected from paper questionnaires*: This is a short questionnaire with 23 questions.
- *Information collected from looking at how you work with each other in the clasroom:* For example, how you help each other and how you speak to each other.

All individual information will not be shared with anyone. All the information will be saved into a secured University computer, which is protected with a password that only the researcher knows. Once all the information is on the computer, names will be replaced with numbers, for example the name Tony will be called 2212. This way no one will know how students worked with each other and how they answered the questionnaire/questions.

Once the information is finalised, other schools and actors will be informed of the overall results. This way other students can start helping each other and make many friends.

You can withdraw from this study at any time, just inform us through any of the following ways: mirjan.zeneli@durham.ac.uk - 078 078 5 5353. Or by simply contacting the school.

Thank you for your help.

Project team member’s name ___ Mirjan Zeneli

Project team member’s signature_______

Date_________________15-05-2013

Figure 30. *Student, parent/guardian information and informed consent sheet 1*

Student/Parent/Guardian Information Sheet for the project called:

"Cooperation among Peer Learning Theories of Cooperation?! A Social Interdependent/Constructivist Position in Peer Tutoring

The school has agreed to work with researchers from Durham University for 6 weeks to improve students' performance in mathematics and attitude towards mathematics and social skills in general. The project has now successfully ended. The main aim of the research is to improve our understanding of how group learning helps students in different ways.

In order to find out *how* and *where* group learning helps students the most, and measure the level which the project was applied properly by the students, the following data needs to be analysed:

- *The completed and uncompleted lesson scripts on which the students have worked for each lesson.*

All individual information will not be shared with anyone. All the information will be saved into a secured Durham University computer, which is protected with a password that only the researcher knows. Once all the information is on the computer, names will be replaced with numbers, for example the name Tony will be called 2212. This way no one will know how students worked with each other and how they answered the questionnaire/questions.

Once the information is finalised, other schools and actors will be informed of the overall results. This way other students can start helping each other and make many friends.

You can withdraw from this study at any time, and the lesson scripts will not be analysed for your case, just inform us through any of the following ways: mirjan.zeneli@durham.ac.uk - 078 078 5 5353. Or by simply contacting the school.

Thank you for your help.

Project team member's name Mirjan Zeneli

Project team member's signature

Date: 07-08-2013

Figure 31. *Student, parent/guardian information and informed consent sheet 2*

Ethics approvals from the School of Education:

Inbox

Thursday, May 16, 2013 10:13 A

Dear Mirjan

I am pleased to inform you that your application for ethical approval has now been granted by the School of Education Ethics Committee in respect of 'Cooperation among Peer Learning theories of Cooperation? A Social interdependent/constructivist position in peer tutoring'.

May we take this opportunity to wish you good luck with your research.

Best wishes

Sheena Smith
Research Office
School of Education
Durham University

Figure 32. *Ethics approval confirmation from the School of Education 1*

Ethics approval for using the lesson scripts.

ZENELI M. Dear Sheena, Thank you for the quick and positive responce. Have a lovely day. Best Wishes, Mirjan Zeneli 8/7/2013

Action:

SMITH J.C.

To: ZENELI M.

Cc: TYMMS P.B.; BOLDEN D.S.; ED-PGRSTUDENTS E.

Inbox

Wednesday, August 07, 2013 4:05 PM

- You replied on 8/7/2013 4:18 PM.

Dear Mirjan

I am pleased to inform you that your application for ethical approval in respect of analysing the data used on the 'Cooperation among peer learning theories of cooperation? A social interdependent/constructivist position in peer tutoring' has been approved by the School of Education Ethics Committee.

May we take this opportunity to wish you good luck with your research.

Best wishes.

Sheena Smith
Research Office
School of Education
Durham University

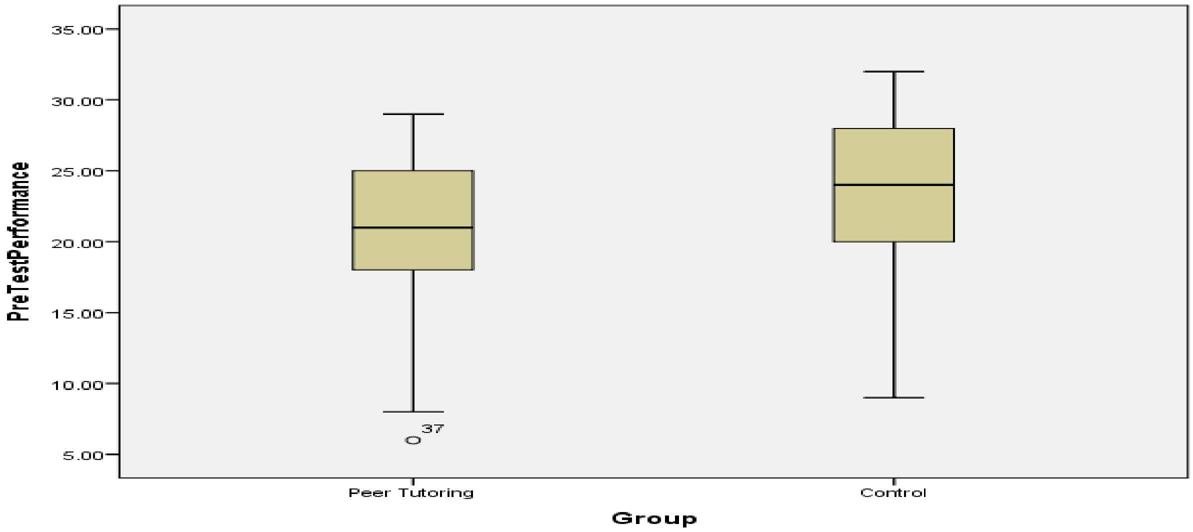
Tel: (0191) 334 8403
www.dur.ac.uk/education

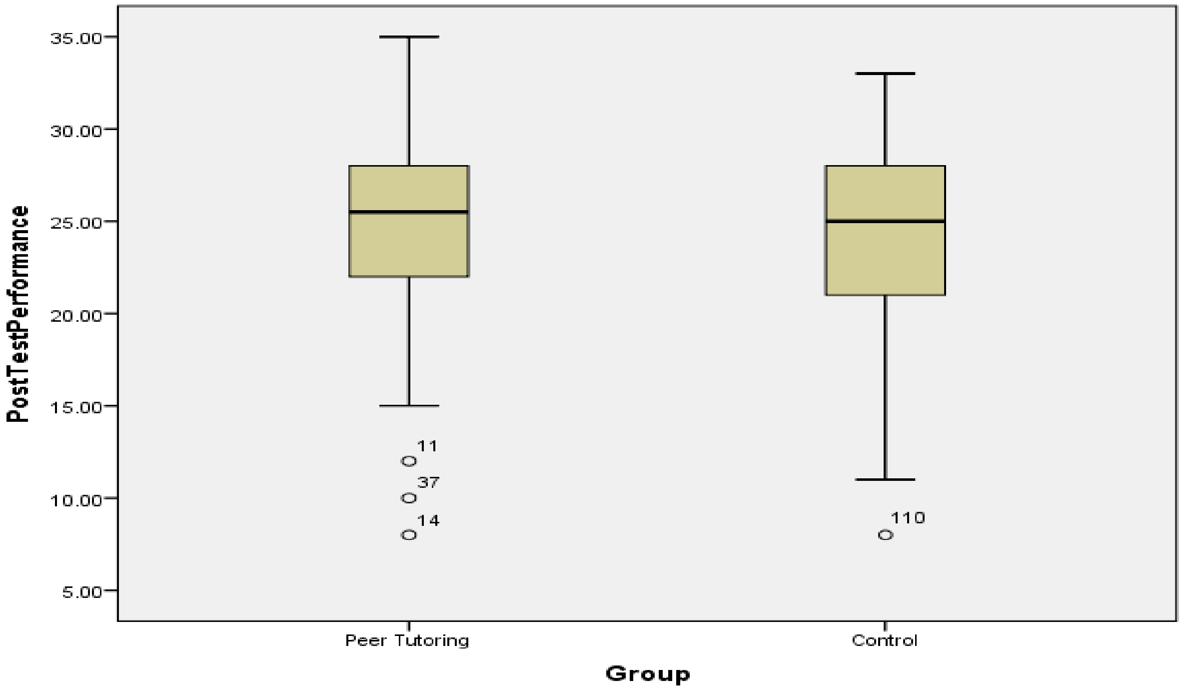
Figure 33. *Ethics approval confirmation from the School of Education 2*

Example of SPSS procedures:

School A year 6 students, testing the impact of ICAT on Mathematics performance

Step 1: Box-plots - Adjusting the outliers and testing for distribution of scores normality



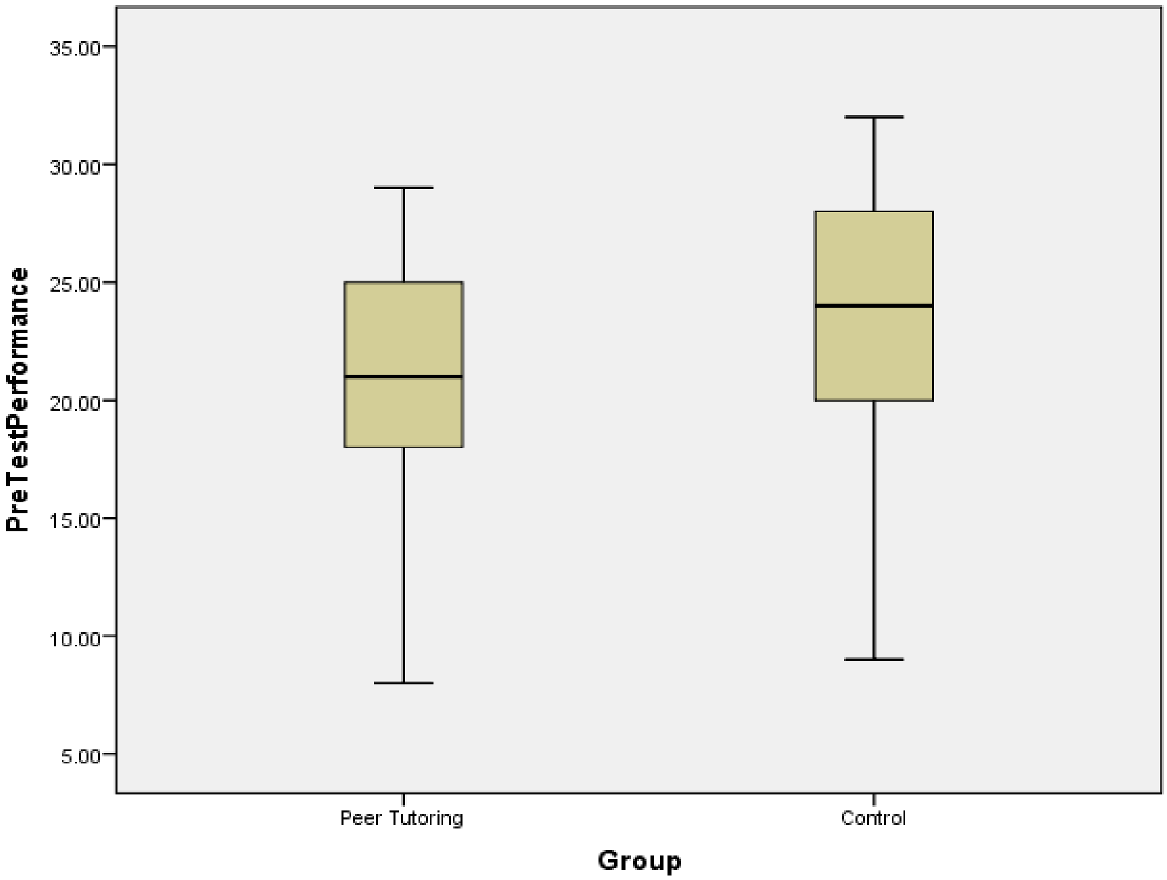


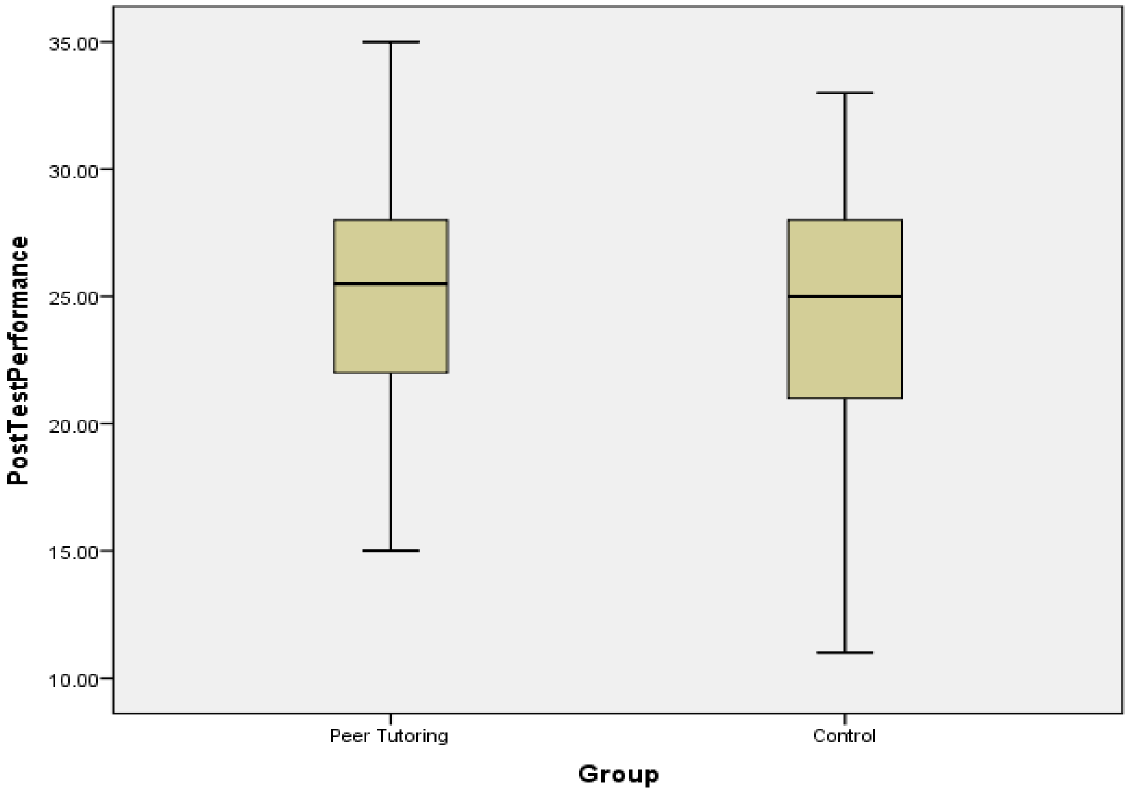
After Adjusting the Outliers:

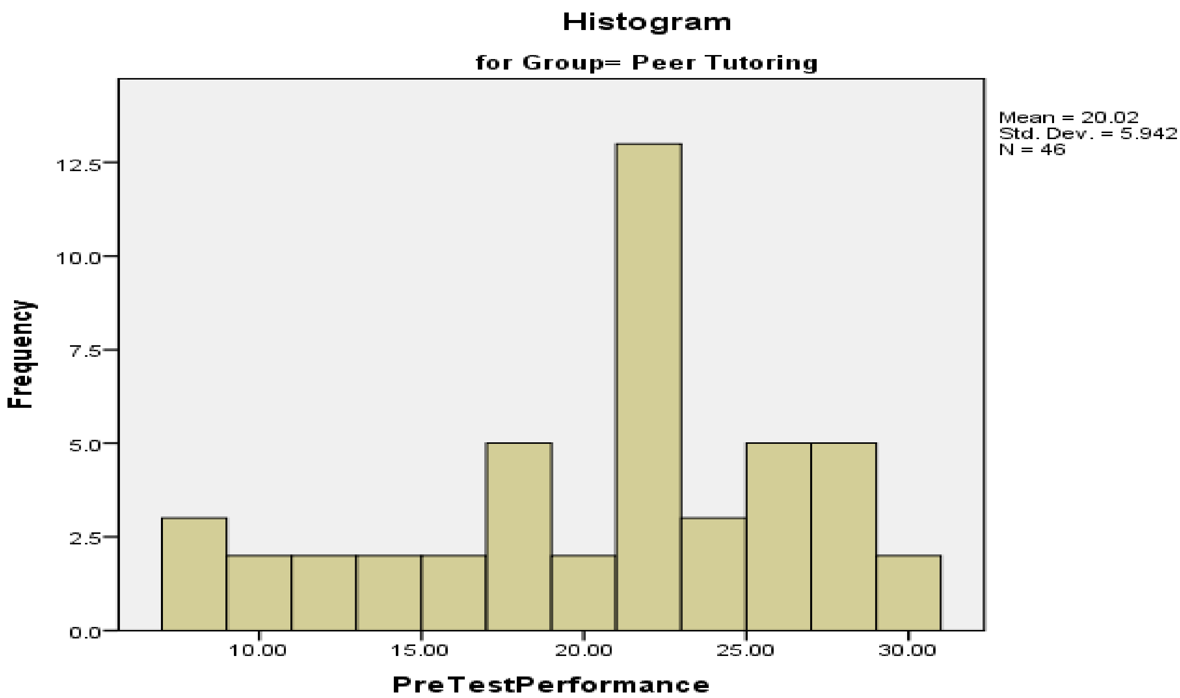
Tests of Normality

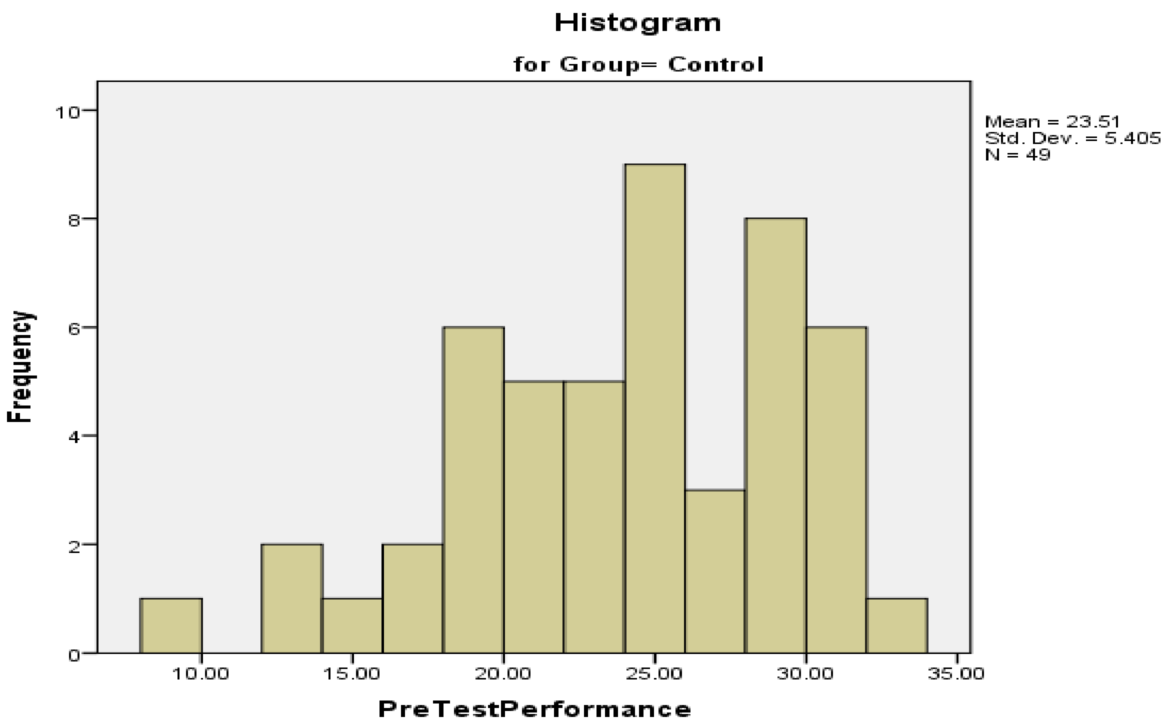
	Group	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
PreTestPerformance	Peer Tutoring	.174	46	.001	.931	46	.009
	Control	.103	49	.200*	.957	49	.074
PostTestPerformance	Peer Tutoring	.112	46	.189	.965	46	.176
	Control	.096	49	.200*	.960	49	.097

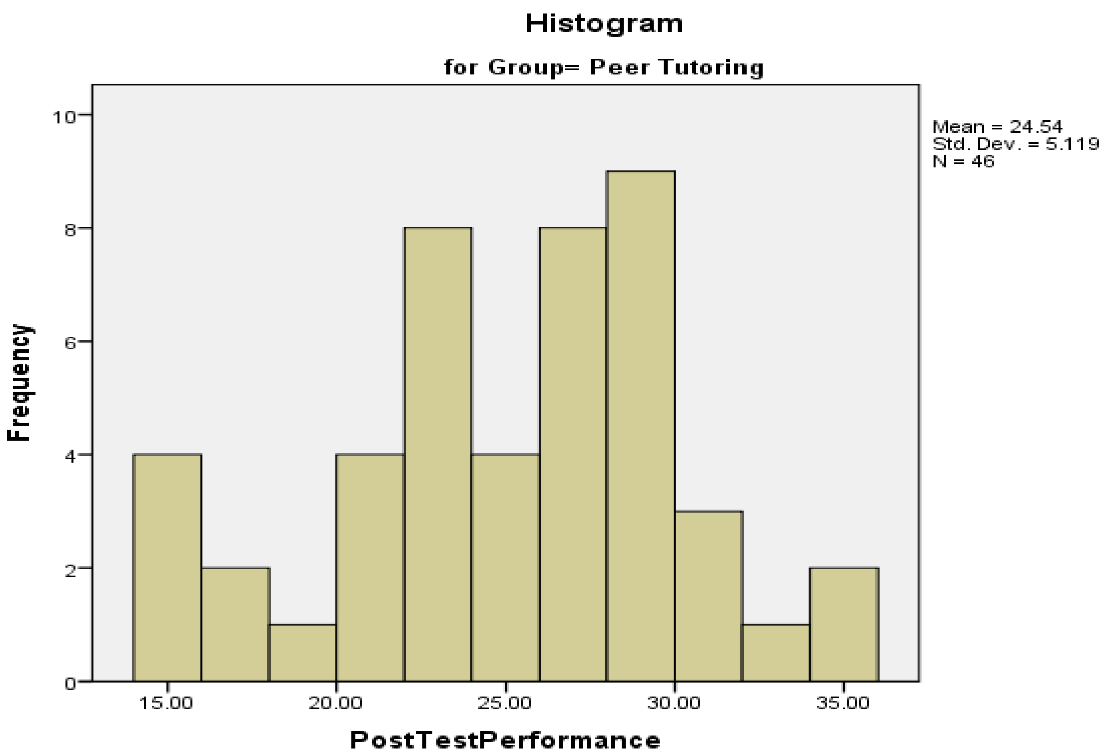
a. Lilliefors Significance Correction
*. This is a lower bound of the true significance.

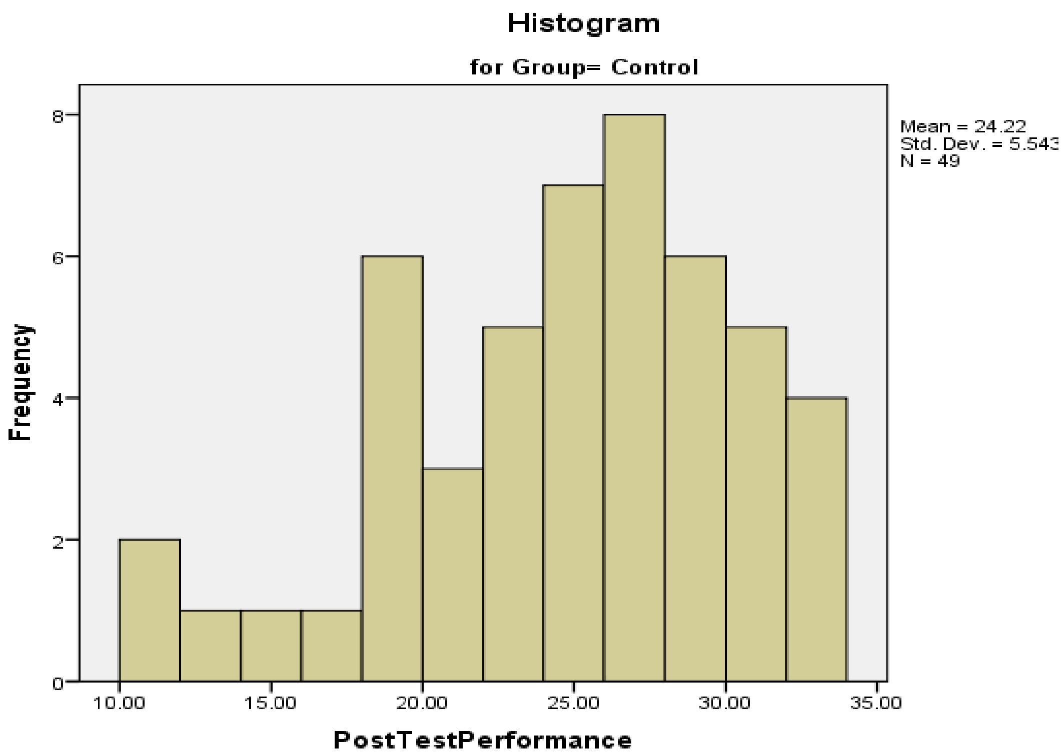



35.00
30.00
25.00
20.00
15.00
10.00
PostTestPerformance
Peer Tutoring
Control
Group

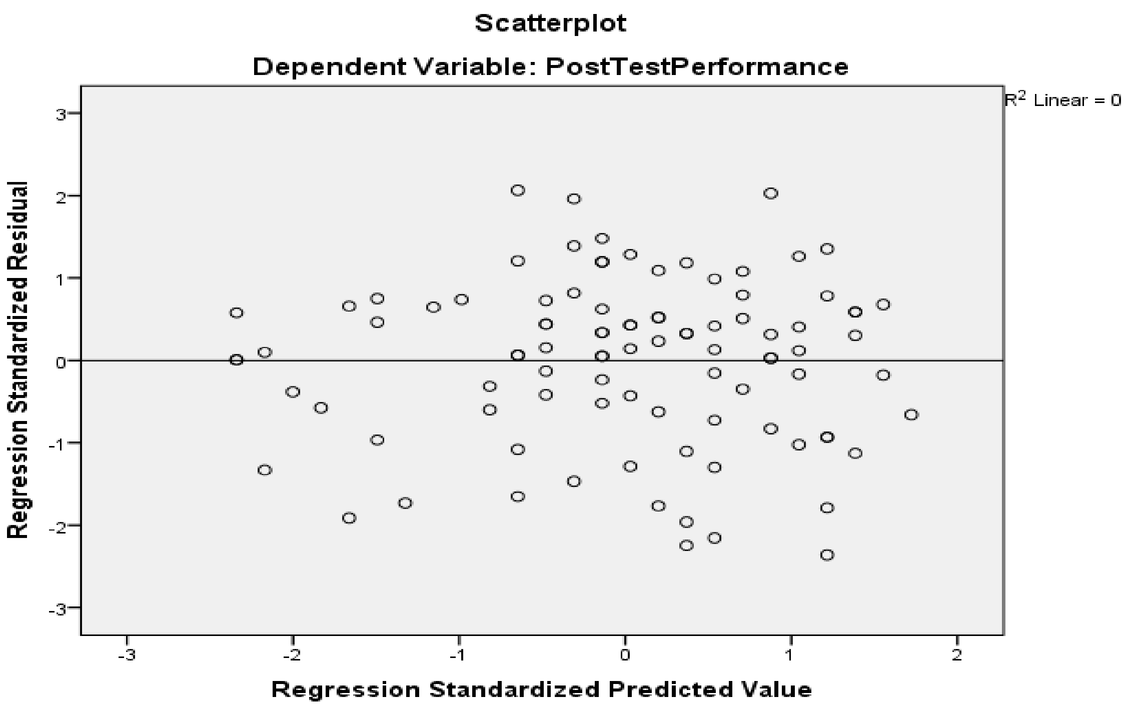

Histogram
for Group= Peer Tutoring
Mean = 20.02
Std. Dev. = 5.942
N = 46
12.5
10.0
7.5
5.0
2.5
0.0
Frequency
10.00
15.00
20.00
25.00
30.00
PreTestPerformance


Histogram
for Group= Control
Mean = 23.51
Std. Dev. = 5.405
N = 49
Frequency
0
2
4
6
8
10
10.00
15.00
20.00
25.00
30.00
35.00
PreTestPerformance

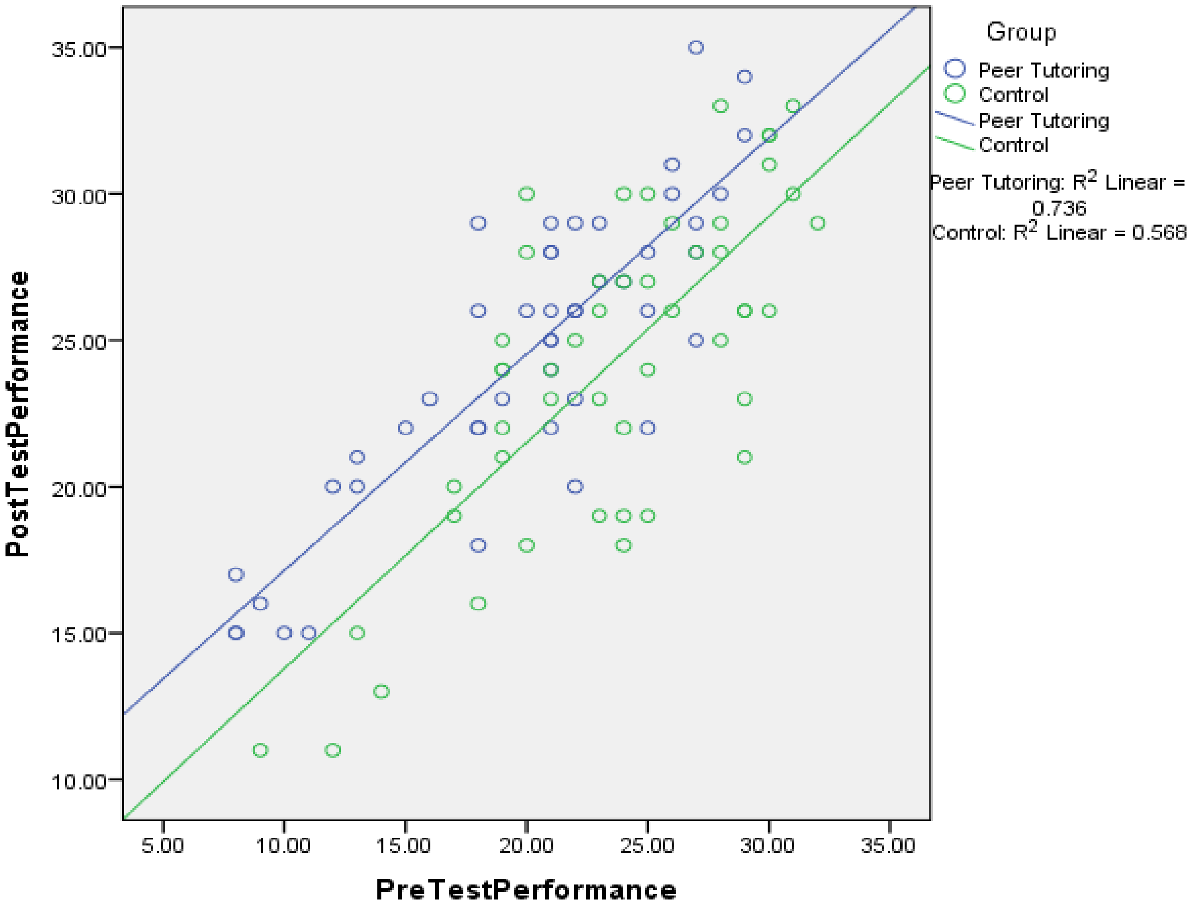

Histogram
for Group= Peer Tutoring
Mean = 24.54
Std. Dev. = 5.119
N = 46
Frequency
0
2
4
6
8
10
15.00
20.00
25.00
30.00
35.00
PostTestPerformance



Step 2: Checking for Homoscedasticity



Step 3: Checking for a positive linear relationship between pre-post test data



Step 4: Checking for residual score normality

Tests of Normality

	Group	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	Peer Tutoring	.108	46	.200*	.978	46	.518
	Control	.077	49	.200*	.979	49	.512

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Step 5: Checking for homogeneity of regression lines between the groups

Tests of Between-Subjects Effects

Dependent Variable:PostTestPerformance

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1707.522[a]	3	569.174	54.588	.000
Intercept	362.767	1	362.767	34.792	.000
Group	19.716	1	19.716	1.891	.172
PreTestPerformance	1702.335	1	1702.335	163.266	.000
Group * PreTestPerformance	.835	1	.835	.080	.778
Error	948.836	91	10.427		
Total	59118.000	95			
Corrected Total	2656.358	94			

a. R Squared = .643 (Adjusted R Squared = .631)

Step 6: Running the analysis

Levene's Test of Equality of Error Variances[a]

Dependent Variable:PostTestPerformance

F	df1	df2	Sig.
6.802	1	93	.011

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + PreTestPerformance + Group

Findings:

Tests of Between-Subjects Effects

Dependent Variable:PostTestPerformance

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	1706.687[a]	2	853.344	82.668	.000	.642	165.336	1.000
Intercept	374.584	1	374.584	36.288	.000	.283	36.288	1.000
PreTestPerfor-mance	1704.273	1	1704.273	165.103	.000	.642	165.103	1.000
Group	188.578	1	188.578	18.269	.000	.166	18.269	.988
Error	949.671	92	10.323					
Total	59118.000	95						
Corrected Total	2656.358	94						

a. R Squared = .642 (Adjusted R Squared = .635)

b. Computed using alpha = .05

Estimates

Dependent Variable:PostTestPerformance

Group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Peer Tutoring	25.902[a]	.485	24.938	26.866
Control	22.949[a]	.470	22.017	23.882

a. Covariates appearing in the model are evaluated at the following values: PreTestPerformance = 21.8211.

Pairwise Comparisons

Dependent Variable:PostTestPerformance

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
Peer Tutoring	Control	2.952*	.691	.000	1.580	4.324
Control	Peer Tutoring	-2.952*	.691	.000	-4.324	-1.580

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Step 7: Residual gain analysis

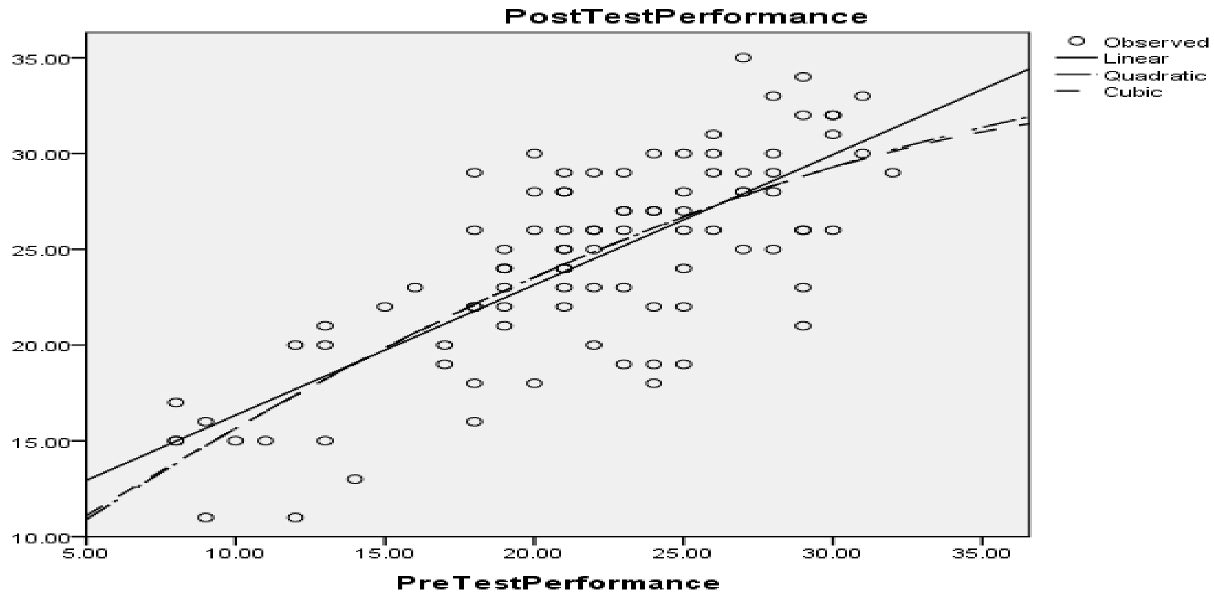
Step 7.1: Identifying the best regression line

Model Summary and Parameter Estimates

Dependent Variable:PostTestPerformance

Equation	Model Summary					Parameter Estimates			
	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Linear	.572	124.036	1	93	.000	9.534	.680		
Quadratic	.579	63.149	2	92	.000	5.603	1.111	-.011	
Cubic	.579	41.646	3	91	.000	6.246	.994	-.004	.000

The independent variable is PreTestPerformance.



Step 7.2: Running a t-test with the best pre-posttest regression line residuals

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Error for PostTestPerformance with PreTestPerformance from CURVEFIT, MOD_1 QUADRATIC	Peer Tutoring	46	1.3573805	2.64822440	.39045939
	Control	49	-1.2742755	3.65030858	.52147265

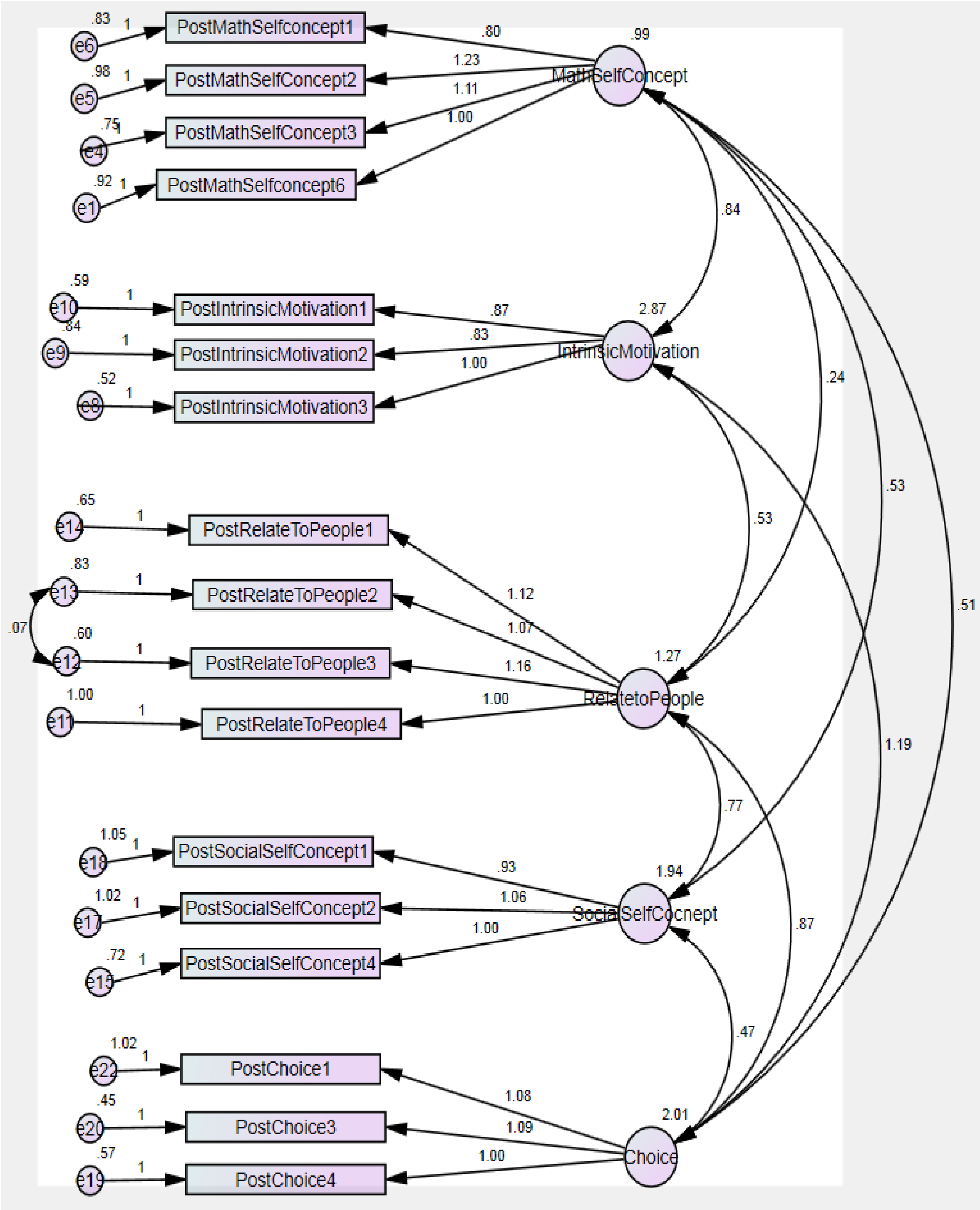
Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Error for PostTestPerformance with PreTestPerformance from CURVEFIT, MOD_1 QUADRATIC	Equal variances assumed	6.823	.010	4.000	93	.000	2.63165600	.65793860	1.32512033	3.93819167
	Equal variances not assumed			4.040	87.554	.000	2.63165600	.65145396	1.33693631	3.92637569

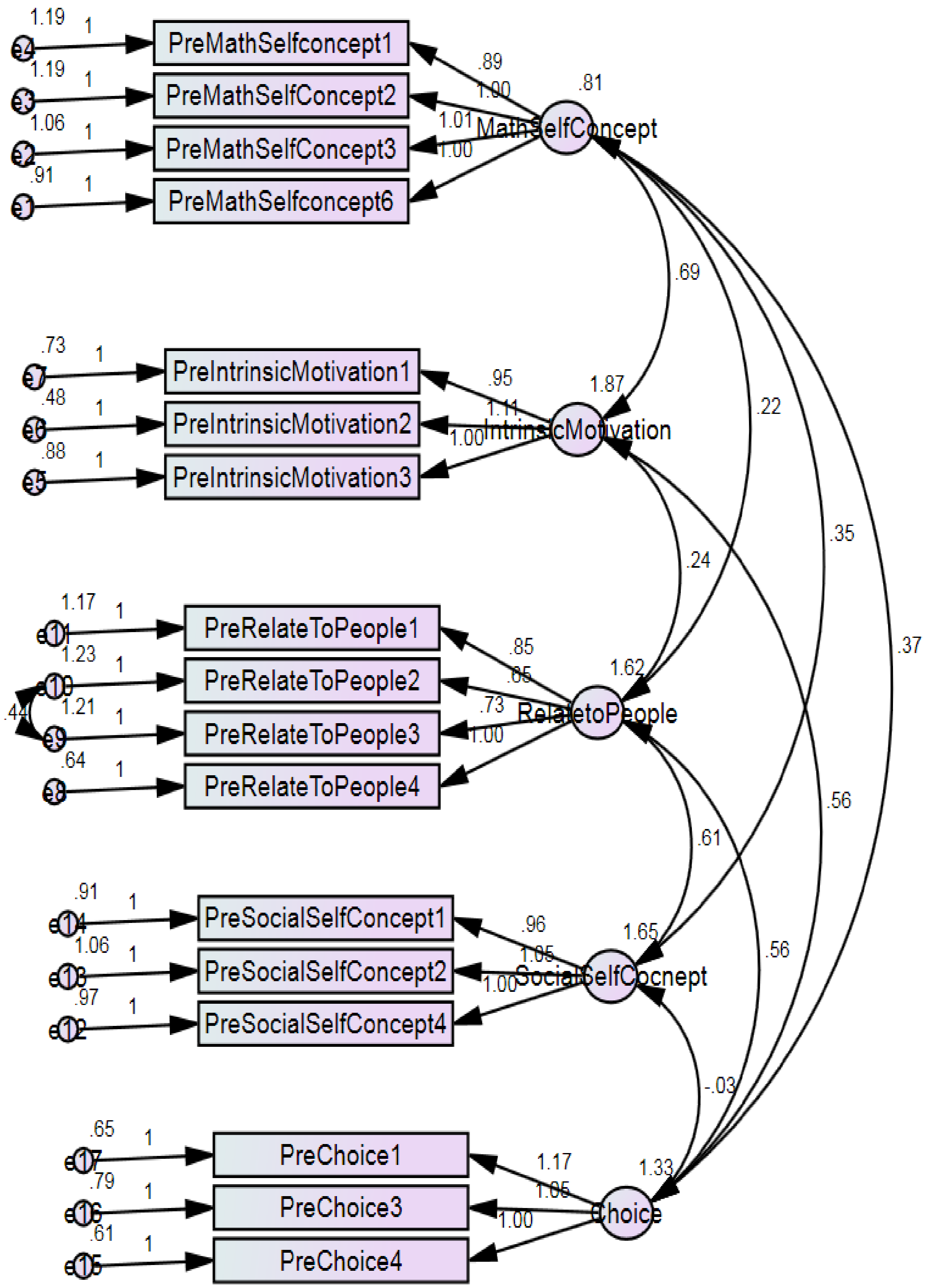
Instrument Development Diagrams: ***Confirming the model***

School A

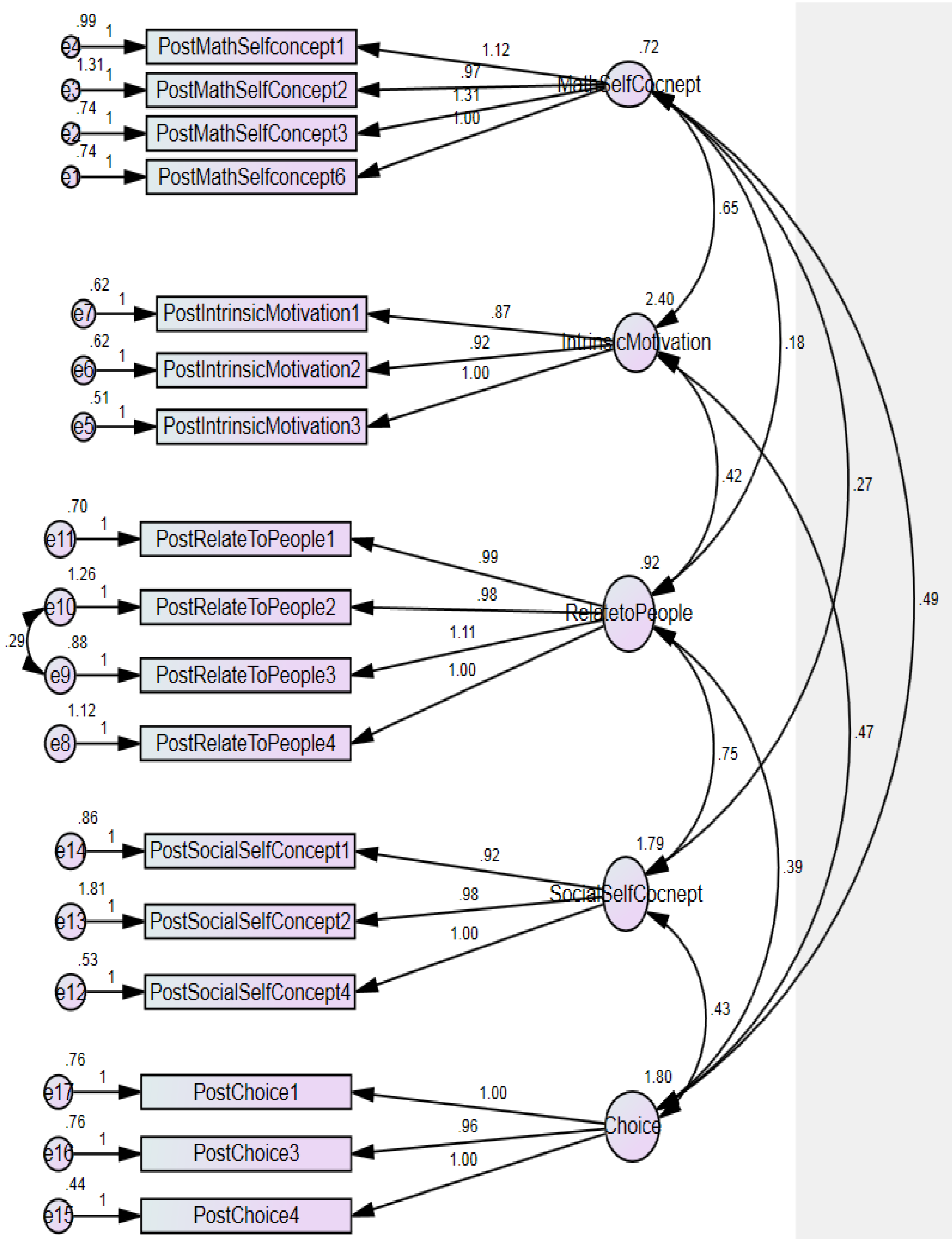
Year 6 post-test



Year 8 pre-test data

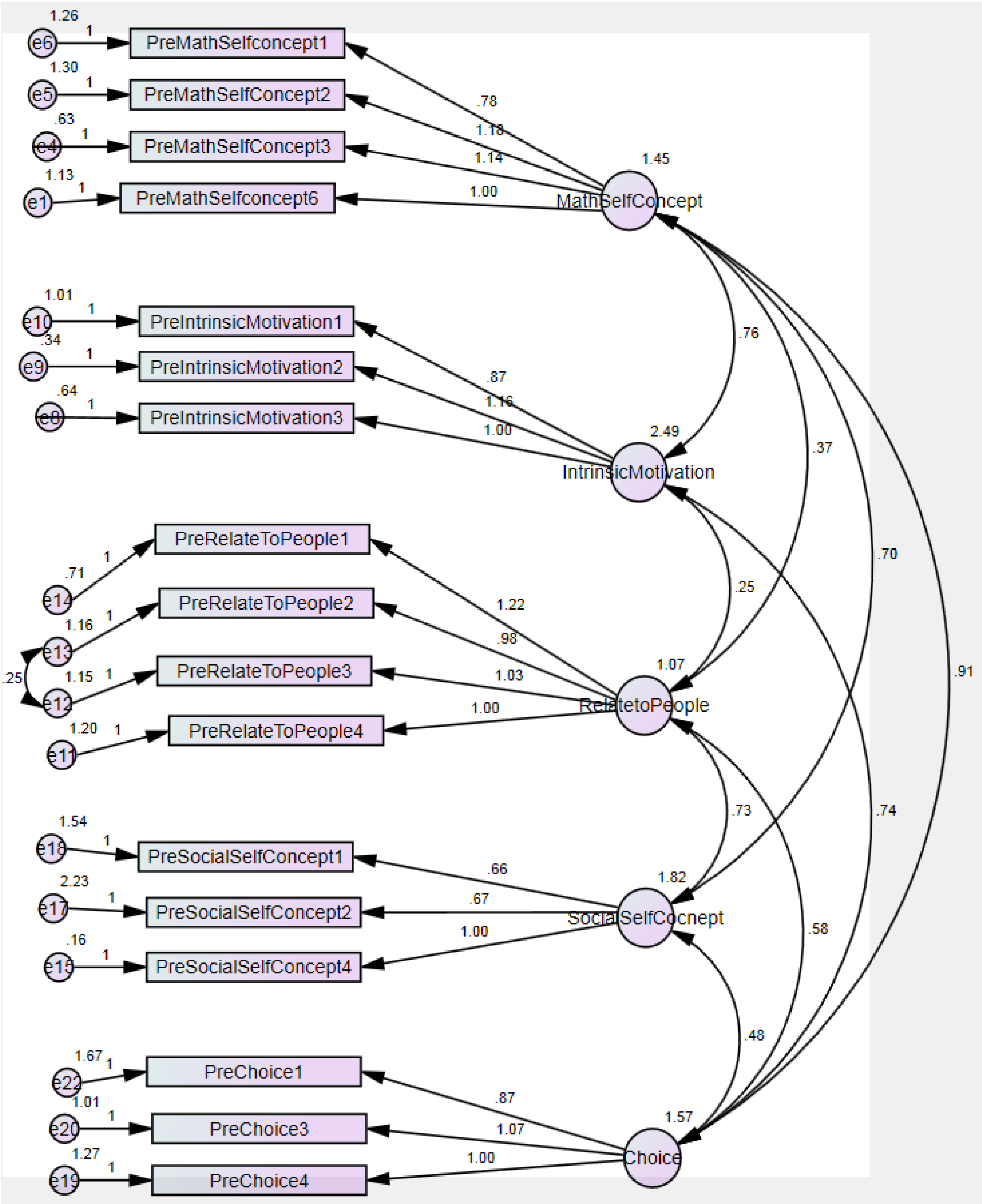


Year 8 post-test

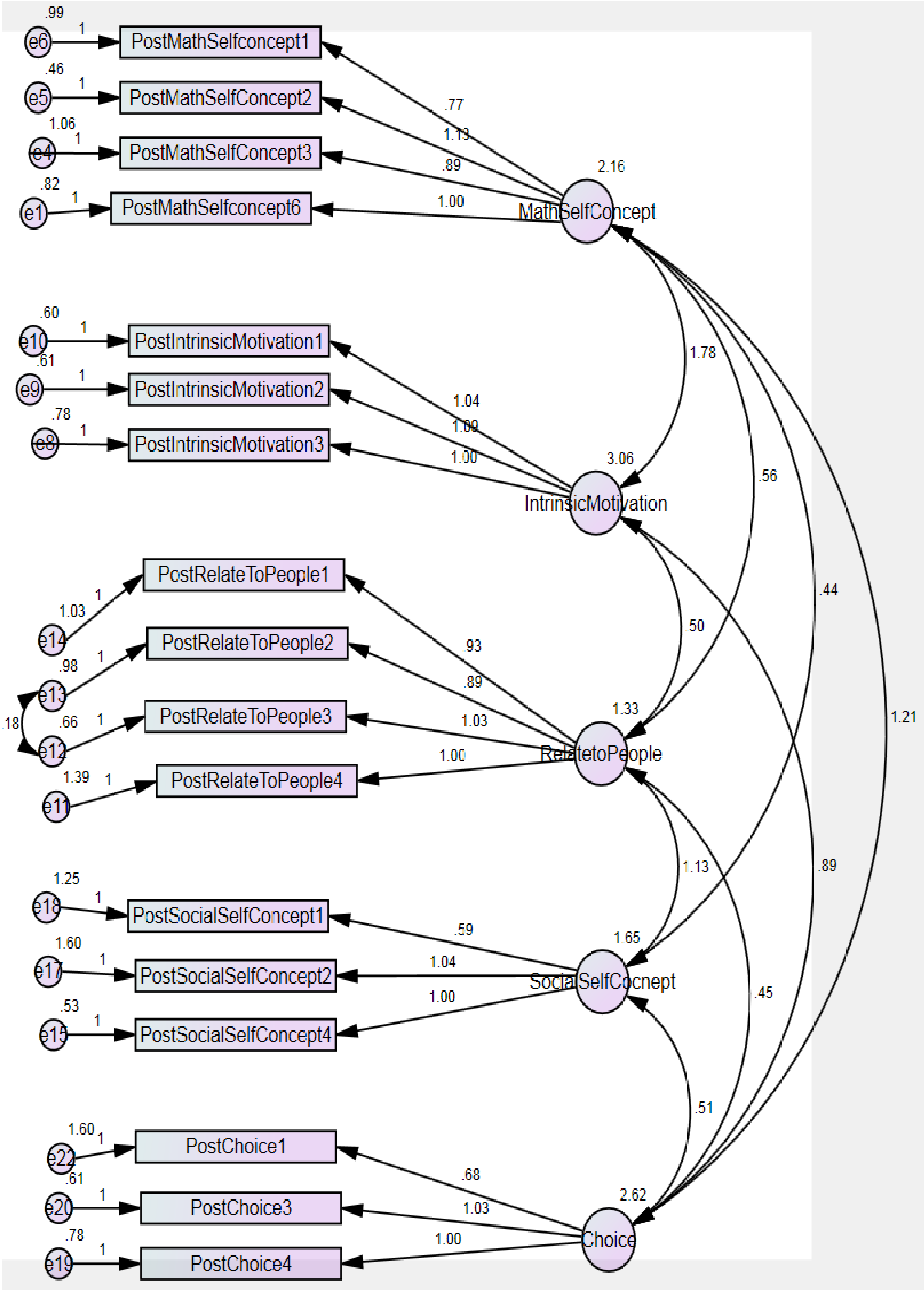


School B

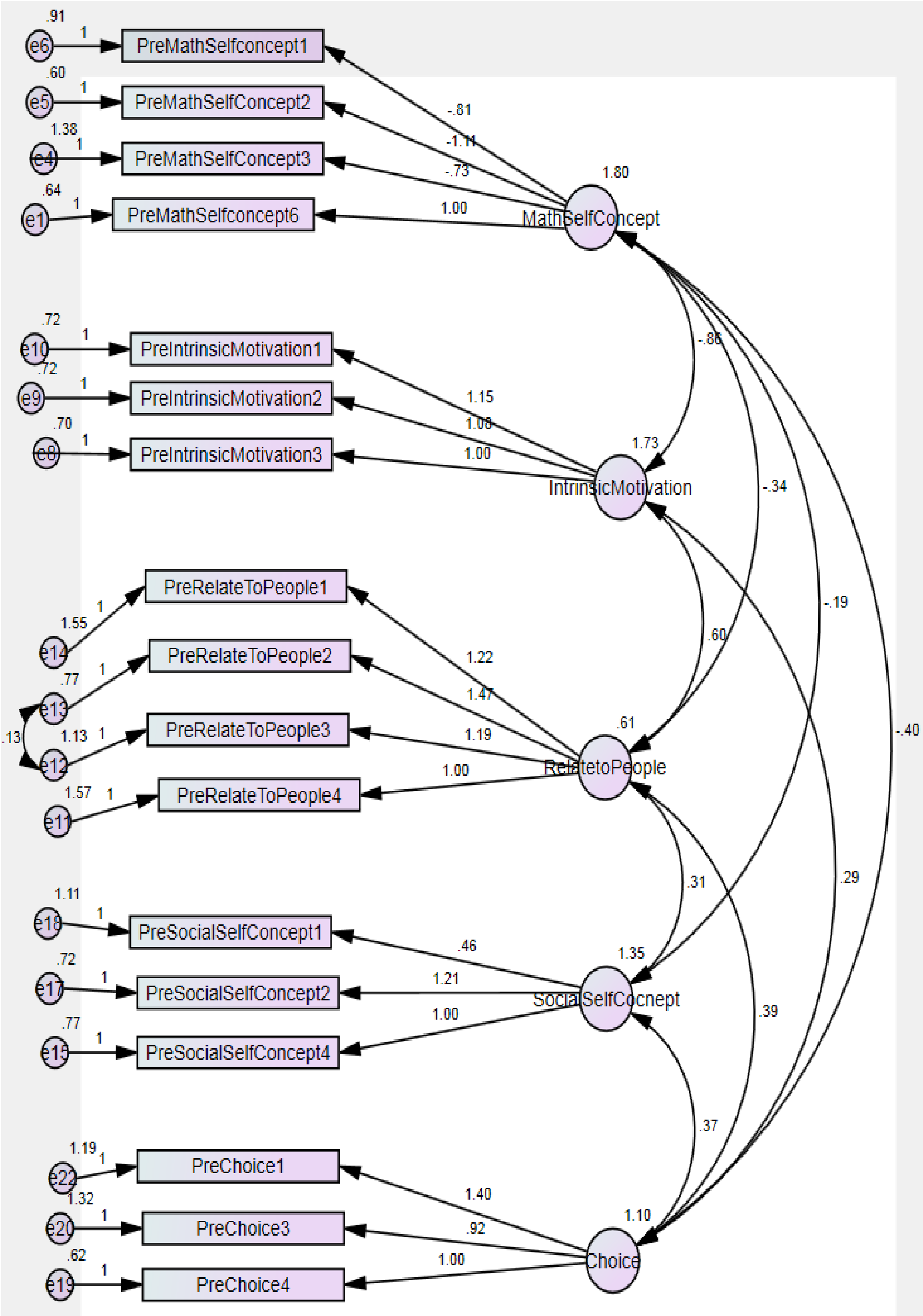
Year 7 pre-test



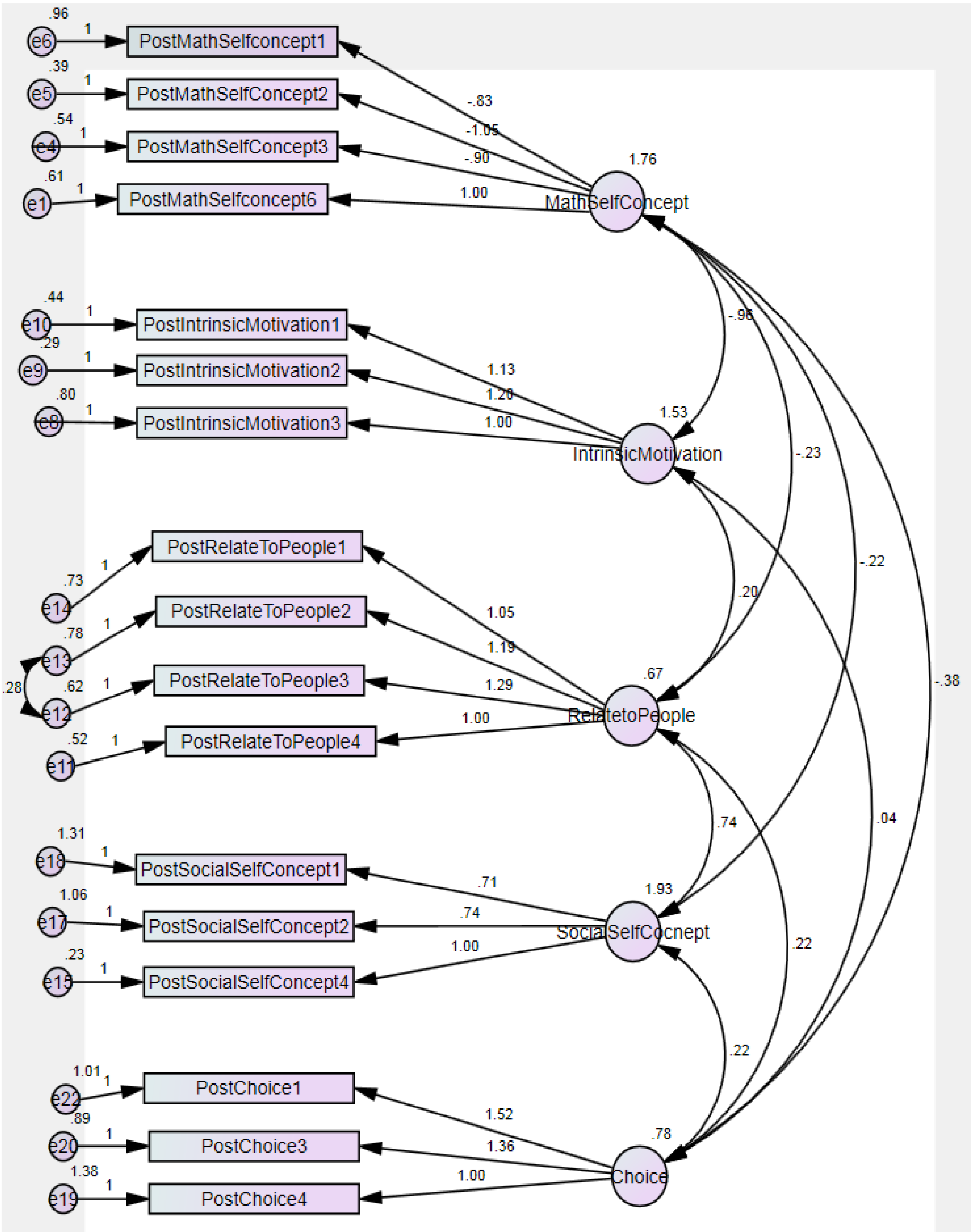
Year 7 post-test



Year 9 pre-test

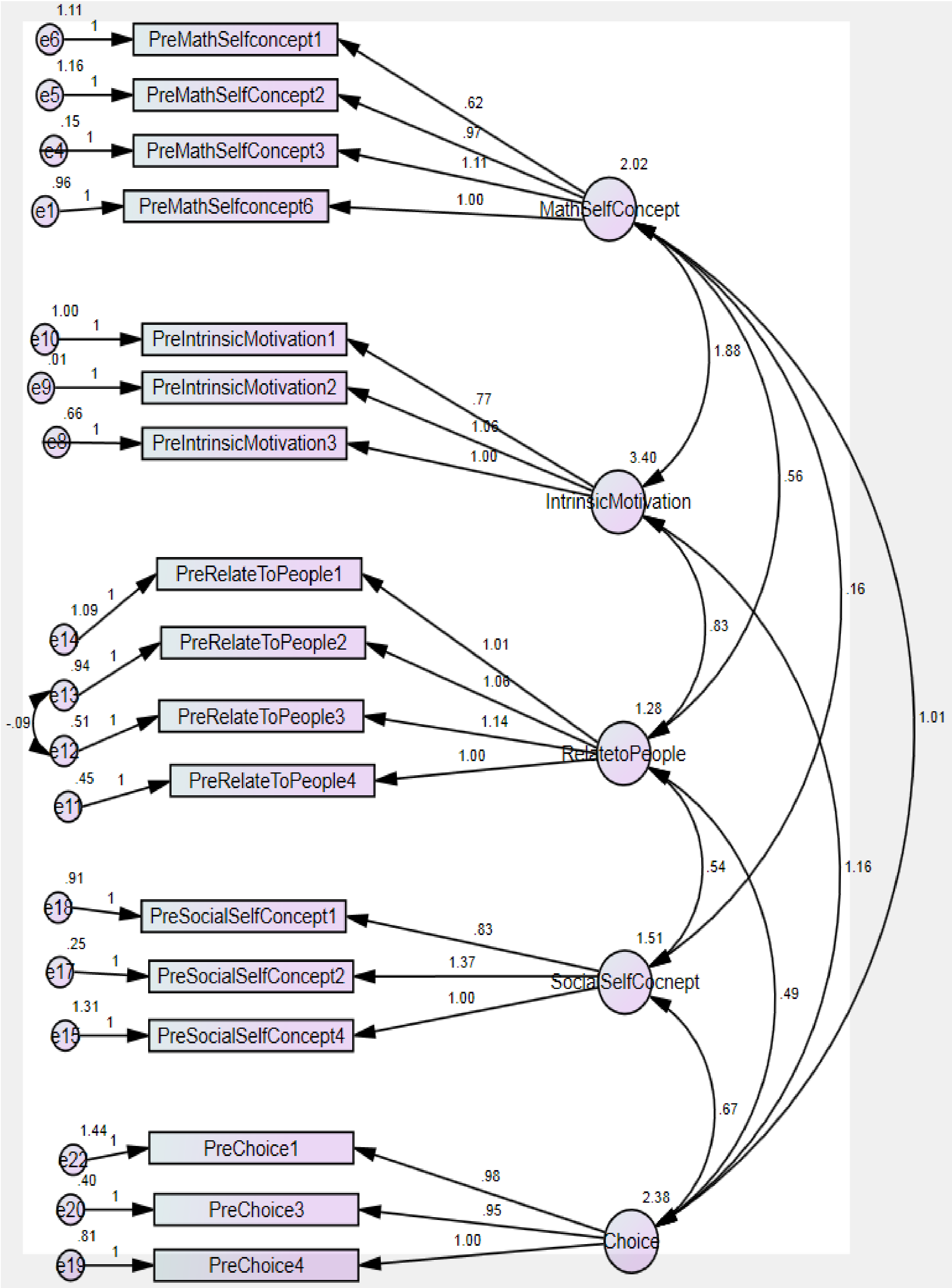


Year 9 post-test



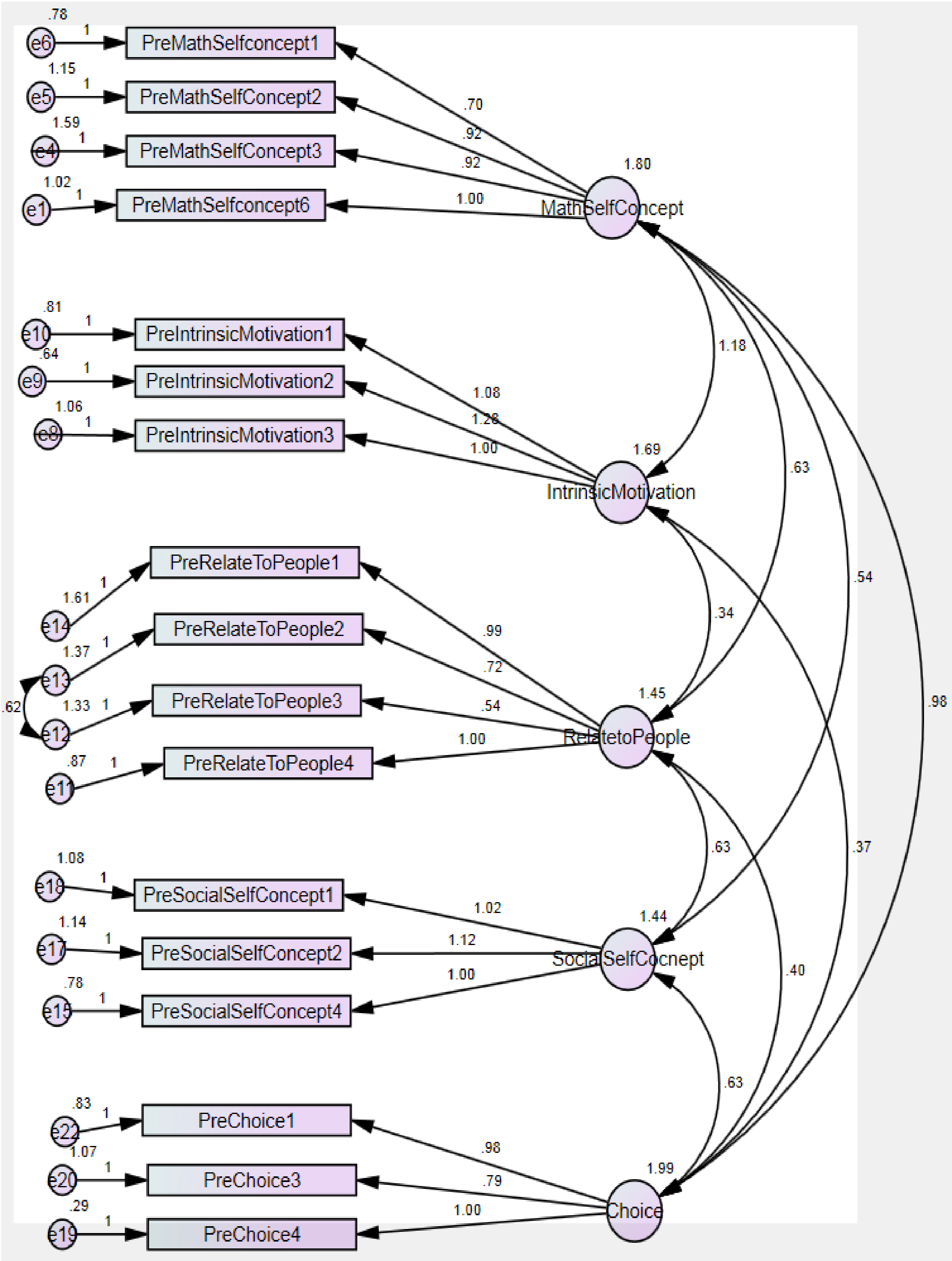
School C

Year 8 pre-test:



Year 8 post-test

Year 10 pre-test



Year 10 post-test