

An Exploratory Evaluation of a Paired Maths
Intervention with Secondary Aged Pupils

Sarah Cooper

Thesis submitted to the University of Nottingham for the
degree of Doctor of Applied Educational Psychology

May 2016

Word Count: 40,269 (excluding titles)

Acknowledgements

The researcher would like to thank all those who have supported them during this research and throughout their doctorate, including:

The tutor team at the University of Nottingham, and particularly Nathan Lambert for his support, guidance and supervision.

All the Educational Psychologists in the team and particularly my supervisors, for their support, encouragement and mentoring, throughout my time at the service.

The schools, parents and children who participated in this research, for their time and commitment to the project.

To my friends and family, particularly my parents, for their constant encouragement, support and understanding.

Contents

Contents of Tables	10
Contents of Figures	15
Abstract	18
1 Introduction	20
2 Literature Review	23
2.1 Mathematics.....	23
2.1.1 Mathematics performance in the UK.....	23
2.1.2 Long-term outcomes for poor mathematics/numeracy.....	23
2.2 Is mathematics shadowed by literacy?	24
2.2.1 Intervention for mathematics difficulties	25
2.2.2 Effective teaching of mathematics	27
2.2.3 Intervention for mathematics difficulties, what works?	28
2.3 Peer- tutoring	30
2.3.1 Definitions	30
2.3.2 Historical and theoretical underpinnings of peer-tutoring	32
2.3.3 Peer-tutoring and links to effective mathematics instruction/intervention	
34	
2.3.4 The evidence base for peer-tutoring	35
2.3.5 Summary: review of evidence for peer-tutoring approaches	44
2.4 Systematic Literature Review	45
2.4.1 Definitions	45
2.4.2 Aims of the systematic literature review	46
2.4.3 Eligibility criteria.....	46
2.4.4 The search and filtering process	48

2.4.5	Weight of Evidence	49
2.4.6	Synthesis of papers	51
2.4.7	Overall synthesis of studies.....	55
2.4.8	Review of key unpublished research	56
2.4.9	Systematic review: Conclusion.....	60
2.5	Intervention within the secondary aged population	60
2.6	Self-esteem.....	61
2.6.1	Definitions	61
2.6.2	Potential outcomes surrounding self-esteem	61
2.6.3	What does this mean for intervention around self-esteem in schools? ..	66
2.7	Conclusion: Rationale and original contribution for the current research	66
3	Method.....	68
3.1	Introduction.....	68
3.2	Theoretical rationale	68
3.2.1	Ontology, epistemology and methodology	68
3.2.2	The chosen paradigm: (Post)Positivism	71
3.2.3	Experiments and RCTs.....	72
3.3	Research Design	73
3.3.1	Investigating effects for tutees	73
3.3.2	Investigating effects for tutors.....	74
3.4	Characteristics of the current study	76
3.4.1	Context: The locality	76
3.4.2	Stakeholders.....	76
3.5	Participants.....	77
3.5.1	Schools	77

3.5.2	Intervention facilitators	78
3.5.3	Pupils	79
3.5.4	Pupil Characteristics	82
3.6	Participant allocation	82
3.6.1	Year 8 participants in the role of the tutee.....	82
3.6.2	Year 10 participants in the role of the tutor	83
3.7	Experimental group: Paired Maths	83
3.7.1	PM in the current research	84
3.7.2	Training.....	86
3.8	Comparison group for year 8 participants	87
3.8.1	Rationale for comparison group	87
3.8.2	Comparison group in this research	87
3.9	Waitlist Control Groups (Year 8s and Year 10).....	87
3.10	Treatment integrity.....	88
3.11	Measures	89
3.11.1	Maths attainment: Basic Number Screening Test	89
3.11.2	Measuring self-esteem.....	92
3.11.3	The Burnett Self-Scale (BSS)(Burnett, 1994)	93
3.11.4	The Behavioural Indicators of Self-Esteem Scale (BIOS) (Burnett, 1998)	94
3.12	Procedure.....	95
3.13	Research considerations.....	96
3.13.1	Internal validity	97
3.13.2	External validity.....	100
3.13.3	Reliability.....	101
3.13.4	Ethical considerations	103

3.14	Approach to Data Analysis	106
3.14.1	Statistical tests	106
3.14.2	Missing data	109
3.15	Summary	109
4	Results	110
4.1	Research Question 1: In a UK secondary school, does a 7 week, cross-age, structured peer-tutor programme (Paired Maths) enhance mathematics attainment in Year 8 students in the role of <i>Tutee</i> ?	110
4.1.1	Basic Number Screening Test.....	110
4.1.2	Summary: Basic Number Screening Test	118
4.2	Summary Research Question 1	118
4.3	Research Question 2: In a UK secondary school, does a 7 week, cross-age, structured peer-tutor programme around maths (Paired Maths) enhance self-esteem in Year 8 students in the role of <i>Tutees</i> ?	118
4.3.1	Behavioural Indicator of Self-esteem (BIOS)	119
4.3.2	Burnett Self Scale	127
4.3.3	Global Self-Esteem Scale.....	127
4.3.4	Inferential statistics.....	133
4.3.5	BSS Learning Self-Concept Scale	135
4.3.6	BSS Mathematics Self-Concept Scale.....	143
4.3.7	Summary: Research Question 2	151
4.4	Research Question 3 In a UK secondary school, does a 7 week, cross-age, structured peer-tutor programme around maths (Paired Maths) enhance self-esteem in Year 10 students in the role of <i>Tutors</i> ?.....	152
4.4.1	Behavioural Indicator of Self Esteem	152
4.4.2	Burnett Self Scale	160

4.4.3	Global Self-Esteem	160
4.4.4	BSS Learning Self-Concept Scale	168
4.4.5	BSS Mathematics Self-Concept Scale	176
4.4.6	Summary: Research Question 3	181
4.5	Summary of Results Section	182
5	Discussion.....	184
5.1	Introduction.....	184
5.2	Research Question 1: Impact on maths attainment for Year 8 tutees	184
5.2.1	Comparison of current findings to existing literature	185
5.3	Research Question 2: Impact on Self-esteem for Year 8 Tutees.....	187
5.3.1	Comparison of current findings to existing literature	187
5.4	Research Question 3: Impacts on Self-Esteem for Year 10 Tutors	189
5.4.1	Comparison of current findings to Existing Literature.....	190
5.5	Overall Summary of Findings.....	190
5.6	Methodological Limitations.....	191
5.6.1	Measures.....	191
5.6.2	Internal validity	194
5.6.3	Treatment integrity	196
5.6.4	Ecological validity	198
5.7	Further Possible Explanations	199
5.7.1	Possible explanations relevant to all research questions.....	200
5.7.2	Possible explanations of findings for Research Question 1: Mathematics attainment for Tutees	202
5.7.3	Possible explanations for Research Question 2: Impact on self-esteem for tutees	204

5.7.4	Possible explanations for Research Question 3: Impact on self-esteem for tutors	205
5.8	Reflections on Real World Research and the Post-Positivist paradigm	206
5.9	Implications for Directions for Future Research	207
5.10	Implications for Educational Psychology Practice	210
5.11	Summary of Discussion	211
6	Conclusion	212
6.1	Unique Contribution of This Research	212
6.2	Final Conclusion	213
7	References	214
8	Appendices	226
8.1	A review of the selected literature, by paper, forming the systematic literature review	226
8.2	Paired Maths Diagram	229
8.3	Additional Guidance for Tutors	230
8.4	Maths aid (original is A3 sized)	233
8.5	Paired Maths Training Presentation	234
8.6	Guidelines for TA Comparison Group	237
8.7	Treatment Integrity Checklist for Paired Maths Group	238
8.8	Burnett Self- Scale	239
8.9	Burnett Self-Scale Scoring sheet	240
8.10	Behavioural Indicator of Self Esteem Scale	241
8.11	Timeline of Research	242
8.12	Confirmation of Ethical Clearance Letter	243
8.13	Information Letter and Consent Form for School	244

8.14	Information Letter and Consent Form for Parents.....	248
8.15	Information Letter and Consent Form for Participants.....	252
8.16	Information Letter and Consent Form for Intervention Facilitators	255
8.17	Debriefing Letter	258

Contents of Tables

Table 2.1A table displaying the eligibility criteria for systematic literature review	47
Table 2.2: A table displaying the outcomes from systematic search of 3 online data bases using the terms and inclusion criteria specified.	49
Table 2.3A table presenting the weight of evidence of each of the four experiments selected for the systematic literature review.	50
Table 3.1: Research Paradigms (adapted from Mertens, 2014)	70
Table 3.2 A Table presenting the key characteristics of the three schools participating in the current research	78
Table 3.3: Table showing the number of Year 8 participants randomly allocated to each group	82
Table 3.4: Table showing the number of Year 10 participants randomly allocated to each group.....	83
Table 3.5 An overview of the assessments <i>considered</i> to measure maths attainment .	91
Table 3.6: Table illustrating the ‘threats’ to the internal validity of the current study and the attempts to guard against them	98
Table 3.7: Continued....Table illustrating the ‘threats’ to the internal validity of the current study and the attempts to guard against them	99
Table 3.8: Table illustrating the ‘threats’ to the reliability of the study and the attempts to guard against them	102
Table 4.1: A table displaying the means and standard deviations, by group, at pre (T1) and post (T2) testing for Basic Number Screening Test scores for Year 8 participants	111
Table 4.2: A table displaying the results of the Shapiro-Wilk test for Raw BNST Scores at T1.....	114
Table 4.3: A table displaying the result of Levene’s test of equality variances for BNST scores at T2	114

Table 4.4: A table displaying the results of the ANCOVA assessing the impact of Paired Maths on Basic Number Screening Test scores for Year 8 participants.	117
Table 4.5: A table displaying the adjusted means of Basic Number Screening Test Scores in relation to the CV.....	117
Table 4.6: A table displaying the means and standard deviations, by group, at pre (T1) and post (T2) testing for BIOS scores for Year 8 participants	120
Table 4.7: A table displaying the results of the Shapiro-Wilk test for average BIOS scores at T1 for Year 8 participants	123
Table 4.8:A table displaying the result of Levene’s test of equality variances for BIOS scores at T2.	123
Table 4.9: A table displaying the results of the ANCOVA assessing the impact of Paired Maths on BIOS scores for Year 8 participants.....	126
Table 4.10: A table displaying the adjusted means of BIOS Scores in relation to the CV.	126
Table 4.11: A table displaying the means and standard deviations, by group, at pre (T1) and post (T2) testing for Global Self-Esteem Scores for Year 8 participants.....	128
Table 4.12:A table displaying the results of the Shapiro-Wilk test for BSS Global Self Esteem scores at T1.	131
Table 4.13: A table displaying the result of Levene’s test of equality variances for BSS Global Self-Esteem scores at T2.....	131
Table 4.14: A table displaying the results of the ANCOVA assessing the impact of Paired Maths on BSS Global Self Esteem Scores for Year 8 students in the role of Tutee.	134
Table 4.15: A table displaying the adjusted means of BSS Global Self Esteem Scores for Year 8 Participants.	134
Table 4.16: A table displaying the means and standard deviations, by group, at pre (T1) and post (T2) testing for Learning Self-Concept Scores for Year 8 participants.....	136

Table 4.17: A table displaying the results of Levene's test of equality variances for BSS Learning Self-Concept Scores at T2.....	139
Table 4.18: A table displaying the results of the ANCOVA assessing the impact of Paired Maths on BSS Learning Self-Concept scores for Year 8 students in the role of Tutee.	142
Table 4.19: A table displaying the adjusted means of BSS Learning Self Concept Scores for Year 8 Participants.....	142
Table 4.20: A table displaying the means and standard deviations, by group, at pre (T1) and post (T2) testing for Mathematics Self-Concept Scores for Year 8 participants ...	144
Table 4.21: A table displaying the results of the Shapiro-Wilk test for BSS Mathematics Self-Concept scores at T1 For Year 8.	147
Table 4.22: A table displaying the results of Levene's test of equality variances for BSS Mathematics Self-Concept scores for Year 8 at T2.....	147
Table 4.23: A table displaying the results of the ANCOVA assessing the impact of Paired Maths on BSS Mathematics Self-Concept scores for Year 8 students in the role of Tutee.	150
Table 4.24: A table displaying the adjusted means of BSS Mathematics Self-Concept Scores for Year 8 Participants	150
Table 4.25: A table displaying the means and standard deviations, by group, at pre (T1) and post (T2) testing for BIOS scores for Year 10 participants.....	153
Table 4.26: A table displaying the results of the Shapiro-Wilk test for average BIOS scores at T1 for Year 10 participants.	156
Table 4.27: A table displaying the result of Levene's test of equality variances for BIOS scores at T2 for Year 10 participants	156
Table 4.28: A table displaying the results of the ANCOVA assessing the impact of Paired Maths on BIOS scores for Year 10 participants in the role of the tutor	159
Table 4.29: A table displaying the adjusted means of BIOS Scores in relation to the CV	159

Table 4.30: A table displaying the means and standard deviations, by group, at pre (T1) and post (T2) testing for Global Self Esteem Scores for Year 10 participants	161
Table 4.31: A table displaying the results of the Shapiro-Wilk test for BSS Global Self Esteem scores at T1 for Year 10 Participants.....	164
Table 4.32: A table displaying the results of Levene's test of equality variances for BSS Global Self-Esteem scores at T2 for Year 10 participants.	164
Table 4.33: A table displaying the results of the ANCOVA assessing the impact of Paired Maths on BSS Global Self-Esteem Scores for Year 10 students in the role of Tutor. ..	167
Table 4.34: A table displaying the adjusted means of BSS Global Self-Esteem Scores for Year 10 Participants.	167
Table 4.35: A table displaying the means and standard deviations, by group, at pre (T1) and post (T2) testing for Learning Self-Concept Scores for Year 10 participants.....	169
Table 4.36: A table displaying the results of the Shapiro-Wilk test for BSS Learning Self-Concept scores at T1 For Year 10.....	172
Table 4.37: A table displaying the results of Levene's test of equality variances for BSS Learning Self-Concept Scores at T2 for Year 10s.	172
Table 4.38: A table displaying the results of the ANCOVA assessing the impact of Paired Maths on BSS Learning Self-Concept scores for Year 10 students in the role of Tutor.	175
Table 4.39: A table displaying the adjusted means of BSS Learning Self-Concept Scores for Year 10 Participants.....	175
Table 4.40: A table displaying the means and standard deviations, by group, at pre (T1) and post (T2) testing for Mathematics Self-Concept Scores for Year 10 participants .	177
Table 4.41: A table displaying the results of the Shapiro-Wilk test for BSS Mathematics Self-Concept scores at T1 for Year 10.	180
Table 4.42: A table displaying the results of Levene's test of equality variances for BSS Maths Self-Concept scores for Year 10 at T2.....	180

Table 4.43: A table summarising the results by question, by measure, against the null hypotheses.	183
Table 5.1: Table displaying the results for the Treatment Integrity observation, general features.	196
Table 5.2: Table displaying the results for the Treatment Integrity observation, features of PM Process within pairs.....	197

Contents of Figures

Figure 3.1 A figure illustrating the 11 weeks of the research study	74
Figure 3.2: A figure showing the eleven weeks of the research study	75
Figure 4.1: A bar chart displaying the mean raw scores, by group at pre (T1) and post (T2) testing for Basic Number Screening Test Scores for Year 8 Participants	112
Figure 4.2: Figure showing the distribution of Basic Number Screening Test raw scores at T1.....	113
Figure 4.3: A scatter graph displaying the relationship between the DV (T2 BNST scores) and CV (T1 BNST scores), by group.....	115
Figure 4.4: A bar chart displaying the mean, by group at pre (T1) and post (T2) testing for BIOS scores for Year 8 Participants	121
Figure 4.5: Figure showing the distribution of BIOS scores at T1	122
Figure 4.6: A scatter graph displaying the relationship between the DV (T2 BIOS scores) and CV (T1 BIOS scores), by group.....	124
Figure 4.7: A bar chart displaying the mean, by group at pre (T1) and post (T2) testing for BSS Global Self-Esteem Scores for Year 8 Participants	129
Figure 4.8: Figure showing the distribution of BSS Global Self-Esteem scores at T1 ...	130
Figure 4.9:A scatter graph displaying the relationship between the DV (BSS Global Self Esteem Scores) and CV (T1 BSS Global Self Esteem Scores) by group for Year 8 Participants.	132
Figure 4.10: A bar chart displaying the mean, by group at pre (T1) and post (T2) testing for BSS Learning Self-Concept Scores for Year 8 Participants	137
Figure 4.11: Figure showing the distribution of BSS Learning Self-Concept scores at T1	138
Figure 4.12: : A table displaying the results of the Shapiro-Wilk test for BSS Learning Self-Concept scores at T1 For Year 8.	139

Figure 4.13: A scatter graph displaying the relationship between the DV (BSS Learning Self-Concept Scores) and CV (T1 BSS Learning Self-Concept Scores) by group for Year 8 Participants.	140
Figure 4.14: A bar chart displaying the mean, by group at pre (T1) and post (T2) testing for BSS Mathematics Self-Concept scores for Year 8 Participants	145
Figure 4.15: Figure showing the distribution of BSS Maths Self-Concept scores at T1	146
Figure 4.16: A scatter graph displaying the relationship between the DV (BSS Mathematics Self-Concept Scores) and CV (T1 BSS Mathematics Self Concept Scores) by group for Year 8 Participants.	148
Figure 4.17: A bar chart displaying the mean, by group at pre (T1) and post (T2) testing for BIOS Scores for Year 10 Participants	154
Figure 4.18: Figure showing the distribution of BIOS raw scores at T1 for Year 10 participants	155
Figure 4.19: A scatter graph displaying the relationship between the DV (T2 BIOS scores) and CV (T1 BIOS scores), by group.	157
Figure 4.20: A bar chart displaying the mean, by group at pre (T1) and post (T2) testing for BSS Global Self-Esteem Scores for Year 10 Participants	162
Figure 4.21: Figure showing the distribution of BSS Global Self-Esteem scores at T1 for Year 10 participants	163
Figure 4.22: A scatter graph displaying the relationship between the DV (T2 BSS Global Self Esteem Scores) and CV (T1 BSS Global Self Esteem Scores) by group for Year 10 Participants.	165
Figure 4.23: A bar chart displaying the mean, by group at pre (T1) and post (T2) testing for BSS Learning Self-Concept Scores for Year 10 Participants	170
Figure 4.24: Figure showing the distribution of BSS Learning Self-Concept scores at T1 for year 10 participants.....	171

Figure 4.25: A scatter graph displaying the relationship between the DV (T2 BSS Learning Self-Concept Scores) and CV (T1 BSS Learning Self-Concept Scores) by group for Year 10 Participants.....	173
Figure 4.26: A bar chart displaying the mean, by group at pre (T1) and post (T2) testing for BSS Mathematics Self Concept scores for Year 10 Participants.....	178
Figure 4.27: Figure showing the distribution of BSS Maths Self-Concept scores at T1 for year 10 participants	179

Abstract

This study presents an exploratory evaluation of a seven week cross-age peer-tutoring mathematics intervention, Paired Maths (based on Topping, Kearney, McGee and Pugh, 2004), as delivered in a UK secondary school. The impact of Paired Maths on the mathematics attainment of tutees, and the self-esteem of both tutees and tutors is evaluated.

Paired Maths involves pairs of students, one student in the role of a tutor and one in the role of a tutee, completing mathematics activities together. Tutors are trained to follow a structured process to scaffold their tutee's learning during the activities. Although existing theory and literature suggests that a cross-age, peer-tutoring intervention around mathematics *could* have a positive impact, no research has experimentally evaluated Paired Maths in the UK secondary school to date.

A total of 78 participants, aged 12-15 years, from three secondary schools situated in a small city in England took part. A Randomised Control Trial was used to consider the impact for Year 8 participants in the role of the Tutee. This compared children randomly assigned to either a Paired Maths (Experimental) Group, a Teaching Assistant Led Comparison Group or a Waitlist Control Group. Results showed no statistically significant differences between groups in mathematics attainment or self-esteem.

Within the same exploratory evaluation, a Randomised Control Trial was also employed to consider the impact for the Year 10 students in the role of the tutor. This compared children randomly assigned to either the Paired Maths (Experimental) group or a Waitlist Control Group. Although Waitlist Control participants saw a statistically significant decrease on measures of Maths self-concept over time, results showed no other statistically significant differences between groups in self-esteem.

Overall, Paired Maths was not considered to have had a positive impact on mathematics attainment for children in the role of the tutee, or on the self-esteem of children either in the role of tutee or tutor. These findings are considered in relation to the literature and methodology adopted, with a number of possible explanations presented. Possible avenues for future research are proposed (within and also beyond

the post-positivist paradigm) with the key conclusion that further exploration is needed around what may support the successful implementation of a Paired Maths programme with *secondary* aged Pupils. Key limitations of the research relate to the measurement of the dependent variables and treatment fidelity.

1 Introduction

Mathematics difficulties are often considered to be overshadowed by literacy difficulties in education (Gersten, Clarke, and Mazzocco, 2007). However, over recent years both the high prevalence and long term detrimental outcomes of poor mathematics/numeracy skills are becoming well documented (Every Child a Chance Trust, 2008). Indeed, with 1 in 12 of England's 15-year-olds found to be failing even to reach "basic levels" in mathematics (PISA, 2012) the need for schools to have access to a clear evidence base of effective mathematics interventions in the UK seems imperative.

Greatest concern around general progress in schools has been expressed for secondary aged pupils (Ofsted and DfE, 2014). While primary schools now seem to be improving outcomes for their students (eighty two per cent are now good or outstanding) improvement in secondary schools seems to have come to a standstill (Ofsted and DfE, 2014). More specifically, in 2012 the Department for Education concluded that there is very little evidence on any mathematics interventions for secondary aged pupils (DfE, 2012).

Many researchers have also called for more research around school-based interventions that promote both academic *and* socioemotional outcomes for students, particularly self-esteem (Durlak, 1995). This may not be surprising given the well-documented potential negative effects of low-self esteem for both children and adults across many areas of life (Baumeister, Campbell, Krueger and Vohs, 2003).

Recently, within the Local Authority where the researcher works as a Trainee Educational Psychologist, a piece of research was conducted assessing the impact of a cross-age peer-tutoring intervention around mathematics: Paired Maths. Paired Maths (Topping, Kearney, McGee and Pugh, 2004) involves pairs of students, one student in the role of a tutor and one in the role of a tutee, completing mathematics activities together. Tutors are trained prior to the intervention to follow a structured process to scaffold their tutees learning during the activities. This research was conducted across three primary schools, and it was found that the intervention led to positive gains in

knowledge of times tables for children taking the role of the tutee in the intervention, and also positive gains in self-esteem for both tutees and tutors when compared to comparison and control groups. When the researcher learnt of the positive results from this study, they considered given the context around: the prevalence and outcomes of mathematics difficulties; the lack of progress in secondary schools; the lack of research with this population and the need for interventions which can address both academic and socio-emotional outcomes, that it may be useful to evaluate the effectiveness of the Paired Maths intervention with secondary aged pupils. The Principal Educational Psychologist welcomed this suggestion, as the service was keen to seek evidence around whether it would be worthwhile to develop and offer the training currently offered to primary schools around Paired Maths to secondary schools.

The current study therefore, aims to evaluate Paired Maths with secondary aged pupils, particularly considering its impact on mathematics attainment and self-esteem.

Previously published evaluations of cross-age peer-tutoring interventions with mathematics have been less robust, with only a handful of studies including a control group for comparison, and with the majority occurring in the US with primary aged pupils.

The key unique contributions of this study therefore, are as follows:

- It presents the implementation of a replicable 7 week Paired Maths intervention and research design
- It involves Randomised Control Trials, which are scarce in the cross-age mathematics peer-tutoring literature
- It evaluates the efficacy of Paired Maths in the UK school context
- It evaluates the efficacy of Paired Maths when implemented with secondary aged pupils
- It evaluates the impact of Paired Maths on self-esteem for both the children in the role of tutee and tutor, as well as on mathematics attainment for children in role of the tutee

In the following chapter, the literature review (chapter 2) aims to provide a review of the current context and the psychological theory and research underpinning the study. The methodology (chapter 3), results (chapter 4), discussion (chapter 5) and overall conclusions (chapter 6) are presented in subsequent sections.

2 Literature Review

2.1 Mathematics

2.1.1 Mathematics performance in the UK

In a worldwide comparison of student performance in 2012, The Programme for International Student Assessment (PISA) found that for mathematics, the UK was ranked only 25th in the world for its performance. This was despite the UK reportedly spending more on education than the average country. Further, 1 in 12 of England's 15-year-olds were found to be failing even to reach "basic levels" in mathematics-there having been a similar picture over the past 6 years (PISA, 2012).

Indeed, raising attainment in mathematics seems to have been a problem in the UK for some time. For example, in 1982, Cockcroft reported that an average class of 11-year olds in England was, at that time, likely to contain the equivalent of a seven-year range in arithmetical ability. Since this time, although the UK has seen developments such as the introduction of the National Curriculum (Department for Education, 2013) and also the National Numeracy Strategy, (Department for Education and Employment, 1999) studies suggest that there has been little change to this picture, with about 6% of children each year still failing to reach expected levels in mathematics at Key Stage 2 (Gross, 2007; Burr, 2008).

2.1.2 Long-term outcomes for poor mathematics/numeracy

In 2006, Parsons and Bynner analysed data from two large longitudinal studies, the National Child Development Study (1958) and the British Cohort Study (1970) in order to investigate the impact of poor numeracy (the ability to use mathematics in everyday life (National Numeracy, 2015)) and literacy on life outcomes. The overall sample included 17,000 participants. Results were analysed separately for both men and women, and indicated that men who had poor numeracy and literacy skills were: more likely to be unemployed or in semi-skilled/unskilled jobs, less likely to use a computer at work, received less training and also less likely to own a home. Further, men with poor numeracy, irrespective of their literacy were: less likely to be in a pension

scheme, more at risk of depression, had little interest in politics and were more likely to have been suspended from school or cautioned by the police.

When considering the results for women, the impact was even larger, where women with poor literacy and numeracy were: less likely to receive work-related training, less likely to own their own home and more likely to suffer depression. Further, women with poor numeracy, irrespective of their literacy, were: less likely to be in full-time employment, had little interest in politics and were more likely to be home-carers or in semi-skilled/unskilled jobs or be in non-working households. They were also more likely to have reported poor health in the previous 12 months, have low self-esteem and feel that they lacked control over their lives (Parsons and Bynner, 2006).

From these quite staggering results, it was concluded that poor literacy does impact on life chances, however, when this is combined with poor numeracy, the impact is compounded (Parsons and Bynner, 2006).

2.2 Is mathematics shadowed by literacy?

It is thought at least 15% to 20% of the population are likely to have difficulties in mathematics which are sufficient enough to cause considerable practical and educational problems (Every Child a Chance Trust, 2008). However, despite the seemingly significant prevalence of difficulties, compared to literacy, numeracy/mathematics is still a relatively neglected area of study. For example, when Gersten, Clarke, and Mazzocco reviewed the ERIC literature base in 2007, they found that the ratio of studies on reading disabilities to mathematics disabilities and difficulties was 5:1 for the decade 1996–2005 -although, this was a significant improvement on the ratio of 16:1 seen in the research of the previous decade.

National Numeracy (an independent UK-based charity run by numeracy experts) suggest a number of reasons why it could be the case that numeracy/mathematics seem to have been "hidden" by literacy including: the role literacy plays in accessing the curriculum, the emotional pull of reading for pleasure, the poor image of maths where it is not seen to be "cool", parental difficulties and reluctance to engage with it, feelings people develop when they are unable to 'do' numeracy (including anxiety) and not knowing what to do to improve their skills (National Numeracy, 2015).

Regardless of the possible reasons why numeracy/mathematics may have been slightly 'left behind', the well documented (and concerning) life outcomes for those with poor numeracy/ mathematics skills suggest intervention should be considered an important focus for educational research.

2.2.1 Intervention for mathematics difficulties

In a research report for the Department of Education and Skills (DfES) in 2004, Dowker carried out an extensive review of the international literature on mathematical interventions, including books and journals in Oxford University libraries, major education and social science databases, the proceedings of conferences of major relevant organisations, a few unpublished documents involving action research in schools and local authorities, and work in progress within the DfES (Dowker, 2004). From this review, Dowker concluded that pupils' difficulties with mathematics are, "highly susceptible to intervention" (p.42) and that interventions can benefit pupils experiencing difficulties.

Dowker (2004) highlighted that before specific interventions are considered, good quality teaching at a whole class/universal level is imperative as this "may prevent some mathematical 'difficulties' from ever becoming apparent; and many mathematical difficulties are undoubtedly mainly the result of limited or inappropriate teaching". (pg.43). The researcher first considers what the literature suggests are the key psychological theories underpinning mathematics learning and what research suggests may constitute effective teaching of mathematics, before then considering, more specifically, interventions for children who may experience mathematics difficulties.

2.2.2 Psychological theories underpinning early mathematics Learning

There are a number of general theories around learning which have implications for the learning of mathematics, as well as some more specific theories around the learning of different mathematical areas. Due to the scope of this literature review, the researcher outlines two learning theories generally highlighted across the literature to have been influential in how mathematics learning has been conceptualised.

2.2.2.1 Piaget's Theory of Cognitive Development

Using a series of experiments with a small number of children (mainly his own) looking at their perception and understanding, Piaget developed his Theory of Cognitive Development (works spanning from 1926-1985). Though much of Piaget's work has been challenged over the years (particularly around the context and number of experiments the theory is based upon e.g. Donaldson, 1978) his stages of learning have been quite influential in how the development of learning, and also specifically mathematics learning is viewed. He argued that all learning takes place through the processes of assimilation (taking in new information and incorporating it into our existing knowledge/ideas) and through accommodation (altering existing cognitive structures as a result of new information or new experiences) , with the two processes interacting throughout life in different ways, according to the stage of mental development (Piaget and Inhelder, 1962).

These 4 universal stages of development include:

- Sensorimotor (0-2 years) - Here children begin to use imitation, memory and thought. They develop object permanence (where they understand that something does not cease to exist just because it cannot be seen), the beginnings of symbolic knowledge and goal-directed behaviours.
- Preoperational (2-7 years) - Language and the ability to think symbolically further develop here. This stage is also characterised by egocentrism (an inability to see things from other's perspectives) a difficulty thinking beyond a one-way logic and they are unable to conserve number.
- Concrete operational (7-11 years) -Here, children are now able to solve concrete "hands-on" problems. They also understand, conservation, classification, seriation and reversibility.
- Formal operational (11-15 years) - Children are now able to use abstract thinking and scientific reasoning to logically solve problems. They also become more focussed on their own identity and social issues. (Piaget and Inhelder, 1962)

The key practical implications for the learning of mathematics are as follows:

- Teacher's should provide an environment for learning based on the idea that action and interaction with others and the environment is key to acquiring new mathematical concepts.
- They also need to consider how 'ready' a child is to take on a certain level/ type of activity, given that children learn at fixed stages and ages, which to a certain extent, no amount of teaching can accelerate. (Ojose, 2008)

2.2.2.2 Vygotky's Social- Cultural Theory of Development

Vygotky's theory of development is underpinned by the concept of social constructivism, where language and shared understandings are key to constructing knowledge and meaning (Vygotky, 1978). Vygotky proposed that true learning can only take place when it occurs within the zone of proximal development. This can be described as the difference between what a learner can do on their own, and what they can do when supported by more skilled peers. When cognitive tasks are placed in this zone, and a more skilled learner and less skilled learner can cooperate together, it is argued they can construct a shared and meaningful understanding (Vygotky, 1978).

When considering mathematics specifically, Vygotky viewed learning as the development of thinking and reasoning. He suggested that mathematical concepts are acquired *externally* through conversation or action before then gradually become *internalised* into thought (Vygotky, 1978)

In terms of the practical implications for mathematics learning, Vygotky's theory indicates the following:

- Children can learn effectively from a teacher, but also from one another.
- Mathematics activities should be meaningful to the child.
- Discussion is key to the internalisation of mathematical concepts.

2.2.3 Effective teaching of mathematics

A Seminal piece of research in the area of effective teaching in mathematics/numeracy, which is often quoted widely within mathematics education

research comes from Askew, Brown, Rhodes, Wiliam and Johnson(1997). They identified key principles of effective teaching of numeracy by looking at mean test scores of students over time who attended a range of schools across the UK. It was a large scale study, where data were collected from over 2000 pupils, with evidence gathered from 90 teachers in total.

The key principles of effective instruction identified by the research, which have been replicated in a number of other studies(e.g. Reynolds & Muijs, 1999;Muir, 2008),are:

- Using teaching approaches that connected different areas of mathematics and different ideas in the same area of mathematics using a variety of words, symbols and diagrams
- Using pupils' own descriptions of their methods and their reasoning to help establish connections and to address misconceptions
- Challenging pupils to think through explaining, listening, and problem solving
- Emphasizing the importance of using a variety of methods of calculation with a particular focus on mental computation
- Ensuring that all pupils are sufficiently challenged
- Encouraging purposeful discussion

This is termed a 'connectionist' approach to teaching mathematics and is closely aligned to constructivist theories of learning (such as Vygotsky). A learner constructs their own knowledge, by selecting information perceived to be relevant and interpreting that information in relation to their existing knowledge (Shuell, 1996). Therefore anyone who is supporting the learning of another will explicitly recognise and work on that individual's misunderstandings (Askew et al., 1997).

2.2.4 Intervention for mathematics difficulties, what works?

From the review carried out in 2004, Dowker was able to draw a number of conclusions around what elements or features make mathematics intervention for children who are experiencing difficulties in mathematics most effective. The key findings were as follows:

- Interventions led by Teaching Assistants (TAs) appear to be more effective when they have had adequate training and support.
- When considering older children and adolescents, intervention around the use of formal operations can have a positive impact on their mathematical development.
- Research indicates that interventions that include training around meta-cognitive skills can have a positive impact on maths difficulties
- Interventions do not need to be very time-consuming or intensive to be effective.
- Considering research that indicates mathematic difficulties can lead to 'mathematics anxiety', it was suggested interventions should take place as early as possible to reduce the likelihood of this anxiety developing and further inhibiting progress.
- For children who struggle in mathematics, interventions which have been informed by a detailed assessment of their learning needs and thus are individualised to their areas of weakness, work best.
- Peer learning and collaborative approaches, such as peer-tutoring and co-operative learning, can promote positive gains children with maths difficulties. However, particularly when considering those with more severe needs, this should not completely replace adult intervention.

This report was later updated in 2009, though with a particular focus on earlier intervention in the primary stages (Dowker, 2009). In this review, Dowker concluded that the evidence base around interventions for numeracy is 'patchy', and just as was highlighted earlier, is much less developed than the evidence base on literacy interventions. However, on a more positive note, it was considered that interventions around numeracy and mathematics are being more widely used across schools in the UK. Further, schools consider that they are seeing positive effects from these interventions, both in achievement and attitudes towards mathematics. Moreover, Dowker highlighted that there is a growing research base, which is using pre- and post-measures, indicating positive effects on achievement. So, it appears that some schools are choosing to use interventions around mathematics, and some people are 'getting it right'. However, it seems that more evidence is needed around the comparative

effectiveness of different types of intervention for numeracy and mathematics difficulties.

Finally, in collaboration with the Education Endowment Fund, the Sutton Trust have put together the "Teaching and Learning Toolkit" (Higgins et al, 2013) which is an accessible summary of educational research providing guidance for teachers and schools on how to use their resources to improve the attainment of disadvantaged pupils, including mathematics. The Toolkit covers 34 areas of intervention, each presented in terms of their average impact on attainment, the strength of the research evidence supporting them and their cost. One intervention which has been classed as having one of the highest impacts (6 months) for the lowest cost, and also has a high rating for the quality of research evidence behind it, is peer-tutoring: the intervention that is the focus for the current research.

2.3 Peer- tutoring

2.3.1 Definitions

Peer tutoring is one of a number of interventions that come under the umbrella term of "Peer learning". Peer learning has been defined as "the acquisition of knowledge and skill through active helping and supporting among status equals or matched companions" (Topping and Ehly 1998, Topping 2005). It involves situations where children or adults learn with or from their peers, rather than professional teachers, who are from similar social groupings (Topping, 2005). Within peer learning, usually individuals take on either a learner or 'tutee' role, or a teaching or 'tutor' role (Topping 2005).

Peer learning can take a number of forms. Indeed, Topping (2005) identified that peer learning can actually vary across 13 different organisational dimensions including:

- Curriculum content/ subject area - including knowledge and or skills
- Constellation of the contact - it can occur in pairs, in or in larger groups, with different numbers of individuals in the role of the "helper" or "tutor".
- Within or between institutions - children from the same or different schools coming to work together, for example.

- Year of study - individuals can be from the same or different years of study, and/or be the same or different ages (often referred to as cross-age).
- Ability - individuals can be of similar or different ability.
- Role continuity – the role as the tutor can be fixed, or there can be structured switching of roles between being the tutor and tutee at strategic moments (this is often labelled Reciprocal).
- When – it can either be supplementary to or in substitution of regular teaching, or it can occur after school hours.
- Place - for example within or outside of the class-room.
- Tutor characteristics - level of Special Educational Needs for example.
- Characteristics of the Tutee (as above)
- Objectives - the activities may target areas such as academic achievement, affective, attitudinal, social and emotional and organisational gains.
- Voluntary or compulsory- it could be a programme which children can sign up to, or a compulsory part of their curriculum.
- Reinforcement - some peer learning may include extrinsic reinforcement, whereas some focus on intrinsic motivation. (Topping, 2005)

Peer-tutoring is just one of many of these possible peer learning approaches. It involves individuals working in pairs where there is specific role-taking as either the tutor (some may think of this as the teacher) or the tutee (or the 'learner'). It focuses on the pair working together to develop skills and knowledge with regard to specific curriculum/academic subjects. It commonly involves training for the tutor around explicit procedures for interaction with their tutee. There are also different configurations within peer-tutoring, including: class wide pairings (where all children within one class are paired up), reciprocal relationships (where both in the pair take it in turn to be the tutor) and same age or cross-age pairings. (Topping, 2005).

Sometimes there is also confusion in the literature between peer "mentoring" and peer-tutoring (Topping 2005). However, mentoring always involves a fixed role where one individual supports another and there is often exploration of aspirations, positive

role modelling, rewards, joint problem solving and counselling techniques and commonly targets disadvantaged groups (Topping 2005).

2.3.2 Historical and theoretical underpinnings of peer-tutoring

Some have traced the origins of peer-tutoring further back than one might expect. For example, Jenkins and Jenkins (1982) argue that "Tutorial instruction (parents teaching their offspring how to make a fire and to hunt and adolescents instructing younger siblings about edible berries and roots) was probably the first pedagogy among primitive societies" (1982, p. 64). Some argue that more formalised peer-tutoring approaches can be first seen in the "Monitorial System" of the early nineteenth century (Bland and Harris 1989), whereas Topping (1988) dates it back to the 1700s. However, it was not until the last several decades that peer-tutoring, along with other peer-assisted learning approaches, gained attention in the scientific literature. Although the focus was initially on the individuals receiving the tutoring (tutees), the potential benefits for both the tutors and the tutees have now been explored quite extensively (Topping, 2005).

Although it appears that peer-tutoring originally stemmed mostly from practice and ancient tradition, authors have also since considered what psychological theories may underpin the approach. Before turning to the literature considering the effectiveness of peer-tutoring approaches, the key underpinning theories will be considered.

2.3.2.1 Piaget's Theory of Cognitive Development

It is argued that peer-tutoring involves a high level of cognitive conflict and challenge for both tutors and tutees, in the tutors trying to explain their existing knowledge to someone else, and tutees trying to take this into their own knowledge (Slavin, 1996; Maheady and Gard 2010; Topping, 2005). Indeed, in Piaget's Theory of Cognitive development, learning is seen as an outcome of interactions which allow cognitive perspectives to be explored. The resulting "sociocognitive conflict" that comes from this is said to lead individuals to 'elaborate' upon and "reconstruct" their own existing knowledge and thinking processes, where connections between new and prior perspectives are integrated or added to through accommodation and

assimilation (as described earlier) (Foot & Howe, 1998; Topping & Ehly, 1998; Slavin, 1996; O'Donnell & O'Kelly, 1994).

2.3.2.2 Vygotky's Social- Cultural Theory of Development

As highlighted earlier, Vygotky's socio-cultural theory of development emphasises the importance of social interactions on development (Vygotky, 1978). It also proposes that true learning can only take place within zone of proximal development, and when cognitive tasks are placed in this zone, and a more skilled learner and less skilled learner can cooperate together it is argued they can construct a shared and meaningful understanding. It could be argued, then, that peer-tutoring should be successful because the more skilled tutor supports the less skilled tutee, and they discuss and come to a shared understanding of how to approach and solve a mathematics activity which is of appropriate difficulty -they work within the Tutee's zone of proximal development.

2.3.2.3 Role Theory

According to Role Theory, the way an individual behaves is at least partially determined by the role they are given to play in a situation (Thomas & Biddle, 1979). It argues that when an individual assumes a role, that individual will be likely to also take on the attitudes and self-perceptions of that role as well as the behaviours, and that this can actually 'spillover' into other situations (Turner, 2002). It is argued that this 'spill over' might point to positive effects from peer-tutoring outside of the peer-tutoring sessions. For tutors, their role involves needing to be competent and have an enthusiastic and positive attitude. Whereas in the role of the tutees it is argued that because they are in a one-on-one instructional settings where there is less anonymity and more immediate responsibility is placed on them to demonstrate their learning, they are more likely exhibit behaviours associated with a "good" student. So, role theory argues that these behaviours, for both tutee and tutor may 'spill over' into other learning situations.

2.3.3 Peer-tutoring and links to effective mathematics instruction/intervention

Peer-tutoring arguably encompasses many of the features highlighted in the preceding literature to be important for effective intervention for mathematics difficulties and for the learning of mathematics more generally.

Firstly, one could argue that peer-tutoring encompasses all the of principles of the optimum teaching approach for mathematics—as identified by Askew and Colleagues(1997)-the 'connectionist' approach :

- It allows for the connection of different areas of mathematics and different ideas in the same area of mathematics using a variety of methods;
- It uses pupils' own descriptions of their methods and their reasoning to help establish connections and to address misconceptions;
- It challenges pupils to think through their explaining, listening, and problem solving;
- It emphasises the importance of using a variety of methods of calculation;
- It is individualised to the pupil, ensuring they are appropriately challenged;
- And finally it encourages purposeful discussion.

Reflecting on the reviews from Dowker (2004;2009) peer-tutoring also involves some of the key aspects identified for effective intervention for children experiencing mathematics difficulties:

- It involves aspects around meta-cognition;
- It involves peer support;
- It involves individualised one-one time;
- It often uses multi-sensory and visual activities;
- And it can be used to supplement rather than be used instead of classroom teaching.

It is also underpinned by two of the major theories which have been influential when considering how children learn generally and also more specifically how they learn mathematics: Piaget's Theory of Cognitive Development and Vygotsky's Social- Cultural Theory of Development.

2.3.4 The evidence base for peer-tutoring

The literature surrounding the effects of peer-tutoring is wide and varied and spans across more than 40 years. It investigates practice in a range of subject areas, and employing many different types of peer-tutoring that vary across a number of the organisational dimensions highlighted by Topping (2005). It considers a variety of outcomes, both academic and psychological, and for both the tutee and the tutor in the peer-tutoring dyad. The researcher will endeavour to review the key findings in the research literature covering these different areas.

A number of meta-analyses have been conducted in the area of peer-tutoring over the years and these will form the basis of the discussion with the individual studies that seem particularly pertinent highlighted.

2.3.4.1 Academic achievement for tutees

The first review looking at the effects of peer- tutoring on academic achievement emerged as early as 1982. Cohen, Kulik and Kulik (1982) completed a review of 65 studies of maths and reading tutoring programs across primary and secondary age groups. They found almost 87% of the studies that measured academic achievement showed that students who received tutoring made significantly more academic achievement gains than students who did not receive tutoring, with an average effect size of 0.4. However, this review does have a number of limitations. Firstly, it is now quite dated as the majority of the studies included occurred before 1980. Second, not all of the tutoring provided was *peer*-tutoring: it also included a small number of teacher-led and other adult-led tutoring, which could have confounded results. Third, it did not take into account how small sample sizes affect effect sizes (Leung 2015).

Britz, Dixon and McLaughlin(1989) reviewed 14 studies around peer-tutoring and mathematics that had been conducted after the review from Cohen and Colleagues (1982). They found that only one of the studies failed to find increase in academic performance for tutees.

Later, when Rohrbeck and colleagues reviewed a large number of studies using group-comparison designs in 2003, they found a positive impact on academic outcomes for

tutees, with an even larger average effect size of 0.59. However, this review included a number of different peer assisted learning techniques, not just peer-tutoring, and so results may be confounded.

One of the most recent meta-analyses considering academic achievement, conducted in 2013 by Bowman-Perrot and Colleagues, focussed on studies involving single case research experiments (where participants serve as their own control). 26 studies, involving 938 students in Grades 1–12 involved in peer-tutoring across a range of subjects were analysed. Again, a significantly positive impact on tutee's academic achievement was found, here with a much larger average effect size of 0.75, suggesting that peer-tutoring can make quite a substantial impact. However, it is important to note that single case experimental designs can arguably lack external validity and thus caution is needed when generalising their findings.

The review the author considers to be most relevant, rigorous and comprehensive, however-and which is also the most recent- comes from Leung (2015). This meta-analysis used the most advanced methodological techniques (procedures that account for small samples for example), focused just on peer-tutoring (so was not confounded by other modes of peer learning), looked at a wide range of participants and subject areas and only included studies that involved experimental group designs (quasi experimental or true experimental). Results from the 72 samples analysed indicated a significantly more positive impact on academic achievement for students who received peer-tutoring, with an unweighted mean effect size of 0.59 ($SD = 0.58$).

When considering whether this positive impact on academic achievement for tutees is similar across different subjects, the literature is a little less clear. Cohen, Kulik and Kulik. (1982) reported greater effect sizes for mathematics compared to other subjects, particularly compared to reading. However, Rohrbeck and colleagues (2003) found no significant differences between the impact on maths and reading. This was echoed in the recent meta-analysis from Leung (2015) where similar effects were seen in interventions around mathematics and English/reading, but it was also found that physical education and arts actually saw the largest impacts.

Leung (2015) also looked at the differential effects for different aspects of mathematics. It was found that peer-tutoring interventions that focused on more basic operation/ computation (e.g., subtraction, addition, multiplication) saw greater impact on achievement than interventions that looked at conglomerate skills or application of concepts (e.g. solve story problems or graphic scenarios).

Some studies have also found that peer-tutoring can lead to greater improvements in academic achievement compared to other forms of intervention, such as small group work (with a teacher). For example, Ehri, Dreyer, Flugman, and Gross (2007) found that students who were struggling to read and receiving peer-tutoring showed greater improvements in their reading compared with students who received small group instruction. It was suggested that this is because the tutoring allowed more focussed and individualised reading practice, where more informative feedback could be given.

In summary, a number of important and reliable reviews suggest that research has generally supported the idea that peer-tutoring can have a significant positive effect on the academic achievement of students. Although the extent of this positive effect appears to differ depending on the content or subject area of the teaching, an overall positive impact is observable.

2.3.4.2 Academic achievement for tutors

Another key element to highlight is that much of the existing research into peer-tutoring and academic achievement (including many of the studies discussed in the previous section) also indicates that there can be positive effects *for tutors* as well.

Although a relatively small portion of studies have assessed academic outcomes for tutors, both Cohen, Kulik and Kulik (1982) and Britz and colleagues (1989) reported on studies where tutors' academic achievement was considered, and in both cases indicated that positive improvements had been found. Indeed some have argued findings indicate that individuals who *tutor* in a subject are more likely to have higher order conceptual understanding of that subject than students who only *study* the subject (Annis, 1983; Benware and Deci, 1984) - though studies in this area have tended to draw on examples of tutoring in higher education contexts.

2.3.4.3 Other outcomes associated with school performance

As well as tangible raised achievement in the target subject, some studies have found that peer-tutoring interventions can have a positive impact on other learning skills and behaviours that arguably enhance the possibility of success at school.

Peer-tutoring has been found to improve on task-time behaviour in class, for both tutees and tutors (Ginsburg-Block, Rohrbeck and Fantuzzo 2006; Ginsburg-Block & Fantuzzo, 1997). For example, In Greer and Polirstok's studies from 1982 and 1986, they found that cross-age peer- tutoring in reading led to an increase in on task behaviours in other subject classes as well as increases in reading scores. This seems particularly important, as some have suggested that the amount of time spent on task can be a useful predictor of academic achievement (Greer & Polirstok, 1982).

Increased attendance rates, and a positive effects on retention rates in older students have also been associated with peer-tutoring (e.g. (House & Wohlt, 1991; Cohen et.al, 1982; Ginsburg-Block, Rohrbeck and Fantuzzo 2006). This appears to be particularly prominent in *tutors*. For example, Nazzal (2002) found that 8th grade tutors in mathematics peer-tutoring programme were absent significantly fewer days during the tutoring program than they were before the tutoring program began (though this particular programme only lasted one week). An increase in attendance seems to be an important potential effect of peer-tutoring, as this could mean that a student will have increased learning opportunities across all subjects, not just where they are being tutored/ are tutoring (Nazzal, 2002). This could also explain the finding described earlier that tutors can see academic improvements in other subject areas.

Another final, school-related outcome area that has been positively associated with peer-tutoring is 'attitudes'. A number of studies have shown that peer-tutoring can, for both tutees and tutors, have a positive effect on attitudes towards both the subject matter being focussed on (Cohen et al.,1982; Brush, 1997; Fogarty & Wang, 1982; Ponzio & Peterson, 1999) and also school life in general. For example, Fantuzzo and his colleagues (Fantuzzo, Riggio, Connely, & Dimeff, 1989) found that students who were involved in a peer-tutoring programme had higher levels of school satisfaction than students who studied alone.

2.3.4.4 Socio-emotional outcomes

Many researchers have argued that there needs to be more school-based interventions that promote both academic *and* socioemotional outcomes for students, particularly self-esteem (Durlak, 1995). However, findings from earlier reviews regarding the impact of peer- tutoring on the self-esteem of tutees and tutors were a little mixed. For example, Cohen and colleagues (1982) in their review found that studies showed only a very small effect on self-esteem in peer tutoring interventions. Further, although Rohrbeck and colleagues (2003), when considering mathematics peer-tutoring, found no studies that showed a decrease in self-esteem, they did not find any that reported significant *increases* in self-esteem for either the tutees or tutors (e.g. Bar-Eli & Raviv, 1982; Sharpley, Irvine and Sharpley, 1983). However, there have been a number of studies that have indeed found that being involved in a tutoring program increases self-esteem, and other socio-emotional outcomes such as social confidence, and perceived academic efficacy in both tutees and tutors (Fantuzzo, Davis, & Ginsburg, 1995; Ginsburg-Block & Fantuzzo, 1998; Heller & Fantuzzo, 1993).

In a more recent meta-analysis, which focussed particularly on social, self-concept (or self-esteem) and behavioural outcomes, Ginsburg-Block, Rohrbeck and Fantuzzo (2006) found that individuals who are involved in peer assisted learning interventions see significantly larger increases in self-esteem (though they term it *self-concept*) than children who are not, and that these are "moderate" positive effects. They also found a positive impact when considering outcomes such as social skills and lower levels of challenging behaviour. However, it is important to note that this meta-analysis did consider a number of different peer assisted learning techniques, not just peer-tutoring, so results may be confounded by other elements of these interventions.

2.3.4.5 Program characteristics related to outcomes

As mentioned earlier, peer-tutoring interventions can vary along a number of different aspects, including their structure and duration, the makeup of the pair (same-age or cross-age), the amount of training given to tutors and who the intervention targets. There is a wealth of research suggesting that there are particular choices around these

variables that can affect the possibility of seeing the positive outcomes described in the above section.

2.3.4.6 Amount and duration of tutoring

Cohen and colleagues (1982) found from their meta-analysis that longer tutoring programs did not necessarily lead to greater improvements in academic outcomes. They actually found that longer programs (19-36 weeks) can lead to less academic improvement than shorter programs (less than 4 weeks). This was echoed in the more recent, rigorous meta-analysis from Leung (2015), who reported that a shorter duration (less than 10 weeks) of the entire tutoring period displayed larger effect sizes than those with a longer duration (more than 10 weeks). Rohrbeck and colleagues (2003), considered how 'dosage' rather than duration impacts on the levels of academic progress with peer-assisted learning strategies (again, note the need for caution around possible confounding variables). Dosage was calculated by duration x sessions per week x hours per session, and they concluded this did not have a significant impact on academic achievement.

When considering the frequency and length of individual sessions, Leung, Marsh, and Craven (2005) found that peer-tutoring interventions involving more (than three) sessions per week were more effective for than fewer sessions. They also found that sessions that were shorter in length were more effective.

Taken together, the research suggest that in terms of time and length, 'optimum' peer-tutoring interventions should be less than 10 weeks long and involve frequent (3 or more times a week) but short sessions. These findings fit with the conclusion from Dowker (2004) that mathematics interventions do not need to be long or intense to be effective.

2.3.4.7 Structuring of the peer-tutoring sessions

Within the vast number of peer-tutoring interventions, there has been much variation in how the individual sessions are structured. Some interventions are highly structured, with set role behaviours and processes to follow, whereas others barely involve any structure at all.

The existing research seems to suggest that interventions that are more structured are more beneficial. For example, the reviews from Cohen and colleagues (1982) and Leung (2015) both showed larger academic achievement outcome effect sizes for structured peer-tutoring interventions compared with unstructured tutoring. Ginsburg-Block, Rohrbeck, and Fantuzzo (2006) found a similar pattern when they reviewed outcomes relating to self-concept (or self-esteem), social skills and behaviour, where tutoring that involved more structured roles led to more positive outcomes.

However, some argue that although having a structured format to peer-tutoring is important, it is also important that room remains for some student creativity, in order to inspire intrinsic motivation, curiosity, and self-worth (Fantuzzo, King, & Heller, 1992; Koestner, Ryan, Bernieri, & Holt, 1984). Indeed, in Rohrbeck and colleagues' (2003) large review in the United States, they found that a structure that allowed some autonomy was associated with larger effect sizes for academic outcomes. Again, though, this review included other peer assisted learning approaches.

2.3.4.8 Type and amount of tutor training

As the use of more structured approaches seem to lead to more peer-tutoring interventions, it seems sensible to consider the training tutors receive. Again, the literature includes examples of peer-tutoring interventions that vary in the amount and type of training given.

In Cohen and colleagues' early meta-analysis (1982) it was actually found that interventions where tutors had received training did not consistently show larger effects on academic performance than those where training had not occurred. Britz and colleagues (1989) also found quite ambiguous results when they reviewed whether tutee performance was mediated by level of tutor training. However, as noted earlier, both of these reviews included possibly confounding study characteristics. Further, the only parameter around tutor training examined in the meta-analysis from Leung (2015) was the impact on the number of weekly training sessions a tutored received. Fewer weekly training sessions of tutor training produced greater effect sizes than did more frequent training sessions (Leung, 2015).

When considering some smaller reviews and individual studies, however, there does seem to be some evidence to suggest that training can be important. When looking at peer-tutoring around reading, Reckrut (1994) suggested that the most successful tutoring programs include training for tutors in interpersonal skills, management skills, and content skills.

Some interesting insights also come from Greer and Polirstok (Greer & Polirstok, 1982; Polirstok & Greer, 1986). When they followed tutors across time - from before being trained, then through and after training- they found that tutors improved their use of reinforcement with tutees in the tutoring. Further, progress in tutees increased at a higher rate after tutors had received more training (though the researchers consider this could have been confounded by affects of maturation).

Finally, in their study looking into mathematics peer tutoring with Primary aged pupils, Fuchs and colleagues (1994) used dynamic assessment measures and video tapes of tutor-tutee interactions. It was found that in pairings where tutors had received more training in a structured approach more interactions that the researchers considered to be instructionally sound occurred, and tutors used a more interactive style of explanation in mathematics operations. Although no achievement outcomes were directly measured, the researchers suggested that the styles adopted by the trained tutors can lead to higher rates of learning in tutees.

2.3.4.9 Which student populations can see the positive effects of peer-tutoring?

When considering the age of participants Leung (2015) found that the studies that included students from secondary schools displayed larger effect sizes for academic outcomes than did those of students at other educational levels, this being followed by university/ college level, primary/elementary school, and then kindergarten/nursery participants. However, care should be taken when considering results from the youngest age group, as only 2 studies were included in the review.

For the academic ability of the tutee, Cohen and colleagues (1982) found there were greater effect sizes –although not statistically significant- surrounding academic achievement for tutees with lower academic ability levels. When Leung (2015)

considered tutee academic ability though, all levels gained positive effects for academic achievement, tutees with high ability displaying larger effect sizes. Interestingly, this was closely followed by children of the lowest ability - children with average ability levels seeing the smallest effects.

Considering children with Special Educational Needs, Scruggs and Mastropieri (1998) carried out a review of the effectiveness of peer tutoring for both tutors and tutees. From their findings, they concluded that: students with special needs benefit academically (whether tutees or tutors) and there are often improved attitudes to the curriculum area, as well as improved interactions with partners outside tutoring sessions. Indeed, in their meta-analyses of single case experiments Bowman-Perrott and colleagues (2013) found that peer tutoring is effective regardless of disability status. Additionally, they found that among students with disabilities those with emotional and behavioural disorders benefitted most (Bowman-Perrott et al. 2013).

Taken together, it seems children of all ages, abilities and needs can reap the potential positive rewards of being involved in a peer-tutoring programme- although, it seems that children who may see the most improvement are children who are of secondary age.

2.3.4.10 Cross-age or same-age peer-tutoring?

One of the more obvious ways peer-tutoring interventions can differ is whether the tutee and tutor are of the same age- and this is another area which has seen mixed research outcomes.

Cohen, Kulik, and Kulik (1982) concluded from their review that cross-age tutoring showed greater effects for both tutees and tutors than same-age. On the other hand, Leung (2015) found same age reciprocal and nonreciprocal peer-tutoring and cross-age peer tutoring displayed significantly similar effect sizes when considering academic achievement of tutees and tutors.

In a more recent large scale study involving 3520 children in a number of Primary schools by Timms and colleagues (2011) it was found that positive effects on academic achievement in the subject area of mathematics were only found to be significant

when cross-age tutoring was used. Overall, the positive effects seen for same-age peer-tutoring were not found to be significant. (This study is explored in more detail later, as it is included in the systematic review).

In their meta-analysis considering the most effective methods of mathematics instruction for students with learning disabilities, Gerston and colleagues (2009) found that cross-age tutoring appeared to work more effectively than same-age peer-tutoring, and any other peer assisted learning approaches. They argue this is because having an older tutor ensures that the tutee is more likely to be engaged in more meaningful mathematical conversation (Gerston et al. 2009). It was also argued, though, that there are relatively fewer studies looking at cross-age tutoring and that more research should be undertaken in this area (Gerston et al. 2009).

Research evidence, it seems, suggests that although both same-age and cross-age peer tutoring can be effective, cross-aged tutoring might be considered a better choice than same-age tutoring.

2.3.5 Summary: review of evidence for peer-tutoring approaches

It appears there is quite a large amount of evidence that peer-tutoring interventions can be effective in improving academic achievement across a range of subjects, for both tutor and tutees, among young people within a range of ages and needs. There is some evidence to suggest though, that the individuals who benefit most are those in secondary school environments and who are lower academic achievers.

Peer-tutoring may also be effective for promoting numerous other attitudes and behaviours likely to be conducive to academic achievement, as well as more socio-emotional domains including self-esteem.

It also seems there is evidence to suggest that peer-tutoring interventions that are structured, are of shorter duration but with frequent sessions each week, involve some but not too frequent training for tutors and involve tutors that are older than their tutees (cross-age) could be the most effective.

Considering these conclusions together with conclusions discussed earlier around the need for more research around mathematics intervention, the researcher now presents their own systematic literature review to consider, in particular, evidence around cross-age peer-tutoring in mathematics.

2.4 Systematic Literature Review

2.4.1 Definitions

There is a general recognition that a key element of Evidence Based Practice is that ,in order to make the best decisions around intervention as practitioners, there also needs to be an integration of the best and most current research evidence (Schlosser &O'Neil-Pirozzi, 2006).

Systematic reviews are a way to bring together evidence using a formally defined set of processes, where it is hoped that by doing so, what is known about a topic of research and how we know it can be made very clear (Gough and Elbourne 2002).

Systematic review is a vital tool to identify what we know, how we know it, and therefore what gaps there are in the research literature –so as to guide the direction of future research (Eagly& Wood, 1994).

Higgins and Green (2011) suggest that a systematic review should consist of five key characteristics:

- A clearly stated set of objectives with pre-defined eligibility criteria for studies
- An explicit, reproducible methodology
- A systematic search that attempts to identify all studies that would meet the eligibility criteria
- An assessment of the validity of the findings of the included studies
- A systematic presentation, and synthesis, of the characteristics and findings of the included studies.

The researcher used these features to guide the systematic review that follows.

2.4.2 Aims of the systematic literature review

This review, adopting a post-positivist methodology, aims to systematically explore the experimental research investigating the impact of cross-age peer tutoring in mathematics.

2.4.3 Eligibility criteria

The inclusion and exclusion criteria employed to select eligible papers are detailed below. A brief explanation for criteria is given where the re-searcher considered necessary (*):

Inclusion Criteria	Exclusion Criteria
The research evaluates a cross- age peer tutoring approach.	The research only evaluates a same-age peer-tutoring approach or another other peer assisted learning approach.
The peer tutoring is based around mathematics activities	The peer tutoring is based around subjects not including mathematics.
The research measures outcomes for the tutee (measuring outcomes for tutor was desirable but not necessary)	Outcomes are not measured for the tutee
The research is an article published in a peer-reviewed journal	The research is unpublished (e.g. report or dissertation, a conference presentation or a book chapter)
The research involves young people	The research has used adults.
The intervention is carried out within a school setting	The intervention is carried out in a setting outside of school
The research involves an empirical investigation with an experimental design. *	The research involves a qualitative investigation, narrative case study, theoretical discussion, personal opinion or is a review (secondary research)
Pre and Post measures are taken.*	Both Pre- and Post- measure are not taken
The full article is accessible OR only the abstract is accessible and provides sufficient detail for it to be included	The full article is not accessible AND the abstract does not provide sufficient detail
The paper is written in English OR the abstract is written in English and provides sufficient detail for it to be included.	The paper is not written in English OR the abstract is written in English, but provides insufficient detail for the researcher to determine its inclusion.

Table 2.1A table displaying the eligibility criteria for systematic literature review

*The research involves an empirical investigation with an experimental design and Pre and Post measures are taken- this step was taken in line with the epistemological standpoint in an attempt to select research that sought to establish causality i.e. where the independent variable (IV) (the peer-tutoring intervention) is manipulated to measure its efficacy.

2.4.4 The search and filtering process

A systematic search of the electronic databases PsycINFO, Education Resource Information Centre (ERIC) and Web of Science was carried out in July 2015. After some initial searches using the thesaurus and mapping onto subject specific terms, it was decided that the following search terms would be used to identify relevant studies (these were slightly adapted depending on the format accepted by the database):

"peer tutoring" OR "peer learning" OR "peer assisted learning" OR "peer instruction"
AND (mathematics OR maths OR numeracy)

Each of the databases used had a search tool which allowed the researcher at the outset to only view papers that had been peer-reviewed and this was applied.

From each database, the total number of initial papers found was recorded. Every paper was then filtered by considering first if the title addressed the research question, and then, if necessary, the *Abstract* was considered. If the researcher felt that the paper *might possibly* fit the eligibility criteria, it was recorded in the data base. These papers were then accessed, in full if possible, and papers which met all of the eligibility criteria were considered as part of the final review. If papers could not be accessed in full, then their eligibility was judged by reading the *Abstract*.

Overall, these search procedures resulted in only 4 final papers being identified.

The two main reasons for discounting many of the papers were:

- They did not involve *cross-age* peer tutoring;
- They did not include a comparison or control group (they did not involve experimental designs).

Below is a table detailing the number of papers identified in each phase of the search. The databases are presented in the order in which the searches were carried out, and therefore only additional (not duplicate) papers are detailed in the numbers presented:

Data Base	Initial Search	After initial filter (title and abstract) excluding duplicates	After secondary filter (full access considering eligibility criteria)
ERIC	120	35	3
PsychInfo	205	4	1
Web of Science	83	0	0
Total Number of final papers:			4

Table 2.2: A table displaying the outcomes from systematic search of 3 online data bases using the terms and inclusion criteria specified.

2.4.5 Weight of Evidence

The quality of the evidence was reviewed by the researcher using Gough's (2007) 'Weight of Evidence Framework'. The weightings of the studies were formed based on judgements of 3 aspects of the studies:

A) Methodological quality

- The trustworthiness and quality of the study in relation to 'normal' standards for studies using the design it has adopted (e.g. clarity of definitions and aims).

B) Methodological relevance

- How appropriate the evidence/research design is in answering the review question.

C) Topic relevance

- The appropriateness of the focus of the evidence for the review question.

These judgements were then combined to give a final overall Weight of Evidence (D)

Each paper was judged to be high, medium or low, in relation to these 4 aspects, which can be seen in table 2.3 below:

Study	A Methodological quality	B Methodological relevance	C Topic relevance	D Overall Weight of Evidence
Johnson and Bailey (1974)	High	Medium	High	Medium/High
Sharpley, Irvine and Sharpley (1983)	High	Medium/ High	Medium/High	Medium/ High
Beirne-Smith (1991)	Low/Medium	Medium	High	Medium
Tymms, Merrell, Thurston, Andor. Topping and Miller (2011)	High	High	High	High

Table 2.3A table presenting the weight of evidence of each of the four experiments selected for the systematic literature review.

2.4.6 Synthesis of papers

In line with recommendations from Higgins and Green (2011) the characteristics of each of the papers are considered including: location, sample characteristics (e.g size, age, educational setting) design, outcomes investigated, measures and intervention characteristics (e.g. length, structured vs. unstructured tutoring) please see Appendix 8.1 for a table summarising these details. The key features and findings from each of the papers will now be considered and an overall synthesis of the systematic literature review will be given, including overall implications.

2.4.6.1 Johnson and Bailey (1974): Cross-age tutoring: fifth graders as arithmetic tutors for kindergarten children

The earliest study found in the systematic review investigated the effects of a cross-age tutoring intervention on basic arithmetic skills, with five fifth-grade students who tutored five kindergarten children. The study involved a quasi-experimental design for the tutees, where a further control group of 5 kindergarten children were included who received no tutoring and were matched with the experimental group on arithmetic ability. The tutoring consisted of 26 sessions of 20 minutes over 7.5 weeks and focussed on teaching skills to be able to "use" the numbers from 0 to 100. Tutors were given three 30 minute training sessions before the tutoring began, and tutors were given a matrix of 'skills' which they were to work through, and used flashcards at the end of each session to test progress. Pre-, mid-, and post- testing was carried out for tutees using a skills-based arithmetic test constructed and administered by the experimenter, and authors also used the Tutor Observation Scale (Niedermeyer, 1970) to carry out systematic observations of tutor-tutee interactions.

Results showed that the tutees made statistically significantly greater gains on the arithmetic test than their comparison control group. Additionally, a sub-analysis showed this significant improvement was only for the specific skills that had been taught within the tutoring. With the systematic observations of the tutor-student interactions, quite wide variability in the amount of praise given by the tutors was found, and the authors considered this to have affected the impact on arithmetic learning. It was concluded overall that cross-age tutoring is a viable approach to

increasing the academic performance of young children, but that when training for tutors included more structured techniques this may be more beneficial.

2.4.6.2 Sharpley, Irvine and Sharpley (1983): An examination of the effectiveness of a cross-age tutoring program in mathematics for elementary school children

This US study investigated the effectiveness of a cross-age tutoring program on mathematics in the areas of "operations" (4 basic operations) and "other mathematics" (estimation, geometry, fractions, time) for tutees and tutors. This study also investigated the impact on the self-esteem of tutors.

76 fifth and sixth grade students were assigned to the role of tutors and 76 second and third graders to the role of the tutee. The study involved a quasi-experimental design with a control group for both tutors (n=63) and tutee (n=95). The peer-tutoring occurred for 30 minutes 4 days a week over 5 weeks. During the time allocated to tutoring sessions, controls remained in their classrooms for normal mathematics lessons with the class teacher, thus the peer-tutoring was instead of normal class-teaching rather than supplementary. Though stated that training did occur for the tutors, no specific information is given in the article around what this training was, or indeed what was the structure or focus of the peer-tutoring sessions, reducing the reliability of this research.

Authors found, using the Yardsticks Criterion Referenced Tests in Mathematics (1975), that both tutees and tutors increased their general overall mathematic achievement scores significantly compared to their control groups, and this was regardless of their initial maths achievement scores (higher and lower ability students performed on statistically similar levels). When considering tutors' self-esteem, using the Coopersmith Self-Esteem Inventory (Coopersmith, 1967) as a measure, findings indicated that there was no significant impact compared to the control group.

2.4.6.3 Beirne-Smith (1990): Peer-tutoring in arithmetic for children with learning disabilities

This study explored the effects of cross-age tutoring on the acquisition of single-digit addition facts in students in the role of tutees and tutors. Tutees were students with

learning disabilities receiving support in a “primary self-contained learning disabilities classroom” and while students in the role of the tutor were from the same schools but with no learning disabilities. Additionally, the study investigated whether two different styles of tutoring procedures would show different effects: Method (A) using a counting-on approach and Method (B) and a rote-memorization approach. A randomised control trial was used for the tutees, where they were randomly allocated to either Method A of tutoring (n=10), method B of tutoring (n=10) or a control group (n=10) with ages ranging from 6.0-10.0. Although tutors were randomly assigned to either method A or B, a control group was not included for tutors. Ages for tutors ranged from grade 3-6. All tutors received two 45 minute training sessions, the first around the use of contingent reinforcement, task and error correction procedure and the second on their respective tutoring methods (either a three-task, counting-on procedure (Method A) or a two-task, rote-memorization procedure (Method B). Sessions were highly structured, and used the same sequence of working through 20 sets of three related addition facts and each session ended with a test using flashcards. The tutoring sessions were 30 minutes long, and occurred over 4 weeks, however it is not clear from the methodology how frequent the sessions were, reducing the reliability of the study.

Using a test around single-digit addition facts constructed by the experimenter, it was found that tutees improved significantly more than students in the control group, though there were no significant differences between the methods used. Using this same measure, no significant differences were found for tutors between Method A and Method B, though overall tutors appeared to improve. However, there was a ceiling effect with tutors on this measure as most scored near 90%, at pre-test.

2.4.6.4 Tymms ,Merrell , Thurston, Andor, Topping and Miller 2011: Improving attainment across a whole district: school reform through peer tutoring in a randomized controlled trial

The most recent study found in the review, and indeed the largest, looked to investigate 3 organisational variables of peer-tutoring in mathematics in 129 UK (Scottish) primary Schools: whether same-age or cross-age tutoring is more effective,

whether a more intensive or lighter approach is more effective and whether it is more beneficial for students to participate in only reading or mathematics peer tutoring or for them to participate in both.

A factorial clustered randomized controlled design was involved, where schools were either allocated to: cross- or same-age tutoring; light or intensive tutoring; maths, reading or maths and reading or a control group. The intensive groups received three sessions a week of tutoring, whereas the light groups received one session a week, where each session was 30 minutes in length for 15 weeks. The tutoring sessions were structured using programme called Duolog Maths, where a structure is given to interactions between tutors and tutees to help the tutees solve mathematics problems. For tutor training, although a training video was given to tutors, no other details were specified.

Using Performance Indicators in Primary Schools (PIPS) and curriculum based assessments around the Scottish 5–14 Curriculum, researchers assessed: mathematics, reading, vocabulary, nonverbal ability, and attitudes to mathematics and reading for both tutees and tutors (also assessing science for tutors) pre- and post-intervention.

Findings from the study suggested that only cross-age peer tutoring was consistent in its impact on attainment in reading and mathematics for both tutors and tutee. No other clear pattern of significant results were obtained. This suggests that same- age peer tutoring did not have a significant impact, and also that results were not significantly affected by the intensity of the intervention, or whether they received *both* reading and mathematics tutoring or just one or the other.

This study did have some limitations associated with its large scale implementation, however, including loss of schools and pupils during the project, schools changing their intervention group and a differential loss to follow up among the older cohort.

2.4.7 Overall synthesis of studies

All of the four studies reviewed found that students involved in a cross-age peer-tutoring intervention in the role of the tutee experienced significant improvements in mathematic achievement when compared to control participants. This included 'use' of the numbers 1-100, knowledge of single digit facts, use of operations and general mathematics skills. One study indicated that this was regardless of initial academic ability (Sharpley, Irvine and Sharpley, 1990). Another study indicated that cross-age peering tutoring was particularly effective when compared to same-age tutoring, and that the intensity of the intervention had no significant impact (Timms and Colleagues, 2011). These results were found across a range of standardised, non-standardised curriculum based, and researcher constructed measures.

Three of the four studies considered impact on mathematics achievement for tutors. All found evidence to suggest the intervention had a significant positive impact, and in two of the studies (Timms and Colleagues, 2011 and Sharpley, Irvine and Sharpley, 1983) this was also significant in comparison to a control group.

Only one study investigated impact on self-esteem, and this was only in tutors (Sharpley, Irvine and Sharpley, 1983): no significant difference was found when compared to students in the control condition.

The researcher considers the following limitations to the research reviewed:

In most of the studies, specific procedures around training given to tutors and the actual nature of the tutoring programmes were rarely detailed, which the researcher considers reduces the reliability of the studies, and makes it hard to comment on the level of structure used in these examples of peer-tutoring.

Further, only half of the studies reported that they had carried out checks around the treatment fidelity of the interventions, reducing their external validity.

Apart from one study from Timms and colleagues (2011), all of the studies were carried out in the United States, so it is difficult to generalise their findings to UK schools.

All of the studies found in the review involved primary or nursery aged children and none in secondary, so this limits the ability to generalise finding to older populations.

Across all the studies, none of the researchers were blind to the allocation of groups, which means that all of the studies could have been subject to researcher bias, reducing the external validity of the findings.

Only half the studies included a control group for tutor participants, thus reducing the ability to draw causal conclusions around the effects of the peer tutoring for these students.

Although studies included a control group for the tutee participants, the nature of the control groups was not always made clear. It was not always clear what was meant by 'instruction as usual' , or whether pupils were removed from lessons and missed out on their 'instruction as usual' for the peer-tutoring intervention.

Finally, no comparison groups were included in the studies. This reduces the ability to conclude that it is the specific nature of the peer-tutoring: could it be that just spending extra time one-one or in a smaller group is the reason positive gains were seen?

2.4.8 Review of key unpublished research

The systematic review above sought only work published in peer-reviewed journals. However the researcher considers that there are two very recent and as yet unpublished pieces of research that are also pertinent to the research question, and these will now be considered.

2.4.8.1 Lloyd, Morris, Edovald, Skipp, Kiss and Haywood (2015a): Durham Shared Maths Project

In June 2015, results from the Durham Shared Maths project funded by the Education Endowment Fund were published online. This large scale project in primary schools, which was carried out over two years in England, considered the use of a cross-age peer tutoring approach called Shared Maths: a structured approach which is a

modified version of the Duolog Maths tutoring model used in the study from Timms and colleagues (2011). Year 5/Year 6 pupils tutored younger pupils in Year 3/Year 4.

Using random allocation, 82 schools were either allocated to take part in the intervention or act as a control. 2,806 students were involved in the peer-tutoring condition (1,426 who started in Year 3 and 1,380 in Year 5) and 2,663 students were in the control group (1,360 who started in Year 4 and 1,303 in Year 6).

Although the primary outcome of interest in this study was maths attainment, they also looked at attitudes and reading. All three of these outcomes were measured pre- and post-intervention using an Interactive Computerised Assessment System (InCAS), which is a diagnostic, computer-adaptive assessment tool (Merrell & Tymms, 2005). For maths the InCAS assessment covered topics of "General Mathematics" and "Mental Arithmetic". The InCAS attitudes module looked at attitudes towards reading, maths and school in general. The peer-tutoring took place for 20 minutes each week over two 16-week blocks, one in each academic year. The questions worked on in the sessions were set by the Year 3s class teacher. At the start of the intervention all pupils are trained by teachers in how to solve problems using the Shared Maths process and the pupils' roles, though specific details around the training is not provided in the report.

Strikingly, contrary to many studies reviewed in this thesis, when considering the maths outcomes no significant difference was found between the intervention and control groups for either tutees or tutors. This was also the same for both the secondary outcomes, attitudes and reading achievement. Thus, evidence suggests that the cross-age peer-tutoring programmes had no additional impact on either tutees or tutors above the maths instruction being received in the normal classroom in this study.

However, the following information was gathered from teachers in the process evaluations element of the research. Teachers felt that:

- The Shared Maths consolidated existing skills and maths knowledge.

- They had not been participating long enough in the programme to judge if maths ability had improved.
- Pupils' reasoning and problem solving abilities improved as the programme increased pupils' confidence in maths and put into context why different numeracy methods are important.
- Tutees got a 'huge boost' in confidence in their maths ability-the effects of this differed between schools and whilst some reported that it was damaging the confidence of tutors, others felt that it made them more determined to learn and remember their maths skills in order to help their tutees in the future.
- Tutoring others showed the tutors how far they had come in the subject, increasing their confidence and self-esteem by making them 'suddenly aware of their own progress'.
- Some teachers believed that Shared Maths would have long-lasting effects on their pupils' confidence in maths and approach to other tasks.
- They felt pupils' problem solving had 'improved significantly' since starting the programme.

The authors also commented that there were some issues with the treatment fidelity of the peer-tutoring in some of the schools, which could possibly explain some of the null findings, and there were some high levels of attrition in the study.

2.4.8.2 Posselthwaite, Semmens and Clarke (submitted for publication, 2015):

Paired Maths: a study exploring the effectiveness of a cross-age peer tutor approach to improve times-table knowledge and increase self-esteem.

Also very recently, some colleagues within the researcher's local authority recognised the need to more rigorously investigate the effectiveness of using a structured peer-tutoring approaches to improving basic maths skills and carried out a piece of research within 4 UK primary schools. Year 3/4 students were randomly allocated to either a Paired Maths intervention where they were tutored by year 5/6 pupils, a waitlist control group, or a comparison group involving small group Teaching Assistant (TA) support. Paired Maths again, is a slightly modified version of the Dialog Math used in the study from Timms and Colleagues (2011), and here it was used to teach knowledge

of the 3,4 and 5 times tables. A number of timed probes designed by the researchers were used to measure improvement in the times tables knowledge, and both the Behavioural indicators of Self-esteem (Burnett, 1998) (a teacher-report measure) and the Burnett Self Scale (Burnett, 1994) (a self-report measure) were used to measure impacts on self-esteem.

Results showed that while both the younger participants in the Paired Maths intervention group and comparison group showed significantly improved knowledge of the 3, 4 and 5 times tables (while the control group did not), only participants involved in the Paired Maths intervention (Tutees and Tutors) showed significant gains on the measures of self-esteem.

The authors of this study considered that this finding around self-esteem was particularly important. Because the group who received TA support also progressed academically, but did not see increases in self-esteem, they argue that other factors to do with the nature of intervention itself could have increased self esteem. For example, it was undertaken with a peer in a 1-1 situation where the tutee is receiving focussed attention and some research has indicated that giving children more attention can lead to increases in learning (Topping, 1995). In addition, the paired maths process promotes the use of praise linked to effort and approach, not just achievement. Praise that is explicitly linked to effort and approach/strategy has been found to be very effective in terms of promoting self-esteem and motivation (Kamins & Dweck, 1999; Muller & Dweck, 1998). In light of the above, the authors deemed that it is possible that the increases in self-esteem might have occurred irrespective of achievement, and that cross-age peer tutoring may be an effective intervention for increasing *both* mathematic achievement and self esteem. It is important to note that, similar to the majority of other studies in the literature, this study did not include a control group for investigating effects for *the tutor* in the paired maths intervention, which significantly reduces the ability to draw causal conclusions around the impact of the intervention for these pupils.

2.4.9 Systematic review: Conclusion

Firstly, considering the vast amount of literature on peer assisted learning approaches and peer-tutoring, the researcher considers that there are a very limited number of examples looking at cross-age peer tutoring in the area of mathematics using rigorous experimental methodology, and many of these are now somewhat dated. This is particularly the case within the UK research base. Although some studies have begun to emerge recently in the UK, inconsistent findings suggest that further clarification is needed around the effectiveness of cross age peer-tutoring in mathematics using rigorous research methods. To the author's knowledge, no research of this kind has been undertaken within secondary populations.

2.5 Intervention within the secondary aged population

Research around interventions within secondary schools seems particularly pertinent at the present time. The annual report from Ofsted in December of 2014 expressed great concern for progress in the secondary phase. Findings suggest that while primary schools now seem to be improving outcomes for their students (82% are now *good or outstanding*) improvement in secondary schools seems to have come to a standstill. Only 71% of secondary schools are at least good and GCSE results have not risen over the last three years (Ofsted and DfE, 2014).

Echoing the review here around cross age-peer tutoring, however, a report for the DfE in 2012 considering literacy and numeracy catch-up strategies found there is actually very little evidence on any numeracy interventions for secondary school pupils, and although it appears ideal for interventions to be carried out in the early primary years, some children for various reasons may not receive such interventions, or may not benefit maximally from them. Also, some mathematical difficulties may only show themselves at a later stage. Indeed, Dowker (2004) reflected on this and concluded that, "it would be desirable to carry out more studies of the effectiveness of arithmetical interventions with older children and adults." (Dowker, 2009)

In terms of teaching for secondary students, a seminal meta-analysis conducted by Swanson and Hoskyn (1998) concluded that the following are important components for teaching skills to struggling secondary students:

- Using small, interactive group instruction
- Using directed questioning and responses
- Breaking tasks down into component parts and fading prompts and cues
- Using extended practice with feedback

The researcher considers that all of these aspects could also be used to describe peer-tutoring.

2.6 Self-esteem

As the literature surrounding peer-tutoring, particularly the recent piece of research involving cross age peer-tutoring in mathematics (Posselthwaite, Semmens and Clarke, submitted for publication, 2015), suggests that these interventions *can* have an impact on self-esteem, the researcher considered it was important to explore this area further.

2.6.1 Definitions

Self-esteem is the evaluative component of self-knowledge. Put more simply, it is how much value people place on themselves, or 'positive or negative attitude toward the self' (Rosenberg, 1965, pg.30). When an individual has a high-self-esteem, they view themselves favourably, and conversely someone who has low self-esteem has an unfavourable definition of themselves. It refers to how a person views themselves and so is more concerned with perception than reality. Therefore, self-esteem does not have to be an accurate evaluation. For example, low self-esteem could be an accurate understanding of one's short coming, or could be a distorted view of inferiority.

Similarly high self-esteem could be a justified appreciation of competency or success, but could also be an inflated sense of grandiose superiority (Baumeister, Campbell, Krueger and Vohs, 2003). Within the literature, the term self-esteem is often used synonymously with the term self-concept.

2.6.2 Potential outcomes surrounding self-esteem

There has been a vast amount of literature surrounding the possible effects or factors of high or low self-esteem for children and adults. One particularly influential paper in

the area comes from Baumeister and colleagues, who in 2003 carried out a large scale meta-analysis into the research around self-esteem and links to a number of life outcomes including school and task performance, interpersonal relations, aggression, violence, happiness, coping, depression, health, sex and eating behaviours. The researcher draws heavily upon this meta-analysis when considering some of the outcomes relating to self-esteem, but focuses on outcomes considered pertinent to the rationale for exploring self-esteem within educational intervention research.

2.6.2.1 Self-esteem and wellbeing

The Department for Education (DfE) (2014) recognises that: “in order to help their pupils succeed; schools have a role to play in supporting them to be resilient and mentally healthy”. Moreover, in line with the Children Act (2004) all schools have a duty to promote the wellbeing of students.

A wealth of research has found a strong link between high self-esteem and increased feelings of happiness across a variety of ages and contexts (Baumeister et. al, 2003). For example, when considering a large sample of 406 young people (aged 14–28), Furnham and Cheng (2000) found that self-esteem was the most powerful predictor of happiness among a range of other potential correlates including parental rearing styles and scores on the Eysenck Personality Questionnaire.

It is not surprising, then, that research has also found that low self-esteem is strongly linked to depression (Baumeister et. al, 2003). In large scale study, Robinson, Garber, and Hilsman (1995) interviewed nearly 400 young people during 6th grade and then again in 7th grade. They found that self-esteem had a direct effect on depression. Interestingly, they also found that levels of self-esteem predicted levels of depression regardless of levels of life stress. Further, Robinson and colleagues (1995) found that with high school students, depression was increased by a combination of external stresses *and* a tendency to blame oneself for failure- which was much more common in those with low self-esteem. So, it could be that it is a negative attribution style associated with those with self-esteem that leads to higher levels of depression. A later study from Furnham and Cheng in 2003, replicated these feelings around happiness

and depression, where global self-esteem remained the most powerful direct predictors of both.

As research has suggested that individuals with higher self-esteem, when presented with negative events, are more optimistic about their coping strategies and see themselves as having more control, Baumeister and colleagues (2003) theorised that these individuals may experience less stress and therefore are less likely to suffer from poor health. Indeed, in a large scale study using prospective data from the Dunedin Multidisciplinary Health and Development Study birth cohort, Trzesniewski and colleagues (2006) find that adolescents with low self-esteem had both poorer mental and physical health during adulthood compared with adolescents with high self-esteem (taking into account gender, socioeconomic status and depression). The researcher considers this study to have particularly high validity, as outcome variables were assessed using objective measures and informant reports rather than just relying on self-report data, unlike much research into self-esteem. It also involves prospective data, which allows more conclusions to be made around the direction of causality.

Relating to this health outcomes, other studies have seen quite strong associations of self-esteem with eating disorders, particularly amongst adolescents (Baumeister et. al, 2003). In one of the largest pieces of research investigating this association comes from French and colleagues (2001). They found in a sample of over 90,000 children and adolescents in grades 6 to 12 that self-esteem had one of the strongest correlations with eating disorders. There are also some prospective studies in this area that suggest self-esteem may play a more causal role with eating disorders. For example, Button, Sonuga-Barke, Davies, and Thompson (1996) found in a sample of 400 females, those who had low self-esteem at aged 11 to 12 were significantly more likely to have developed an eating disorder at age 15 to 16.

Although some inconsistent findings are seen in the literature, there is some research supporting a link to between self-esteem and levels of externalising behaviours (lashing outward at others, such as aggression, violence, defiance and criminal behaviours (Jianghong, 2004)). One study, which Baumeister and colleagues (2003) argue provides some of the most robust evidence to support this link, comes from

Trzesniewski and colleagues in 2002. In this longitudinal study, over 700 students were followed from age 11 to age 13, and it was found that self-esteem was significantly correlated with objective measures of externalising behaviours (including teacher report, parental report and interview) at both time points, and also that self-esteem at age 11 predicted externalizing behaviour at age 13. Further, this was still the case when taking into account other possible confounding variables such as relationships to parents, relationships with peers, IQ, and socioeconomic status. This finding was also replicated later in the study mentioned above from Trzesniewski and colleagues in 2006 using prospective data from Dunedin Multidisciplinary Health and Development Study birth cohort.

2.6.2.2 Self-esteem and academic achievement

A generally accepted standpoint within educational communities is that if young people have a higher self-esteem, they are more likely to do well in school. Further, self-esteem is often talked about in relation to children who are underachieving and/or experiencing difficulties with school life. However, when one looks at the existing research, only relatively modest correlations have been found between general self-esteem and school performance (e.g Byrne, 1984; Byrne and Shavelson, 1986; Hansford and Hattie, 1982; Marsh and Yeung, 1998 and Mintz and Muller, 1977). Indeed, in a more recent meta-analysis from Valentine, Dubois, and Cooper, (2004) found that that positive self-perceptions and academic achievement only have on average across studies a relation of around 0.08. Further, a number of longitudinal studies that have tried to look more at the direction of causality in this modest relationship, have not found that general self-esteem plays a role in improvement, but instead found results to suggest that it is increased academic achievement that increases global self-esteem (Bachman and O'Malley, 1977; Helmke and van Aken, 1995, Maruyama et al., 1981, Pottebaum et al., 1986, Skaalvik and Hagtveg, 1990).

A good illustrative example comes from Skaalvik and Hagtvet (1990). This large longitudinal study was carried out in Norway across consisted of 600 children in two cohorts, one in third grade and the other in sixth grade. Data was collected at the start of the school year and then again a year and a half later. Findings suggested that

children who did well in school one year led to much higher levels self-esteem the next year, whereas high self-esteem did not lead to increased academic achievement.

Actually, quite strikingly, high levels of self-esteem in grade 6 students predicted lower levels of academic achievement the following year. More specific measures were also taken of students' self-perceptions of their academic ability (academic self-concept).

Skaalvik and Hagtvet (1990) suggests that results gained point to there being a relation between academic performance and global self-esteem which is mediated by academic self-concept. In more simple terms, when children academically achieve, they see themselves as better at work in school, which in turn boosts their overall self-esteem.

Finally, there are also quite a large group of studies where findings have suggested that interventions and programmes which specifically target and try to and enhance self-esteem have limited effects, or in some cases have even led to a decrease in academic achievement (e.g. Scheirer and Kraut, 1979; Forsyth and Kerr, 1999).

Baumeister and his colleagues (2003) therefore concluded from their meta-analyses that self-esteem has no significant impact on subsequent academic achievement. If any, they argue high self-esteem can have a detrimental effect on academic achievement, and therefore intervention which focuses solely on enhancing self-esteem could be damaging on academic performance. However, a later meta-analysis from O'Mara, Marsh, Craven, and Debus (2006) found when interventions target targeted specific domains of self-esteem rather than just overall global self-esteem (i.e. academic self-concept) that there can be positive impacts on self-esteem.

Also, higher levels of this more specific domain of self-esteem, academic self-concept, have also been linked with behaviours that could also arguably encourage academic success. For example, persistence (Lent, Brown, & Larkin, 1986; Skinner, Wellborn, & Connell, 1990), intrinsic motivation (Bandura, 1997; Skaalvik, 1997), the adoption of task and achievement goals (Bong, 2001; Skaalvik & Skaalvik, 2005), and better self-regulation strategies.

2.6.3 What does this mean for intervention around self-esteem in schools?

It is clear that self-esteem plays an important role in ensuring a child's happiness and wellbeing, and that low levels of self-esteem can be associated with a number of negative life outcomes including depression, poor health and a number of externalising possibly criminal behaviours. Given the duty schools have to promote the wellbeing of students, intervention around self-esteem seems an important area to consider.

Research around academic performance and self-esteem has suggested that increasing academic performance can increase self-esteem, whereas interventions focussing solely on enhancing self-esteem in isolation are not seen to be effective unless they specifically target academic self-concept. The researcher considers that the literature around self-esteem suggests that interventions that focus on increasing academic performance and also if possible an individual's academic self-concept are those that may have the most positive effects on self-esteem.

2.7 Conclusion: Rationale and original contribution for the current research

Mathematics difficulties in the UK are prevalent and potentially lead to negative life outcomes. However, although research is limited compared to literacy intervention, intervention can have a positive impact on these mathematical difficulties, and peer-tutoring approaches could be a highly cost effective option.

Research suggests that peer-tutoring as a general approach, particularly cross-age peer tutoring, can be effective in terms of both academic achievement and also elements of self-esteem. Low self-esteem can be an important predictor of a number of negative outcomes, but can be positively affected by improving academic achievement.

The systematic literature review indicated cross-age, structured peer-tutoring with mathematics could be an effective intervention for increasing mathematics achievement and promoting self-esteem. However rigorous experimental studies that include a control comparison group for both the tutee and tutor are limited - particularly in the UK. Further, no research of this kind (to the researcher's knowledge) has been undertaken with secondary aged students.

Therefore the current research aims to investigate the effectiveness of using a cross-age peer-tutor programme (Paired Maths) with secondary aged pupils using a rigorous experimental design, involving control and comparison groups with Randomised Control Trials (RCT).

The investigation addresses the following research questions:

In a UK secondary school, does a 7-week, cross-age, structured peer-tutor programme around maths (Paired Maths):

1. Enhance mathematics attainment in Year 8 students in the role of tutees?
2. Enhance self-esteem in Year 8 students in the role of tutees?
3. Enhance self-esteem in Year 10 students in the role of tutors?

3 Method

3.1 Introduction

The previous chapter reviewed the body of literature highlighting the rationale for the current research study. This chapter will outline the design and methodology adopted to address the research questions set out above. Broader areas of science and research are first considered, including critical discussion of the philosophical assumptions underpinning various research paradigms and a focused consideration of the post-positivist stance assumed in this research. The Randomised Control Trial (RCT) design, characteristics, measures and procedures used are then outlined, with a final review considering questions of validity, reliability and ethics.

3.2 Theoretical rationale

3.2.1 Ontology, epistemology and methodology

It is recognised that research is just one of many ways that we can begin to know or understand something (Mertens, 2014). However, in the increasing pursuit of evidence-based practice (Oakley, 2002) there has been much more emphasis on the use of research to establish "What Works?" in both education and psychology (Kratochwill et al, in Gutkin & Reynolds, 2009 p.497).

Cohen, Manion, & Morrison (2011) conceptualise research as an approach that considers both reasoning and experience in an attempt to discover truth. It is considered crucial that on this road to discovery researchers' thinking and practice are guided by a world view (or paradigm) that is as close to their own conceptions about this 'truth' as possible (Mertens and Wilson, 2012; Cohen, Manion & Morrison, 2011). Mertens (2014) identifies that there are four key research paradigms cited in the literature: Post-positivism, constructivism, pragmatism and transformatism. Mertons (2014) also suggest that there are four key areas of philosophical assumption underpinning each paradigm, relating to : reality (ontology), knowledge (epistemology), systematic enquiry (methodology) and the nature of ethics (axiology). Each assumption is linked to another, where fundamentally, a researcher's conception

about the nature of reality (ontology) and what we can know about it (epistemology) influences how they can and should interact with it (methodology and axiology). Table 2.1 outlines the key features of the different paradigms informing the axiological, ontological, epistemological and methodological perspectives assumed within this study (adapted from Mertens, 2014).

Basic Belief	Post-positivism (evolved from positivism)	Constructivism	Transformative	Pragmatic
Axiology	Privacy Informed consent Minimal harm Justice/ Equal Opportunities	Balanced representation of views Increased participant's awareness Community rapport	Respects cultural norms and promotes human rights Increase social justice	Researcher's values and politics effect the process
Ontology	One reality known within a degree of probability (critical realism)	Multiple realities, socially-constructed	Rejects cultural relativism Considers social positioning and privilege	Each person has their own interpretation of one true reality
Epistemology	Objectivity essential	Interactive link between researcher and participants	Interactive link between researcher and participants, but acknowledge social and historic power imbalances	Relationships determined by what is required
Methodology	Typically Quantitative Interventionist De-contextualized	Mostly Qualitative Hermeneutical Context considered	Typically uses, Qualitative but Mixed methods possible Context and history important	Method driven by research question Often uses Mixed Methods

Table 3.1: Research Paradigms (adapted from Mertens, 2014)

3.2.2 The chosen paradigm: (Post)Positivism

Post-positivism stems from positivism, which assumes that the social world can be studied in the same way as the natural world (Mertens, 2014). It adopts a realist ontology, where there is one true reality that exists independently of the person observing it (Fielding, 2009). Epistemologically, knowledge of this reality is gained from direct experience or observation only, and research involves the following features:

- Objectivity
- Value-free
- Quantitative data
- Hypothesis testing
- Causality between variables is established
- General / universal laws are proposed (adapted from Robson, 2011)

However, a number of criticisms of adopting a purely positivist standpoint within research in the social world have been raised, for example, Ritchie and Lewis (2003) highlighted concerns around whether, when studying human behaviour, we can:

- Completely 'control' variables?
- Disregard contextual variables?
- Disregard meaning and purpose of behaviour?
- Neglect the importance of discovery through alternative understandings, by only focussing on hypothesis testing?

These criticisms led to most within the social sciences moving away from positivism, and to the development of the post-positive perspective (Byrne, 1998).

Post-positivism adopts a 'critical realist' ontology, which still upholds the belief that there is only one true reality. However, it argues that because of limitations of human experience and behaviour, this reality can only ever be known within a certain degree of possibility (Mertens, 2014). It is recognised that any research cannot directly 'prove' a theory, but to get as close as possible attempts should be made to try and control any variations that arise from the social world. Fixed designs and quantitative

measures are most commonly adopted to try and achieve this (Mertens, 2014). Indeed, the present study has been conducted within a post-positivist paradigm, reflected in the selection of a Randomised Control Trial (RCT) experimental methodology, the use of quantitative data and the statistical analyses of findings.

3.2.3 Experiments and RCTs

Experiments are often the chosen methodology for those following the post-positivist paradigm as they are considered the best way to establish causality and generalisability (Cohen, Manion & Morrison, 2011). This involves controlling for extraneous variables, which influence the ability to link the manipulation of independent variables (IVs) to variability in the dependent variables (DVs) -internal validity (Cohen, Manion & Morrison, 2011). However, it is also important that experiments, particularly within the social sciences, are carried out ethically and have external validity, where findings can be generalised to real life. How this was managed in the current research is considered in section 3.13.

Randomised Control Trials (RCT's) are considered by some to be the *gold standard* of experimental approaches and are ranked highly in the 'hierarchy of evidence' (Burns, Klingbeil, Ysseldyke & Peterson-Brown, 2012; Petticrew & Roberts, 2006). This is because when participants are randomly allocated to experimental and control groups, one can arguably somewhat limit the threat of other extraneous variables influencing the research conclusions.

The RCT method was also selected here since it is considered appropriate to the research question. RCTs can be used to assess effectiveness and some argue are important for the development and evaluation of evidence-based educational interventions (Robson, 2011; Slavin, 2002). Indeed the use of an RCT methodology is also beneficial because of the limitations of previous research into cross-age peer-tutoring in mathematics. For example, although much of the previous research reported positive outcomes in mathematics attainment and attributed these to the intervention, the lack of control or comparison conditions meant that it was not possible to establish that they were due to cross-age mathematics peer-tutoring interventions specifically, or to additional mathematics support generally.

3.3 Research Design

In this exploratory research, experimental methods in the form of two RCTs were adopted to test for causality between the intervention -Paired Maths (PM)- and mathematics attainment in participants in the role of the tutees and self esteem in both tutees *and* tutors. The two RCTs looking at the impact for each side of the pair (tutee and tutor) are outlined below.

3.3.1 Investigating effects for tutees

This part of the research adopted a three group pre- and post-test RCT. Year 8 participants were randomly allocated to one of three groups: Experimental (PM, Role of tutee); Comparison (TA led); and Wait list Control. Pre-measures (T1) were taken from all participants one week prior to the intervention commencing. Each group intervention lasted for seven weeks, with one additional week in the middle with no intervention due to school closure. Post-intervention measures (T2) were then taken from all participants during the week following the intervention. This process is illustrated in figure 3.2 to follow.

3.3.1.1 Independent Variables

The RCT for Tutees was a mixed design measuring the effect of two Independent Variables (IVs):

1. Treatment
2. Change over time

The treatment IV was between groups and had three levels (Experimental (PM), Comparison and Wait-list Control Group) examining whether there is a difference between the groups that could be attributable to the intervention. The Change over time IV was within group and had two levels (T1 and T2) examining whether there is a difference between pre- and post- testing within each group that could be attributable to the intervention.

3.3.2 Investigating effects for tutors

This part of the research adopted a two group pre- and post-test RCT. Year 10 Participants were randomly allocated to one of two groups: Experimental (PM, Role of Tutor) or Wait list Control.

Again, pre-measures (T1) were taken from all Year 10 participants one week prior to the intervention commencing. As the Year 10s in the Experimental group were tutoring the Year 8s assigned to their experimental group, again each group lasted for seven weeks (with one additional week in the middle with no intervention due to school closure). Post-intervention measures (T2) were then taken from all participants during the week following intervention. This process is also illustrated in figure 3.2.

3.3.2.1 Independent Variables

The RCT investigating the impact for Tutors was again a mixed design measuring the effect of two IVs:

1. Treatment
2. Change over time

The treatment IV was between groups and had two levels (Experimental (PM) and Wait-list Control Group) examining whether there is a difference between the groups that could be attributable to the intervention.

The Change over time IV was within group and again had two levels (T1 and T2) examining whether there is a difference between pre- and post- testing within each group that could be attributable to the intervention.

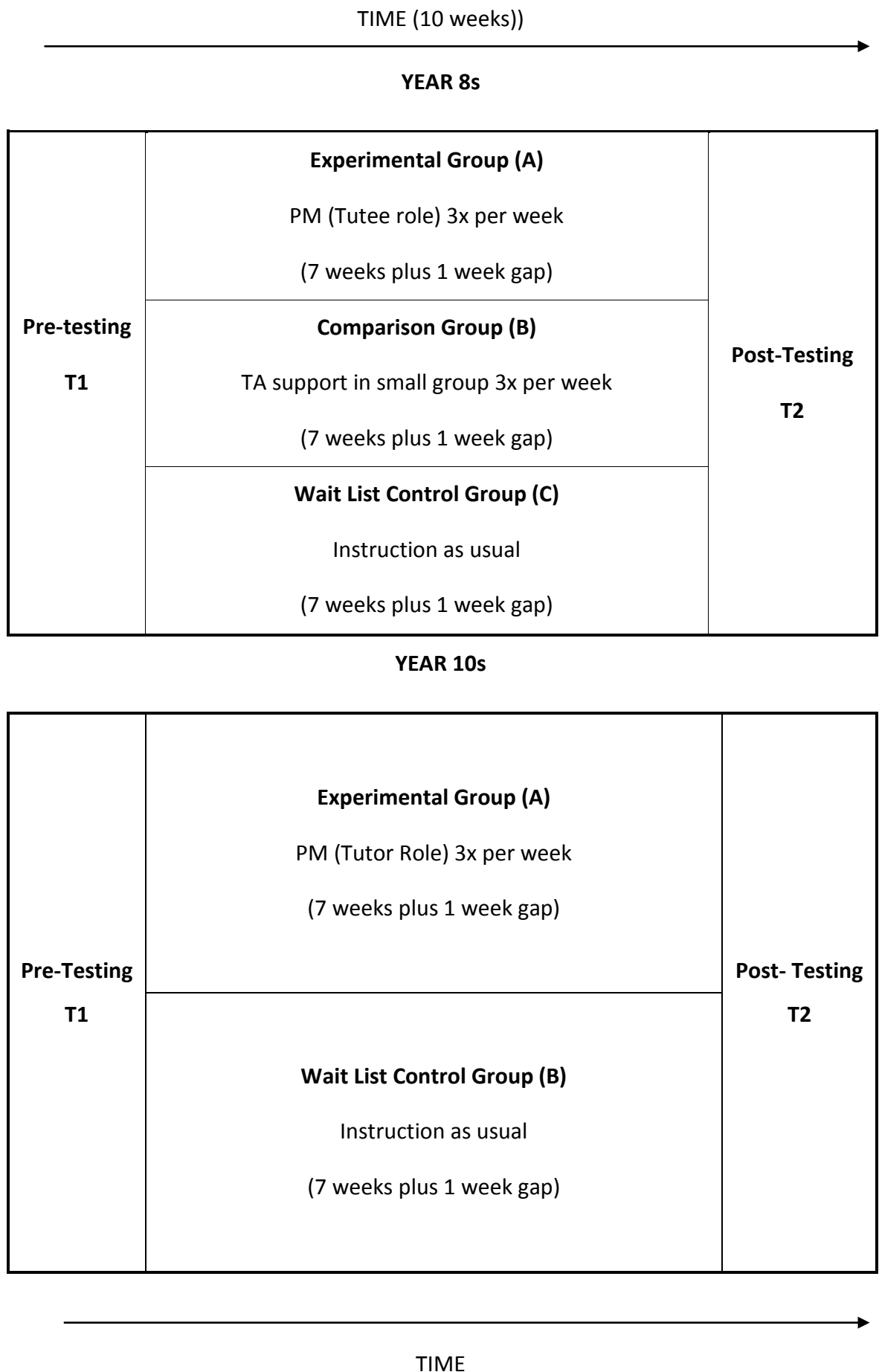


Figure 3.2: A figure showing the ten weeks of the research study

3.4 Characteristics of the current study

3.4.1 Context: The locality

The researcher was, at the time of the research, employed by an East Midlands UK local authority Educational Psychology Service as a *Trainee Educational Psychologist*. The research took place in a small city within this area with a population of around 100,000 (reference withheld to protect the anonymity of the local authority).

3.4.2 Stakeholders

Within applied research, there are a number of stakeholders involved, and it was important to keep the needs and rights of these stakeholders in line with the British Psychological Society's ethical codes of practice throughout (2009;2014) the key stakeholders in this research study were as follows:

- The Local Authority

The research was undertaken within the researcher's professional role as part of the Educational Psychology Service (EPS) for the Local Authority.

PM was an intervention for which the EPS had already provided training to a number of primary schools in the area, and had carried out a research project (as discussed earlier in chapter 1) (Posselthwaite, Semmens and Clarke, submitted for publication 2015). This research was therefore important in supporting the EPS to consider evidence locally as to whether the intervention, training and resources for school should be promoted to secondary schools.

- University of Nottingham

This research was completed as a compulsory part of the Doctorate in Applied Educational Psychology (DAEP) (professional training) at the University of Nottingham. This is a three year full time course that prepares students to work as Educational Psychologists (EPs) in a professional capacity. The researcher was supervised by academic staff within the University of Nottingham and outcomes from research undertaken on the course at the University can potentially influence the direction of

future research and practice in relation to the academic field of Educational Psychology.

- Research Schools

The schools involved in the research were key stakeholders due to their role in providing access to the participant samples, intervention facilitators, involvement in assessment procedures and resulting time and resource considerations. They also had an interest in the results, which may influence the use of numeracy programmes or peer-tutoring approaches within the school (more details about the schools are provided below).

- Parents

The parents of the children had an interest in the wellbeing and educational achievement of their children, and also provided consent for their children to take part.

- Children

The children themselves had an interest in their own wellbeing and educational achievement. The study may impact upon the mathematics support or peer approaches that they experience after the research period.

3.5 Participants

3.5.1 Schools

3.5.1.1 Selection criteria

Recruitment procedures for schools are explored in more detail later (section 3.12), however, schools had to meet the following criteria to be included in the research:

- The school demonstrated commitment to the study by protecting the time for the interventions three times a week at the agreed times.
- The school was able to provide two members of staff to support the PM and Comparison intervention groups.

- The school had not previously implemented the PM intervention.

3.5.1.2 School characteristics

Three secondary schools participated in the research, all of which taught pupils from 11- 18 years of age. A summary of the key characteristics of each school is provided in table 3.2 below.

School	Size	Pupils from ethnic minority backgrounds	Pupils on the SEN register	Most recent Ofsted Grading	Most recent Ofsted Inspection
A	Smaller than Average	Below Average	Average	Requires Improvement	Monitoring Visit 12/15
B	Third of the Average size	Below Average	Above Average	Inadequate	Monitoring Visit 01/06
C	Larger than Average	Below Average	Above Average	Good	11/14

Table 3.2 A Table presenting the key characteristics of the three schools participating in the current research

These are taken from each school's most recent Ofsted report. References for this information and the school's names are withheld to ensure their anonymity.

3.5.2 Intervention facilitators

From each school, two Teaching Assistants (TAs) were required to be involved in the research: one to supervise the PM intervention group (involving both Year 8 and Year 10 experimental groups), and one to lead the Year 8 comparison intervention group. Each TA had been employed by the school for at least 3 years and all were female. In two of the schools, the TAs typically supported students within class across a number of different subjects including maths. None of the intervention facilitators had previously been trained or run a PM intervention.

3.5.3 Pupils

3.5.3.1 Selection/ inclusion criteria

As it has been highlighted that cross -age peer-tutoring can be more beneficial, it was decided to include participants from different year groups for the role of the Tutee and Tutor. This is because research has shown that cross-age pairings tend to show more positive gains for both tutees and tutors than same-age pairings (Gerston et al., 2009). A gap of two school years has most commonly been used in existing research (including the recent RCT carried out in Primary schools from Posselthwaite, Semmens and Clarke (submitted for publication)).

Gender was not included as part of the selection criteria, and both cross and same-gender pairings of the Year8s and Year 10 occurred. To the researchers knowledge, the existing research around peer-tutoring, particularly focussing on mathematics, does not highlight any differential outcomes for same or cross gender pairings. It was also felt by the school staff that it was more important to match the students on whether they were likely to work together well.

Inclusion criteria for each set of participants is set out below:

3.5.3.1.1 Criteria for participants in the role of the tutee (year 8s)

The research aimed to consider whether a PM intervention could benefit children in the role of the Tutee who were considered to be struggling in mathematics. To identify these children, the Special Educational Needs Co-ordinator (SENCO) in each school was given a set of inclusion criteria. Each part of this criteria- with the researcher's rationale for its inclusion- is set out below:

The pupil would be in Year 8 at the time of the intervention

It was decided to include year 8 participants in the role of the tutee, rather than year 7 participants, in order to allow time in the previous academic year (while the participants were still in year 7) to both identify and gain consent for all participants before the training and interventions began at the beginning of the next academic year.

The pupil's routine school assessments show that they are at least 2 sub-levels below age-related expectations in numeracy/ mathematics.

Mathematic attainment was considered based on school assessment data. At the time of selection procedures in the research, this was based on the National Curriculum levels (DfE, 2013). Pupils were selected if they were 'below' age related expectations. This was considered to be at least 2 sub-levels below in their last assessments in the Summer term before the intervention began.

Although it was considered, a screening test for mathematical level was not undertaken. This is because it was considered by the researcher to be unethical to gain consent and screen a large number of children with the intention of only supporting those who reached a specified criterion. Also, it is argued that school data which is gathered within natural settings, and which is triangulated over time, is actually more ecologically valid and reliable than a one-off screening test (Axworthy, 2005).

The pupil has >90% attendance and would be able to attend the selected time slots each week

This was considered important to reduce the risk of mortality (participants leaving the study) and ensure that the intervention was accessed adequately. Five year 8 pupils who were originally selected and participated in the interventions (1 in the PM intervention and 3 in the TA Comparison group) were not included in the final sample used for analysis as their attendance was below 90%, however all other participants attended >90% of the sessions which were offered during the research period.

The pupil is not receiving any other additional maths intervention or peer-tutoring

This was to try to increase the internal validity of the research, to increase confidence in the ability to attribute any changes to the PM intervention being examined.

3.5.3.1.2 Criteria for participants in the role of the tutor (year 10s)

To identify the children for the role of the tutor, again, the SENCO in each school was given a set of inclusion criteria. Each part of this criteria –along with the researcher's rationale for its inclusion- is set out below:

The pupil will be in Year 10 at the start of the intervention

After discussion with the key staff in the schools, it was decided to involve participants who would be in Year 10 for the role of the tutor, as older students in the school would have more pressures around examinations, and the school staff thought that Year 8 participants may not feel comfortable if they were to be tutored by participants closer to their own age in Year 9.

The pupil's routine school assessments are at least at or above their age related expectations for mathematics

As highlighted in Chapter 1, research has shown that Peer-tutoring can show more positive gains for *tutors* when they are of lower ability compared to those of other academic ability levels (Leung, 2015). However, after discussions were held between the researcher and the mathematics staff within the schools, it was clear that the schools felt more comfortable and confident in including pupils in the role of the tutor who were at least at age expected levels for their mathematics. It was expressed that this would be a key factor in pupils being willing to take part in the research and also to feel confident enough to engage with the PM framework.

The pupil would benefit from intervention that could enhance self-esteem

As previous research has indicated that cross-age peer tutoring can enhance self-esteem for both the tutees and tutors, it was thought that this inclusion criteria could aid school's selection processes, and reach those that may benefit most from the intervention. A formal selection criteria was not imposed with this, as it was not considered ethically appropriate to screen children for self-esteem.

3.5.4 Pupil Characteristics

3.5.4.1 Participants in the role of the tutee

A total of 47 year 8 pupils participated and were included the final analysis. 15 from School A, 16 from School B and 16 from School C. 22 participants were male and 25 participants were female. As participants were in year 8, they were all aged between 12 and 13 years. Two participants were Black or Minority Ethnicity (BME). As research reviewed earlier in chapter 2 suggests that peer-tutoring can be effective regardless of whether an individual is considered to have Special Educational Needs (e.g.Bowman-Perrott et al. 2013) three children with autism were included.

3.5.4.2 Participants in the role of the tutor

A total of 31 Year 10 pupils participated and were included in the final analysis. 10 from School A, 11 from School B and 10 from School C. 13 participants were male and 18 were female. As participants were in Year 10, they were all aged between 14 and 15 years. Three participants were Black or Minority Ethnicity (BME).

3.6 Participant allocation

3.6.1 Year 8 participants in the role of the tutee

Year 8 pupils were randomly allocated to the experimental (PM) group, the comparison (TA led) group or Waitlist Control Group. The random allocation process was conducted in *each school separately*; by picking names out of a hat.

Group	Number of Participants
Experimental (PM)	16
Comparison (TA led)	15
Waitlist Control	16

Table 3.3: Table showing the number of Year 8 participants randomly allocated to each group

3.6.2 Year 10 participants in the role of the tutor

Year 10 pupils were randomly allocated to the experimental (PM) group or Waitlist Control Group. Again, the random allocation process was conducted *in each school separately*; by picking names out of a hat.

Group	Number of Participants
Experimental (PM)	16
Waitlist Control	15

Table 3.4: Table showing the number of Year 10 participants randomly allocated to each group

3.7 Experimental group: Paired Maths

This research used a structured peer-tutoring intervention called Paired Maths (PM), the approach that was also used in the previous research carried out within the same local authority with primary aged pupils (Posselthwaite, Semmens and Clarke, submitted for publication). During PM interventions, pairs of students, one student in the role of a tutor and one in the role of a tutee, complete maths activities together. Tutors are trained to follow a structured process to scaffold their tutees learning during the activities. The process tutors are trained to follow, which was originally devised by Keith Topping and his colleagues in 2004, is detailed below:

1. Understanding the question

- Read: Your tutee might be having trouble reading a word question. If so, read it for them.
- Identify: Make sure that your tutee understands what the question is asking.
- Listen: Give your tutee time to think about the question and then ask them to explain how they might solve it.

2. Finding an answer to the question

- Question: Ask helpful and intelligent questions that give clues, to stimulate and guide your tutees thinking. Challenge things that you think are wrong.
- Praise: Give your tutee praise and encouragement very often, even for a very small success with a single step in solving a question. Keep their confidence high.
- Think out loud: Give your tutee some thinking time, before expecting an answer. Encourage them to tell you what they are thinking all the time. Then you will find out where and how they are going wrong. Remember tutors also need time to think! If you are not sure, say so. You are not supposed to know everything.

3. Finishing the question by asking themselves what have they done and how it links to things they have done in the past

- Check: Check that your tutee eventually gets the right answer. But remember there is probably more than one “right” way to solve the question.
- Sum-it-up: Have your tutee summarise the key steps in doing the maths question. Point out any errors or gaps, then summarise the key steps to yourself.
- Link-it-up: The tutor and tutee should talk about how the learning might be used to do another similar question (generalize it to another maths question) or how the learning might be useful in a wider context.

(Topping, Kearney, McGee &Pugh, 2004)

3.7.1 PM in the current research

As identified in Chapter 2, existing research suggests that in terms of time and length, 'optimum' peer-tutoring interventions should be no longer than 10 weeks long and involve frequent (3 or more times a week) but short sessions. In this research the interventions took place for 7 weeks (with a week's break for school holidays), and it was planned that there would be 3 sessions per week (actual numbers of sessions

were less than this). It was planned that the sessions would last for 20 minutes in the morning on Monday, Wednesday and Friday during tutor time before lessons started. It was decided that the sessions would not take place during mathematics lessons so that the intervention did not substitute normal mathematics teaching, as this has been an important element highlighted by Dowker(2004) for effective mathematics intervention.

Each PM intervention group in each school consisted of 6 pairs of students, with year 10 students (in the role of the tutor) each paired with a year 8 student (in the role of the tutee). Each tutor was supplied with a “Paired Maths Pack”. This comprised of a distilled diagram (laminated) of the PM process (Appendix 8.2), a booklet of additional guidance for each section of the process (Appendix 8.3), and A3 maths mat aid (Appendix 8.4) including a number line , a fraction wall and mathematical language definitions. This study used the same PM diagram as used in the study from Posselthwaite, Semmes and Clarke (submitted for publication) which was adapted from a diagram that was published online by the University of York Institute for Effective Education (2015), distilling the PM process into a simple diagram, whilst retaining the key features that are necessary for optimal learning . The guidance booklet given was also adapted from the resources provided by the University of York Institute for Effective Education (2015).

Each PM group had a bank of maths activities (ordered and labelled in sequence of difficulty). Part of the tutor role was to help their tutee select activities in each session. The mathematics activities used in this research were the SMILE (Secondary Mathematics Individualised Learning Experiment) Wealth of Worksheets. These were initially developed as a series of practical activities for secondary school students by practising teachers in the 1970's but were accessed by the researcher online via the National STEM Learning Centre (2015). These activities were selected for the following reasons:

- The activities included focussed on the number system and basic operation/ computations (e.g., subtraction, addition, multiplication) areas where peer-tutoring interventions have seen more positive effects (e.g. Leung (2015))

- They covered a range of difficulty appropriate for the selected participants (Curriculum Levels 2- Level 6 were made available)
- They were activities that could be attempted as a pair but also would be appropriate to be attempted as an individual (as they were included in the comparison condition, details to follow)
- They are readily available and free to download online
- They were included as a suggested useful resource for mathematics peer-tutoring interventions in the 'Paired Maths Initiative Evaluation Report' from the Junior Certificate School Programme in Ireland (McCarthy,2008)

3.7.2 Training

The researcher delivered a 1.5 hour training session to each school involving the Year 10 students (in the role of the tutor) and the member of staff overseeing the intervention. This was delivered in the Autumn term one week before the PM intervention began to the Year10 participants allocated to the experimental group, and then again for the participants in the wait-list control group after the research phase. The presentation used as the basis of this training (written by the researcher, adapting some of the materials presented by the University of York Institute for Effective Education, 2015) can be found in Appendix 8.5, but the key elements included in the training are outlined below:

- Overview of the rationale and aims of the study
- Details of the PM materials and the PM process; outlining the roles of the tutor and tutee
- Opportunities for the tutors to observe the PM process in action (as modelled by the researcher)
- Opportunities to practice PM using the materials in pairs
- Feedback and guidance from the researcher was given to the tutors throughout the practice element of the training

3.8 Comparison group for year 8 participants

3.8.1 Rationale for comparison group

A comparison group was included in this research in order to strengthen conclusions that it is the specific nature of the PM that has an impact on the desired dependent variables. Without the inclusion of a comparison group, the researcher suggests it would not be possible to rule out whether it was merely spending extra time on maths activities in a one-one or in a smaller group context that caused any changes on the dependent variables.

3.8.2 Comparison group in this research

The participants allocated to this group worked individually on the same maths activities given to the PM intervention group but with the support of a TA in the context of a small group (6 participants in total). These groups occurred at the same time during tutor time on a Monday, Wednesday and Friday before the commencement of lessons for the same amount of time (20 minutes) and occurred for the same 7 week period. Prior to the intervention, the researcher held discussions with each of the TA's selected from the 3 schools to discuss and agree guidelines around the type of support that would be given in the small groups. This was to increase treatment fidelity across the groups and provide a valid comparison for the experimental group (please see appendix 8.6 for these guidelines). The TA's were also given the opportunity to familiarise themselves with the maths activities.

3.9 Waitlist Control Groups (Year 8s and Year 10)

Participants randomly assigned to the wait-list control conditions, for both year 8s and year 10s, did not receive any additional intervention during the 7-week research intervention stage, but continued to receive the typical whole-class mathematics teaching alongside the participants in the other conditions. Following the conclusion of the main research phase, to ensure that the wait-list control conditions were not disadvantaged, participants were offered to take part in a PM intervention (Year 8s in role of the tutee, Year 10s in role of tutor) and training was delivered again by the researcher.

3.10 Treatment integrity

Dane and Schneider (1998) argue that any robust evaluation of an intervention should include assessment of integrity. They identified five aspects to intervention integrity:

- Adherence -the extent that the intervention is delivered as it was designed
- Exposure -the amount of the intervention that the participants experience (e.g. number of sessions, frequency, duration)
- Quality of delivery -the skills that the implementer brings to the delivery of the intervention
- Participant responsiveness- the level of participant response or engagement with the intervention
- Programme Differentiation- the extent to which treatments in different conditions are differentiated

To consider adherence, quality of delivery and participant responsiveness in this study, the researcher carried out three observations of the PM sessions in each school. A checklist of PM features (see Appendix 8.7) was used to help this process. This involved both general elements of the PM intervention group as a whole, but then also the opportunity to observe one Tutor- Tutee pair more closely during each integrity check. As there were only 3 integrity checks in each school, and therefore not all tutor-tutee pairings could be directly observed, the researcher acknowledges this as a limitation.

The exposure component was assessed by asking each TA running the interventions (experimental and comparison) to keep a register for the sessions documenting which participants attended them and how many sessions took place.

In order to ensure programme differentiation between the experimental PM and the intervention phases, as well as the guidance given to all TAs at the beginning of the research, two observations of the TA comparison group were completed in each school. The researcher also used the agreed guidelines as a basis for these observations (see Appendix 8.6)

3.11 Measures

Measures are used to detect changes on dependent variables, which in this study are:

- Mathematics Attainment in Year 8 tutees
- Self Esteem in Year 8 tutees and Year 10 tutors

All measures selected (apart from the Behavioural Indicator of Self Esteem, details below) were group administered by the researcher and this was done separately for each age group (one group for Year 8s and one group for Year 10s) in each school. For both T1 and T2, assessments took place over three consecutive days, with a day in each school. For Year 8 participants, self-esteem measures were completed before the mathematics measure, reducing the risk of artificial reduction in scores. Data was recorded by hand on the questionnaires/test papers in accordance with standardised procedures.

3.11.1 Maths attainment: Basic Number Screening Test

3.11.1.1 Rationale for selection

In her review of "What works for children with Mathematical Difficulties" Dowker (2009) reports that standardised assessments can be useful in both deciding who should benefit from an intervention programme and also monitoring the effectiveness of interventions. A number of factors were considered to select the most appropriate standardised assessment, details of which can be seen in table 2. The Basic Number Screening Test (BNST) (New Edition), Originally developed by Gillham and Hesse and restandardised by McCarthy in 2015, provides an assessment of a pupil's grasp of the basic principles underlying the number system and the processes involved in calculation. It was chosen for the current research because:

- It provides two parallel forms that:
 - are suitable for pre- and post- testing within research time frame
 - could be counterbalanced to avoid order effects
- It can be group administered, saving time
- It is presented orally and so does not rely on any literacy skills

- It can be used with the Year 8 age group (although originally designed for ages 6-12, the authors advocate the use of the test with older less able children)
- The difficulty of the items were hoped to be at a level where the participants could experience some level of success, yet still with some room for improvement (an attempt to remain ethical *and* avoid floor/ceiling effects).
- It was highlighted as a useful assessment by Dowker (2009)

Test	Age Range	Parallel forms?	Time Needed Between Administrations	Group or Individual Administration	Presented Orally?
Basic Number Screening Test-New Edition (Hodder Education)	Designed for 6-12 years (BUT can be used for older less able pupils)	Yes	No minimum	Group or individual	Yes
Graded Arithmetic-Mathematics Test (Hodder Education)	5-12	No	Unknown	Group	No
Mathematics Assessment for Learning and Teaching	5-14	No	Unknown	Group or Individual	No
Progress in Maths (GL Assessment)	6-14	No	1 year	Group or Individual	Some, not all parts
British Abilities Scales (3rd Edition) Number Skills Test	5:00-17:11	No	6 months	Individual	Some, not all parts

Table 3.5 An overview of the assessments *considered* to measure maths attainment

3.11.1.2 Administration

The test was administered to all year 8 participants as a group in each school. The test was presented orally by the researcher following standardised instructions and lasted around 30 minutes. As the instructions are exactly the same for both parallel forms of the test this allowed for all Year 8 participants in each school to be assessed together. The assessment yielded a maximum raw score of 30.

3.11.1.3 Reliability and validity

The BNST was originally standardised on 3000 children within one large city. The test was further re-standardised in March of 2011 on 8250 pupils aged 5:0 to 12:11 from 59 schools across England and Wales, and included children with English as an additional language.

The test-retest reliability (Pearson product-moment correlation co-efficient) and the internal reliability (Cronbach's Alpha) of the BNST are reported to be "exceptionally high" by the authors (Where $r=0.94$ and $\alpha=0.05$ respectively) and they argue that this indicates that it "is a very robust test" (pg.26 of Manual).

Regarding validity, when the test was originally developed, teachers from one of the schools used in the standardisation process were asked to rate pupils on a 7 point scale before administering the test. These scores were then correlated with the test-scores and these correlations (Spearman's rho) averaged at +0.82, which the authors argued was "very satisfactory". In addition, during the most recent standardisation process, pupils' teacher-assessed national curriculum sub levels were reported to show "very close alignment with test performance" (pg.27 of Manual).

3.11.2 Measuring self-esteem

Two measures were used to assess the impact on self-esteem for both tutees (year 8s) and tutors (year 10s):

- The Burnett Self-Scale (BSS, Burnett, 1994) which looks at self-reports and

- The Behavioural Indicators of Self-Esteem Scale (BIOS, Burnett, 1998) which looks at teacher-reports.

The following factors informed the decision to use the two selected measures of self-esteem:

- They are both standardised measure- which increases their reliability
- They both have good content and construct validity and reliability (details to follow)
- They are appropriate for the nationality and age of participants
- There were a feasible length and difficulty
- The same measures were used in the recent research carried out with Primary School aged children by Posselthwaite, Semmens and Clarke (submitted for publication 2015)-which could increase the validity in comparing results between studies- these authors also recommended the measures to the researcher as they found them to be relatively easy and reliable to administer
- Using two measures of self-esteem, from two different perspectives, provides more triangulation of measurement and therefore more external validity.

3.11.3 The Burnett Self-Scale (BSS)(Burnett, 1994)

In its full form, this is a 40-item self-report scale providing information on individual pupils' thoughts and feelings about different aspects of their lives including: physical appearance; physical ability; peer relations; relations with mother; relations with father; reading self-concept; maths self-concept; learning self-concept; in addition to a global self-esteem sub-scale. For the purposes of this study, only those items pertaining to the sub-scales for maths self-concept; learning self-concept and global self-esteem were employed, as these were felt to be most relevant and important for assessing the effectiveness of a mathematics intervention. For each item on the scales, there are five possible response choices from "*really good/like*" to "*really do not like/really am not good*". Responses are scored from 1 – 5, as appropriate. Please see

Appendix 8.8 for the form used with the 3 scales mixed together and Appendix 8.9 for the corresponding scoring sheet.

3.11.3.1 Administration

In groups, participants were provided with the BSS and a pencil. Pupils were required to answer each question independently, but the researcher and a TA offered support with reading the items if participants had literacy difficulties. Participants were given as long as needed to complete the measure, but this mostly took no longer than 10 minutes. The researcher introduced and gave instructions for the BSS using the standardised script written by the authors of the measure for each group.

3.11.3.2 Reliability and validity

Burnett (1994) reports alpha reliability figures for each sub-scale ranging from 0.67 (peer relations sub-scale) to 0.88 (for the global self-esteem sub-scale), indicating moderate to high reliability.

3.11.4 The Behavioural Indicators of Self-Esteem Scale (BIOS) (Burnett, 1998)

The BIOS is a 13-item scale completed by a class teacher (See Appendix 8.10). It is based on observations of behaviours thought to be indicative of self-esteem, and it requires the teacher to rate each item on a 5 point scale describing how often they have seen a behaviour in the past few weeks, from **1** (Never) through to **5** (Always).

3.11.4.1 Administration

This measure were given to all students' (Year 8s and 10s) mathematics teachers 1.5 weeks prior to the commencement of the intervention (they were all completed before the Year 10s were given training) . They were then given to the same teachers the day after the intervention had finished. Each school was 1.5 weeks to complete the measures and get them back to the researcher after the end of the intervention period. Mathematics teachers were chosen to complete the measures for two reasons. Most importantly, in secondary schools, children are taught by a number of different staff, and therefore it was considered appropriate to choose a teacher to complete the measure who saw the children on a regular/daily basis if possible. Secondly, it was

considered that considering the intervention was surrounding mathematics, it may be that any positive impacts on self-esteem may have been more likely to present within mathematics lessons.

3.11.4.2 Reliability and validity

Burnett (1998) reported reliability coefficients for the BIOS of 0.94 and test re-test reliabilities of 0.82 over a 10 week period and 0.73 over a 20 week period . It was noted as a limitation by the researcher, however, that this measure maybe have been subject to observer bias, as the teachers may have expected a change following the intervention. In addition, as a window of a week and a half was given, it is acknowledged that different schools could have completed the measure at slightly different times, which could slightly reduce validity.

3.12 Procedure

After gaining ethical approval from the University of Nottingham Ethics Committee, the researcher contacted secondary schools that their visiting EP thought would be interested in taking part in the research. A more detailed and dated timeline of the research for each of the three participating schools can be seen in Appendix 8.11, however the key actions are briefly detailed below.

1. Initial meeting (May 2015): After verbal interest had been sought via telephone, the researcher visited the SENCo, Head Teacher (HT) and Head of Mathematics of each school and gave a short presentation about the research and discussions were held around the age of participants to be included. All three schools that were visited agreed to participate in the research, and once they had given written consent to take part, they were then sent the agreed inclusion criteria to consider participants.

2. Meeting two (June 2015): A second meeting with the SENCo and Head of Mathematics in each school was held, and this involved planning dates and pragmatic considerations and any issues around pupil selection were discussed (this was done anonymously).

3. Meeting three (July 2015): The researcher met with SENCos and the TAs in each school to discuss the features of the TA comparison group. Participants who had given informed consent were also randomly allocated to groups.

4. T1 Measures taken and Teacher Questionnaires given (September 2015) All measures, including the BSS and BNST were taken in the 3 schools within the same week, and Maths Teachers were given 1.5 weeks to complete the BIOS.

5. Deadline for Teacher Questionnaires and Paired Maths Training(End of September, 2015) Training for the Year 10s allocated to the experimental group and the Teaching assistants supporting was given to all three schools within the same week. The BIOS questionnaires were collected from teachers *before* the training was given.

6. Interventions Period with fidelity checks (28th September- 20th November 2015)

7. T2 Measures and Teacher BIOS Questionnaires given (End of November 2015) Again, all measures administered by the researcher were carried out within the same week for all schools.

8. Training offered for waitlist control year 10s and commencement of second PM intervention groups (January 2016)

9. Debriefing meetings (Scheduled July 2016).

3.13 Research considerations

When implementing a fixed design such as an RCT, in order to establish if we can assume that findings are true, it is considered imperative to scrutinise the procedures used (Robson, 2011). Mertens (2014) suggests that researchers particularly need to consider two aspects of their research to do this:

1. Internal validity: can we make causal inferences about the relationship between the independent and dependent variables? Basically, in *this situation*, are research findings *likely* to be a result of IV manipulation or other variables?

and 2. External Validity – can findings be generalised to other contexts?

Chen & Rossi (1987) highlight that designing research that can come to a satisfying balance between these two areas can be a difficult task. Increasing control in an attempt to increase internal validity can come with increasing difficulties to generalise, and vice versa. A study that is highly generalisable increases difficulties in assuming causality linked to the manipulated IV.

The following section considers how the procedures involved in this research address this balance between internal and external validity.

3.13.1 Internal validity

Potential threats to internal validity assumed by Cook and Campbell (1979) were identified, and how these were addressed in the current research is considered in Tables 3.6 and 3.7

'Threat' to internal validity (based on Cook and Cambell, 1979)	Level of threat/ steps taken to address
History: Were there factors in the environment impacting on findings over time?	Low: Randomisation guards against findings being attributable to changes in the environment.
Testing: Were there practice effects impacting on the responses given by the participants on the measures?	Medium: Although the BNST has two forms, the measures for self-esteem only have one form repeated and this could potentially have resulted in some bias- this is an accepted limitation. However, there was a time period of 12 weeks between pre- and post- testing which should reduce this level of threat.
Instrumentation: Was there a difference between pre- and post-tests?	Low/ Medium: The researcher administered the BNST and BSS themselves and followed the standardised procedures and scoring guidelines for each measure. However it is acknowledged that there is a higher level of threat for the BIOS, as although specific guidance and supervision was given, it was not the researcher that completed this measure.
Regression: Were extreme groups used?	Low: Random allocation guarded against this.
Mortality: Participants leaving the study.	Low-Medium: Random allocation should account for effects of <i>selective</i> drop-out.

Table 3.6: Table illustrating the 'threats' to the internal validity of the current study and the attempts to guard against them

'Threat' to internal validity (based on Cook and Cambell, 1979)	Level of threat/ steps taken to address
Ambiguity about Causal Direction Did x cause y? or did y cause x?	Low: As the research is not correlational, and there were pre- and post- measures, ambiguity about causal direction is not considered a threat.
Selection Were Initial group differences prior to the research?	Low. Random allocation was considered to minimise this threat. Groups were statistically tested to check that groups were homogenous at pre-testing, and the Analysis of Covariance inferential test used accounts for difference in pre-test scores
Selection by Maturation Predisposition for groups to 'grow apart.'	Low: Random allocation was also considered to minimise this threat. If maturation occurs, both groups could be assumed to mature 'equally' due to being randomly allocated.
Diffusion of Treatment Can the control/ comparison groups benefit from the intervention through the diffusion of treatment from those involved in the experimental group?	Low: As it is the interactions between the tutor and tutee that is key to the PM intervention, and this element cannot possibly be present for the control or comparison conditions in both RCTs, this was not considered to be a threat.
Compensatory Equalisation of Treatment Did the control or comparison group receive 'special' treatment to compensate for not receiving the intervention?	Low/Medium: The TA facilitating the comparison Group was made aware that all groups would receive PM, <i>and</i> that the purpose of the research is focusing on the PM intervention not the TAs themselves. The inclusion of a treatment fidelity observation for the TA comparison group and guidelines for support also reduced this threat. As the control group received no additional intervention this also reduced this threat.
Compensatory Rivalry Did the control or comparison groups try and compensate/try extra hard?	Medium Although the control groups and the TA comparison group were offered the PM intervention after the initial experimental intervention stage, thus reducing the need to compensate, the researcher considers that there still could have been some level of threat around this to the validity of research, particularly on the self-esteem self-report measures.

Table 3.7: Continued....Table illustrating the 'threats' to the internal validity of the current study and the attempts to guard against them

3.13.2 External validity

Cohen, Manion and Morrison (2011) argue that for any finding to be useful, it needs to be generalisable beyond the scope of that specific research. While on the one hand it has been suggested that where studies have a good level of internal validity, the findings are more likely to be generalisable (Anderson, Lindsay and Bushman, 1999) it has been argued that actually some of the features that contribute to high internal validity, such as using an RCT, can also compromise the external validity (Roth and Fonagy, 1996). With these tensions acknowledged, the researcher of the current study has taken the following steps, informed by critical questions posed by Mertens (2014, p. 154) to try and enhance the external validity of the findings:

- Clear descriptions of the PM intervention and comparison condition, and the procedures used within both RCTs increase the ability to replicate what has been done in other situations.
- Participants from mainstream secondary schools were selected for the research in consultation with schools using an explicit and accessible selection criteria. It is feasible therefore to assume that this research can be applied to a similar population in a similar context.
- In both RCTs, randomisation controlled for multiple treatment inference.
- With the RCT investigating effects for the tutee in PM, the inclusion of the TA comparison group reduced the threat of a Hawthorne Effect. However, as a comparison group was not involved in the RCT investigating effects for the role of the tutor, it is acknowledged as a potential limitation.
- The interaction effect of testing assumes that participants may be more sensitive to the DV after pre-testing. It is difficult to guard against this threat where self-report measures have been used, as with the BSS measure of self-esteem, and is therefore acknowledged as a potential threat to external validity.
- While the PM intervention and TA Comparison group were 'novel', peer-approaches and also intervention groups with TAs are not too uncommon in secondary schools, so it is not considered to impact significantly differently on what pupils may 'typically' experience.

3.13.3 Reliability

Reliability refers to the dependability and consistency of the research (Cohen, Manion and Morrison, 2011). Though the reliability for each individual measure has already been considered, Table 3.8 illustrates the potential threats to the reliability of the research as a whole and identifies what attempts were made by the researcher to address these.

Threats to reliability	Consideration in this research and steps taken to address
<u>Researcher Bias</u> When the researcher inadvertently affects the outcomes by unconsciously behaving in a different way towards each group of participants.	Both the BNST and the BSS self-report measures were administered in groups that consisted of participants in both conditions and participant allocation was not explicitly known to the researcher during the assessments and when grading them. Scripts for each researcher administered assessment were strongly adhered to. Although maths teachers who completed the BIOS self-esteem assessments were initially blind to which groups participants had been allocated, at post-test this was not the case. Therefore it is acknowledged that there could be some threat of bias here.
<u>Participant bias</u> When participants try to please the researcher.	Randomisation procedures reduce this threat. Also, during the instructions for the BSS measure it was emphasised to the participants that there are no right or wrong answers. However it is acknowledged there may have been a risk of participant bias with the BSS. Group effects were also avoided by administering measures in 'mixed' groups.
<u>Participant Error:</u> Factors such as tiredness or mood, influencing a participant's responses to a test on a particular occasion.	As only one pre- and one post-test were carried out this could have been a threat. However, random allocation should have distributed this threat between the groups. In addition, considering effects on self-esteem, using the BIOS to triangulate with the BSS reduces this further. This is because the scale requires the teacher to consider the participants' behaviour over a two week period.
<u>Treatment Integrity:</u> This refers to the 'treatment' being delivered as intended	As this intervention relies on each Tutor within each intervention group following the PM process, this threat was considered to be quite high. See section 3.10.

Table 3.8: Table illustrating the 'threats' to the reliability of the study and the attempts to guard against them

3.13.4 Ethical considerations

The research adhered to the ethical principles outlined by the University of Nottingham Research Ethics Committee and also the standards from the British Psychological Society: Code of Ethics and Conduct (2009) and The Code of Human Research Ethics (2014). Approval for the research was granted from the Ethics committee at the University of Nottingham in February 2015 (Appendix 8.12). The key ethical areas outlined by the BPS and how these were addressed are outlined below:

Risk: Considering that both the PM and Comparison conditions involved an out of class intervention, supervised by TAs and focusing on mathematics, this was considered low risk as this is something pupils may ‘typically’ experience in their school day. However, as the PM intervention sessions also involve the added element of working closely with another student, the following steps were taken to try and reduce any potential for conflicted interactions that *may* cause some upset:

- Staff were asked to match tutees and tutors considering their likely compatibility and risk of conflict.
- Tutors were asked during training to follow the set PM framework in their interactions with tutees, which aims to encourage positive and enjoyable interactions and should keep discussions focussed on the mathematics. The supervising members of staff were asked to monitor this.
- It was regularly stated to the participants that participation was voluntary, and how important it was that the participants report if they become uncomfortable in their pairings.
- The TAs supervising the PM group were available at the end of each session for the students to speak to them in confidence if needed.
- The TAs supervising the PM group were asked that if either member of the pairings seemed to be uncomfortable with working in their pair, or the relationship appeared to become inappropriate in some way, they were to cease the pairing immediately. If necessary, following discussions with the either side of their pair separately, a decision would be made in conjunction

with parents around whether it was suitable for the pair to remain working together in the intervention.

It was acknowledged that some items on the BSS measure of self-esteem that require the participants to consider how much they 'like' themselves *could* potentially cause some upset. However, the BSS is a standardised assessment tool which has been widely used with the age groups involved in this research. Also, the researcher and SENCO were readily available after these measures were completed to debrief and support participants if required.

Valid consent: Written consent was gained from the school (signed by Head-teacher), staff facilitating the interventions, parents and the individual children participating in the research. Information letters and consent forms, written to suit their audience, can be seen in Appendix (8.13, 8.14, 8.15, 8.16).

For parents, information and consent forms were sent out directly in the post by the school. However, if it was thought that parents may have difficulty accessing the information letter, the SENCO made a phone call to read it out to them. The SENCO also called parents if they indicated on the consent form that they felt they had not had the opportunity to ask questions about the research, with the follow up offer of a phone call with the researcher if it was required.

As the pupils undertaking the research are 11-14 years of age, the researcher felt that they are at an appropriate age to understand the processes and give their own consent. However, three measures taken to minimize the chances of pupils giving uninformed consent:

- By signing the consent form, parents indicated that they had checked the student was aware of the procedures involved in the research.
- On their information sheet, students were asked twice to speak to an adult if they were unsure of any of the information given to them.
- When the information sheets were initially given to students after parental consent was sought, the SENCO read the information sheets out with them as a group before they were taken home to ensure understanding.

For all involved in the research, it was made clear in the information letters and consent forms and verbally throughout that they had the right to withdraw, at any time.

Confidentiality: All participants were assigned a code, and only the SENCO had access to the child name-code correspondence which was kept in a safe, confidential place in school. When the researcher assessed participants, before leaving the school, the researcher consulted the name-code list and: (1) permanently erased the child's name from the assessment sheets/booklets and (2) added their assigned code to it. All completed assessments were kept in a locked filing cabinet by the researcher office, and all data was then digitized and stored by the researcher on an encrypted memory stick using the pupil codes. The data can therefore only be accessed by the researcher, however, should it ever be accessed by another individual (which is highly unlikely) only the codes of the pupils will be available. If during the selection process the SENCO wished to discuss any children for their suitability, this was done anonymously.

Giving advice: The researcher did not directly interface with pupils during intervention groups. Any difficulties encountered during groups were considered to be addressed through procedures that the school would 'typically' use when facilitating intervention groups.

Deception: At no point in the research were schools, parents or pupils deceived. All information sheets and consent forms were accessible and procedures made clear, with the opportunity to discuss or ask further questions provided.

Debriefing: All stakeholders, including pupils, parents and staff are due to be informed of the main outcomes of the study in writing through a debriefing letter (see Appendix 8.17). This also outlines that stakeholders will have an opportunity to meet with the researcher to discuss any aspects of the research in July of 2016.

3.14 Approach to Data Analysis

Using statistical analysis to test hypotheses is considered necessary in research that exists within the post-positivist paradigm (Mertens, 2014). Null hypothesis testing (the most common approach used within psychological experiments) involves statistical testing to find the probability of obtaining differences/relationships in the data gathered if the null hypothesis were true (Howell, 2013). For both year 8 tutee participants and year 10 tutor participants, data gathered at pre-testing (T1) and post-testing (T2) for all conditions (experimental conditions, 'wait-list control' conditions and comparison condition for Year 8s) was statistically analysed to test the null hypotheses. This was to determine whether there had been any statistically significant changes in participants' responses on the dependent variables that could be attributed to the PM intervention. For positive findings to be attributed to the PM programme, participants in the experimental conditions should have statistically significant positive improvements compared to participants in the wait-list control conditions and, in the case of the Year 8 tutees, the comparison condition.

A statistically significant effect was considered to be achieved if the probability that the result would be obtained by chance were less than 5% ($p < 0.05$). This $p = 0.05$ level is often referred to as the 'Type I error' rate, and is what is most frequently used in research (Howell, 2013). Within Psychological research, both descriptive and inferential statistics form the basis of statistical analysis.

3.14.1 Statistical tests

3.14.1.1 Descriptive statistics

Descriptive statistics help the researcher to concisely summarise the data in a way that will aid them to understand findings and visualise any patterns that have emerged (deWinter and Cahusac, 2014). In the current research, descriptive statistics are used to summarise data associated with each research question, whereby the mean and standard deviations are displayed for the total sample and for each condition. Bar graphs are used to present these visually.

3.14.1.2 Preliminary checks

As descriptive statistics can only offer limited information about data, statistical tests are used to provide more detailed insight into whether there are any significant differences between groups. When choosing the most appropriate test to use, preliminary checks have to be carried out to clarify whether data meets the assumptions for parametric or alternatively non-parametric testing (deWinter and Cahusac, 2014). Parametric tests are usually preferred as they have more statistical power. When the data does not meet the required assumptions, non-parametric tests can be used. Although less statistically powerful, non-parametric tests can arguably be used more broadly as they make less assumptions about the data (deWinter and Cahusac, 2014). The following preliminary checks were applied to determine the most appropriate test to use:

3.14.1.2.1 Normal distribution

Distribution considers if the sample data is normally distributed as would be presumed in the general population. The data are displayed visually using histograms; a normal distribution can be observed as a 'bell-shape' curve with the mean at the peak (Dancey & Reidy, 2014). Further, as it has been highlighted that visual methods provide only limited information and should be supplemented with formal normality tests, the researcher elected to use the Shapiro-Wilk test of normality. A non-significant result of this test ($p > 0.05$) indicates that the data are normally distributed. This was chosen over statistical tests of kurtosis and sphericity, as it is argued that these can be overly sensitive (Pallant, 2013). If data is in violation of the normality assumption, it is recommended that researchers conduct a non-parametric test. However, it has also been argued that when research involves an RCT, some parametric tests (such as Analysis of Co-Variance) should still be the favourable statistical test used, even when data is not found to be normally distributed, as long as the data can meet the other necessary assumptions are met (Vickers, 2005).

3.14.1.2.2 Homogenous groups

As this research aimed to look at differences between groups, it was important to assess whether groups were homogenous, and thus comparable. Levene's test of

equivalent variances was used to assess homogeneity of groups; if a result is >0.05 the assumption of homogenous groups has not been violated. This test is particularly important where there are unequal sample sizes in each group (Wilmot, 2015) (as is within the current research) to ensure that post-test scores do not show significant differences due to group differences that exist at the pre-test.

3.14.1.3 Inferential statistics: Analysis of Covariance

An Analysis of Covariance (ANCOVA) was the inferential statistic selected to assess the difference between (Experimental PM, comparison and control for year 8 tutees and Experimental PM and Control for year 10 tutors) and within each group (T1 vs T2) on all of the DVs. In addition, unique to the ANCOVA, this test can also remove the effect of another variable (Brace, Kemp & Snelgar, 2012), the co-variate. ANCOVA was therefore chosen as it allowed for the inclusion of the covariate of the T1 results, so individual differences in pupils' performances at pre-test (not highlighted by the preliminary checks) could be accounted for. It has been argued that this is particularly helpful when research includes an RCT design (Brace, Kemp & Snelgar, 2012) as any differences in pre-test scores are more likely to have come about by chance (Miller and Chapman, 2001). In addition to the preliminary checks already highlighted for parametric testing, preliminary checks specific to ANCOVAs were undertaken to ensure that the following assumptions were met:

- The relationship between the dependent variable (T2 data) and the covariate (T1 data) should be linear. Scatter plots of the data are used to check this.
- The relationship between the dependent variable (T2 data) and the covariate (T1 data) should be similar for each condition (homogeneity of regression). An Analysis of Variance was used to determine if there is a significant interaction between the condition and the covariate.

3.14.1.4 Non-parametric tests

Where the data violated the assumptions necessary for ANCOVA, the Mann Whitney U Test was used. This test is used to determine if the two groups of data differ

significantly at either pre-testing or post-testing. Following the use of the Mann Whitney U Test, the Wilcoxon Signed Rank test was used to examine if there was a significant difference within each group between pre and post-testing (T1 and T2 scores).

3.14.2 Missing data

If a participant was absent during the group testing sessions, the researcher was able to assess the participant individually on a different day within the same week. This was not considered to have an impact on the research as the same standardised scripts were used for administration. No participant was absent for a whole week, and all BIOS questionnaires were completed by the mathematics teachers within timeframes specified. So, the researcher was able to gather a full data set for both the year 8 tutee and year 10 tutor participants who attended the minimum sessions of the interventions and all the participants in the control conditions.

3.15 Summary

This chapter considered the research methodology used to address the research questions, as well as the approach to data analysis. The following Chapter 4 presents the results of the research.

4 Results

Throughout the following results section data from all three participant schools have been combined.

Each of the research questions set out in Chapter one is considered in turn.

4.1 Research Question 1: In a UK secondary school, does a 7 week, cross-age, structured peer-tutor programme (Paired Maths) enhance mathematics attainment in Year 8 students in the role of Tutee?

The Dependent Variable (DV) in relation to *Research Question 1* is the *pupils' performance* as indicated by the raw scores achieved on the Basic Number Screening Test (BNST). The Independent Variable (IV) is time (T1 and T2).

The hypotheses associated with this research question are outlined below:

Experimental Hypothesis 1: *Participants in the PM intervention group will show significantly higher gains in mathematics attainment than those in a comparison group and/or control group.*

Null hypothesis: *After completion of the PM intervention, there will be no statistically significant differences in mathematics attainment between the PM group, a comparison and/or control group.*

4.1.1 Basic Number Screening Test

4.1.1.1 Descriptives

Table 4.1 displays the means and standard deviations (SD) for the raw scores from the BNST by group at pre- (T1) and post- (T2) testing.

The data shows a small increase across time in all groups, with the control group making marginally larger improvements compared to the other intervention groups: Experimental (PM) (+1.125), Comparison (+0.866) and Control (+1.436).

Figure 4.1 presents this data using a bar chart.

Time	Group	Mean	Standard Deviation
Pre Test (T1)	Experimental (Paired Maths) (n=16)	20.063	4.203
	Comparison (n=15)	20.067	4.605
	Waitlist Control (n=16)	18.500	6.928
	Total (n=47)	19.532	5.332
Post Test (T2)	Experimental (n=16)	21.188	3.799
	Comparison (n=15)	20.933	3.575
	Waitlist Control (n=16)	19.936	4.500
	Total (n=47)	20.681	4.129

Table 4.1: A table displaying the means and standard deviations, by group, at pre (T1) and post (T2) testing for Basic Number Screening Test scores for Year 8 participants

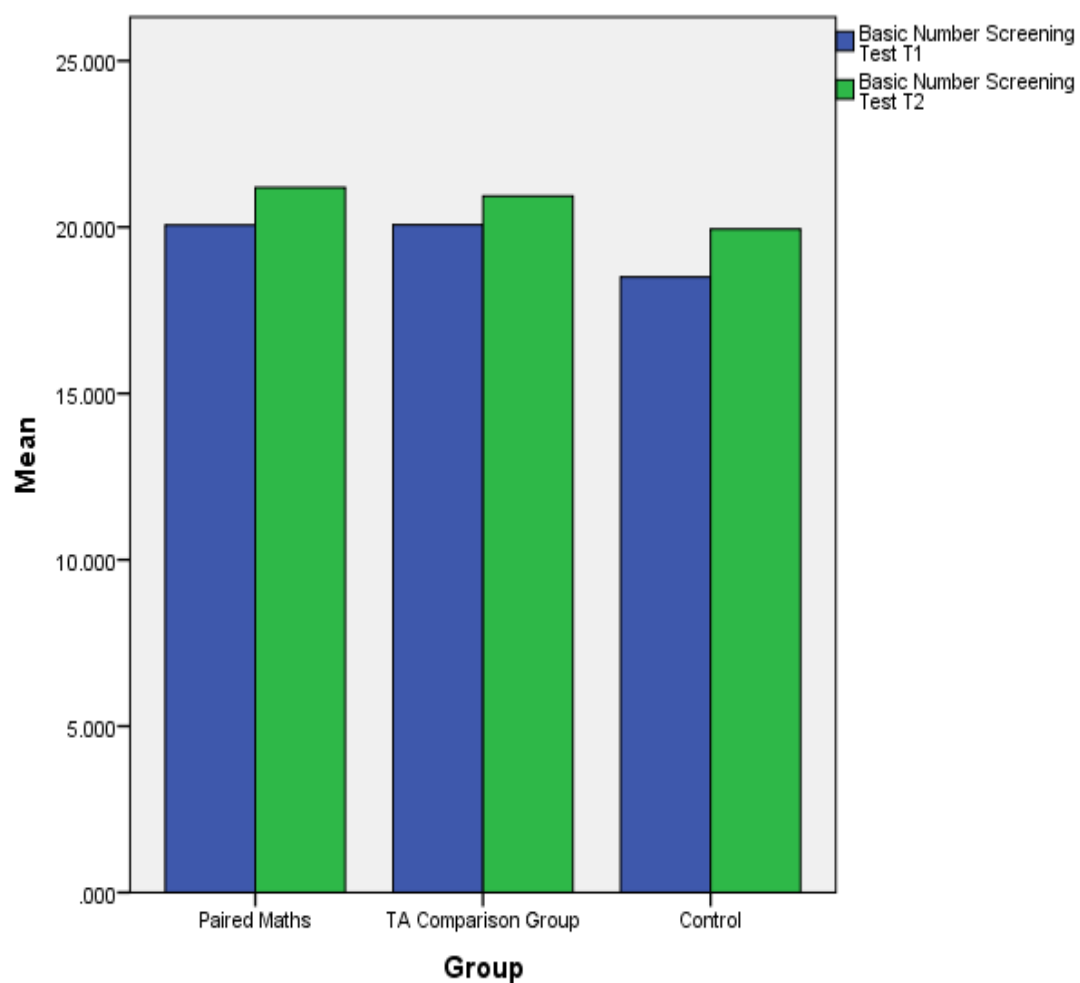


Figure 4.1: A bar chart displaying the mean raw scores, by group at pre (T1) and post (T2) testing for Basic Number Screening Test Scores for Year 8 Participants

4.1.1.2 Assumption Testing

4.1.1.2.1 Normality

Figure 4.2 illustrates the distribution of BNST raw scores at T1.

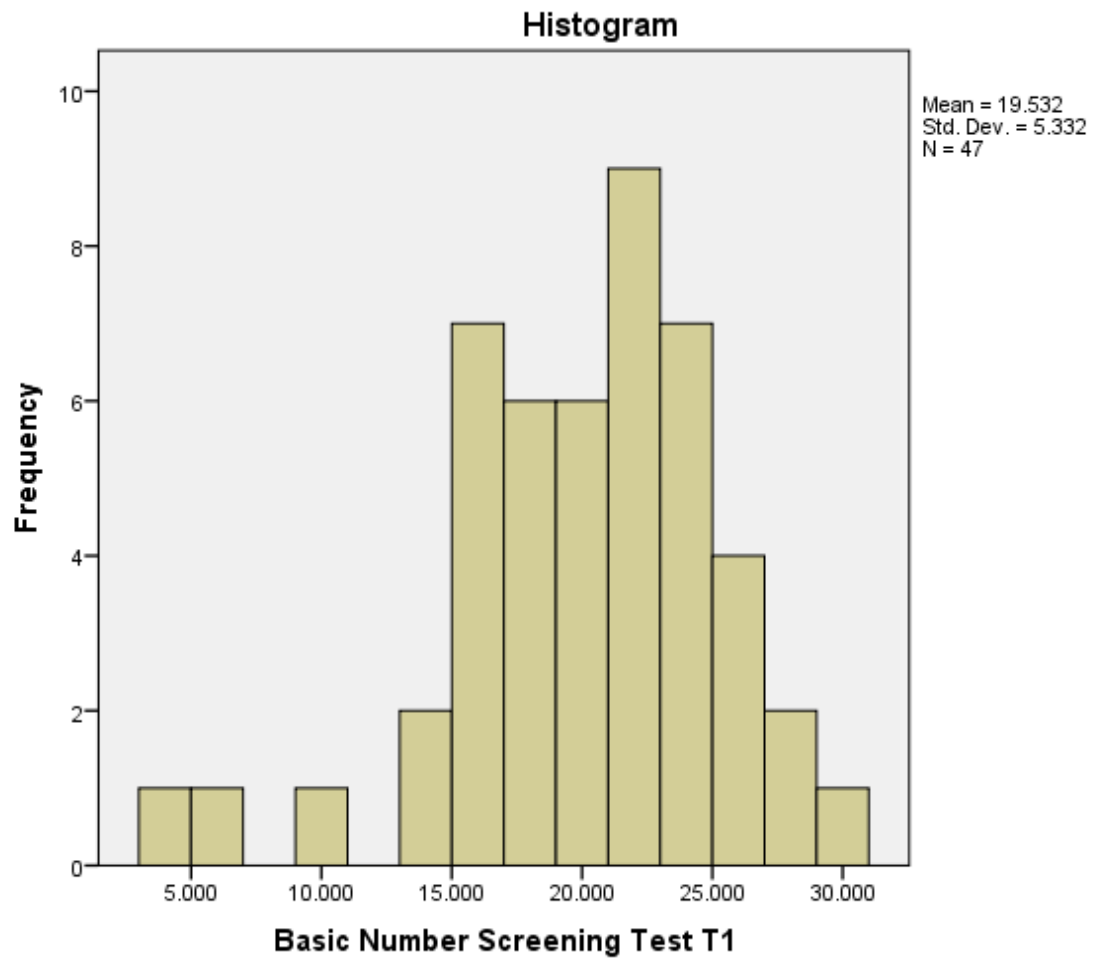


Figure 4.2: Figure showing the distribution of Basic Number Screening Test raw scores at T1

The Shapiro Wilk statistic ($p > 0.05$) is reported in table 4.2. This was not statistically significant. BNST data at T1 was therefore assumed to be normally distributed.

Shapiro-Wilk		
Statistic	Degree of Freedom	Significant
0.946	47	0.029

Table 4.2: A table displaying the results of the Shapiro-Wilk test for Raw BNST Scores at T1.

4.1.1.2.2 Homogeneity of variance

To check homogeneity of variance, Levene's Test of Equality Variances (table 4.4) was conducted on the DV: raw Basic Number Screening Test scores at T2. This was not statistically significant ($p > 0.05$). The groups were therefore considered to have homogeneity of variance.

Levene's Test		
Statistic	Degree of Freedom	Significant
0.706	44	0.499

Table 4.3: A table displaying the result of Levene's test of equality variances for BNST scores at T2

4.1.1.2.3 Linearity

The relationship between the DV and CV (Covariate), by group, is displayed in figure 4.3. As the relationship at each level is linear (and not curvilinear), (Pallant, 2013), the data was considered acceptable for ANCOVA.

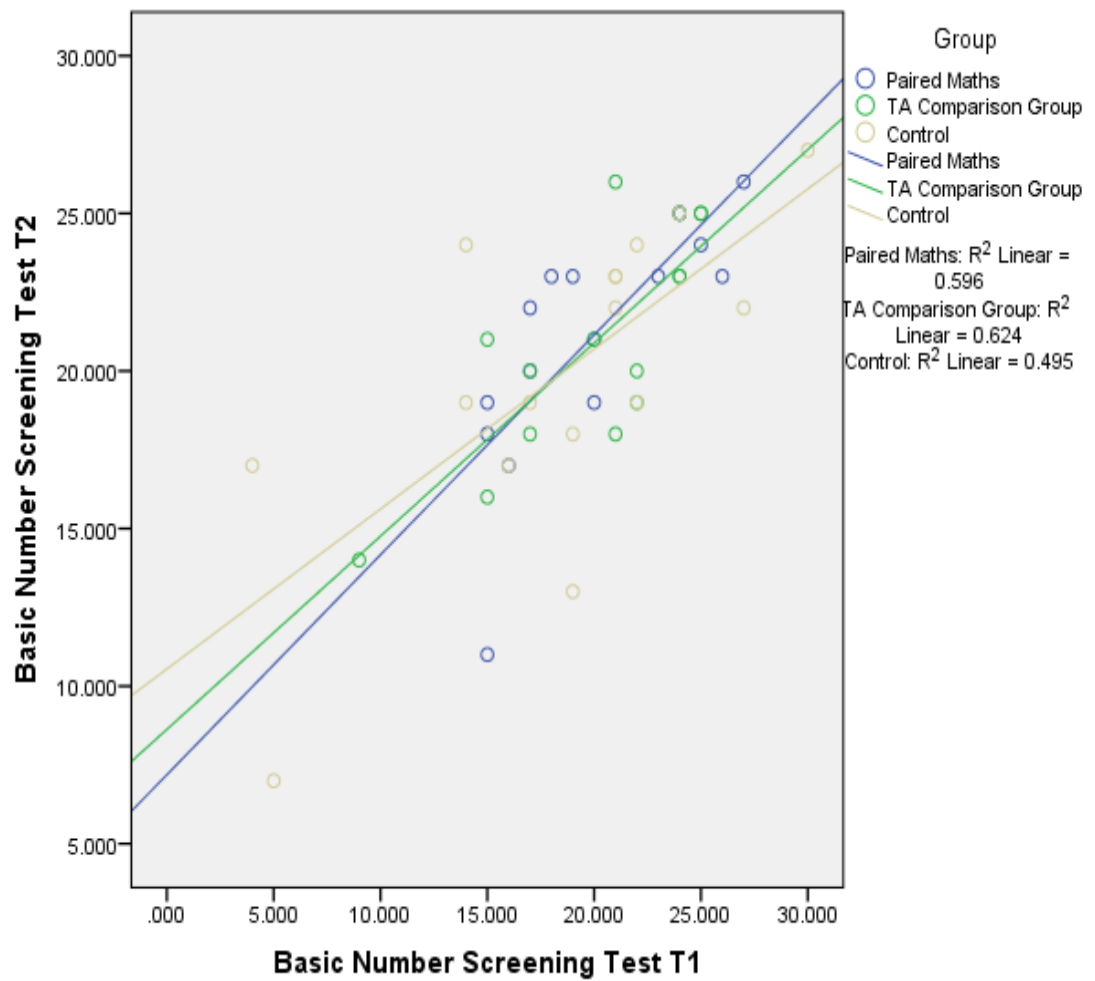


Figure 4.3: A scatter graph displaying the relationship between the DV (T2 BNST scores) and CV (T1 BNST scores), by group

4.1.1.2.4 Homogeneity of regression

Homogeneity of regression examines whether there is an interaction between the covariate and treatment groups (Pallant, 2013). Statistical analysis indicated that the test for homogeneity of regression was not statistically significant ($F(2, 47) = 0.456$, $p = 0.637$). This satisfies the requirement to perform an ANCOVA.

4.1.1.3 Inferential statistics

An ANCOVA was conducted to assess the differences in BNST scores at T2 (DV) between groups, after controlling for pre-test differences (CV). The results showed that there was a statistically non-significant effect between groups, $F(2, 47) = .065$, $p = .937$. Output tables and adjusted mean scores when considered against the covariate are displayed in tables 4.4 and 4.5 respectively.

Tests of Between-Subjects Effects

Dependent Variable: Basic Number Screening Test T2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	432.339 ^a	3	144.113	17.611	.000	.551
Intercept	283.918	1	283.918	34.696	.000	.447
BNST1	418.435	1	418.435	51.134	.000	.543
Group	1.072	2	.536	.065	.937	.003
Error	351.874	43	8.183			
Total	20886.000	47				
Corrected Total	784.213	46				

a. R Squared = .551 (Adjusted R Squared = .520)

Table 4.4: A table displaying the results of the ANCOVA assessing the impact of Paired Maths on Basic Number Screening Test scores for Year 8 participants.

Group

Dependent Variable: Basic Number Screening Test T2

Group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Paired Maths	20.884 ^a	.716	19.440	22.329
TA Comparison Group	20.628 ^a	.740	19.136	22.120
Control	20.527 ^a	.720	19.075	21.979

a. Covariates appearing in the model are evaluated at the following values:
Basic Number Screening Test T1 = 19.53191.

Table 4.5: A table displaying the adjusted means of Basic Number Screening Test Scores in relation to the CV.

4.1.2 Summary: Basic Number Screening Test

The descriptive statistics presented highlight some very small differences between groups, favouring the Control Group in BNST scores at T2. However, this is no longer evident after controlling for pre-test differences and the difference between groups was not found to be statistically significant.

4.2 Summary Research Question 1

While all groups showed increases in scores on the BNST, the descriptive statistics show a very small difference between the groups, suggesting that the Control increased slightly more. These differences were not, however, statistically significant.

The findings indicate that one cannot reject the null hypothesis, which states: *After completion of a PM intervention, there will be no statistically significant differences in mathematics attainment between Year 8 students in the PM group, a comparison group and/or control group.*

4.3 Research Question 2: In a UK secondary school, does a 7 week, cross-age, structured peer-tutor programme around maths (Paired Maths) enhance self-esteem in Year 8 students in the role of Tutees?

The DVs in relation to *Research Question 2* are the pupils' self-report scores on the Burnett Self Scale (BSS) (split into three subscales of Global Self Esteem, Learning Self-Concept, and Mathematics Self-Concept) and scores from the Behavioural Indicators of Self Esteem instrument (BIOS), as completed by their maths teachers. The IV is time (T1 and T2).

The hypotheses associated with this research question are outlined below:

Experimental Hypothesis: *Year 8 Participants in the role of the Tutee in the PM intervention group will have significantly higher gains in measures of self-esteem than those in a comparison group and/or control group.*

Null hypothesis 1: *After completion of an PM intervention, there will be no statistically significant differences in measures of self-esteem between Year 8 participants in the role of the tutee in the PM group, a comparison group and/or control group.*

4.3.1 Behavioural Indicator of Self-esteem (BIOS)

4.3.1.1 Descriptives

Table 4.6 displays the means and standard deviations (SD) of the average BIOS score by group at pre (T1) and post (T2) testing. The data shows that while the alternative and control groups showed an increase in BIOS scores across the experimental period, (+0.207) and (+0.164) the PM group showed a very slight decrease (-0.04).

Figure 4.4 presents this data using a bar chart.

Time	Group	Mean	Standard Deviation
Pre Test (T1)	Experimental (Paired Maths) (n=16)	3.200	0.512
	Comparison (n=15)	3.314	0.852
	Control (n=16)	3.365	0.984
	Total (n=47)	3.292	0.792
Post Test (T2)	Experimental (Paired Maths) (n=16)	3.196	0.373
	Comparison (n=15)	3.521	0.712
	Control (n=16)	3.529	0.700
	Total (n=47)	3.528	0.623

Table 4.6: A table displaying the means and standard deviations, by group, at pre (T1) and post (T2) testing for BIOS scores for Year 8 participants

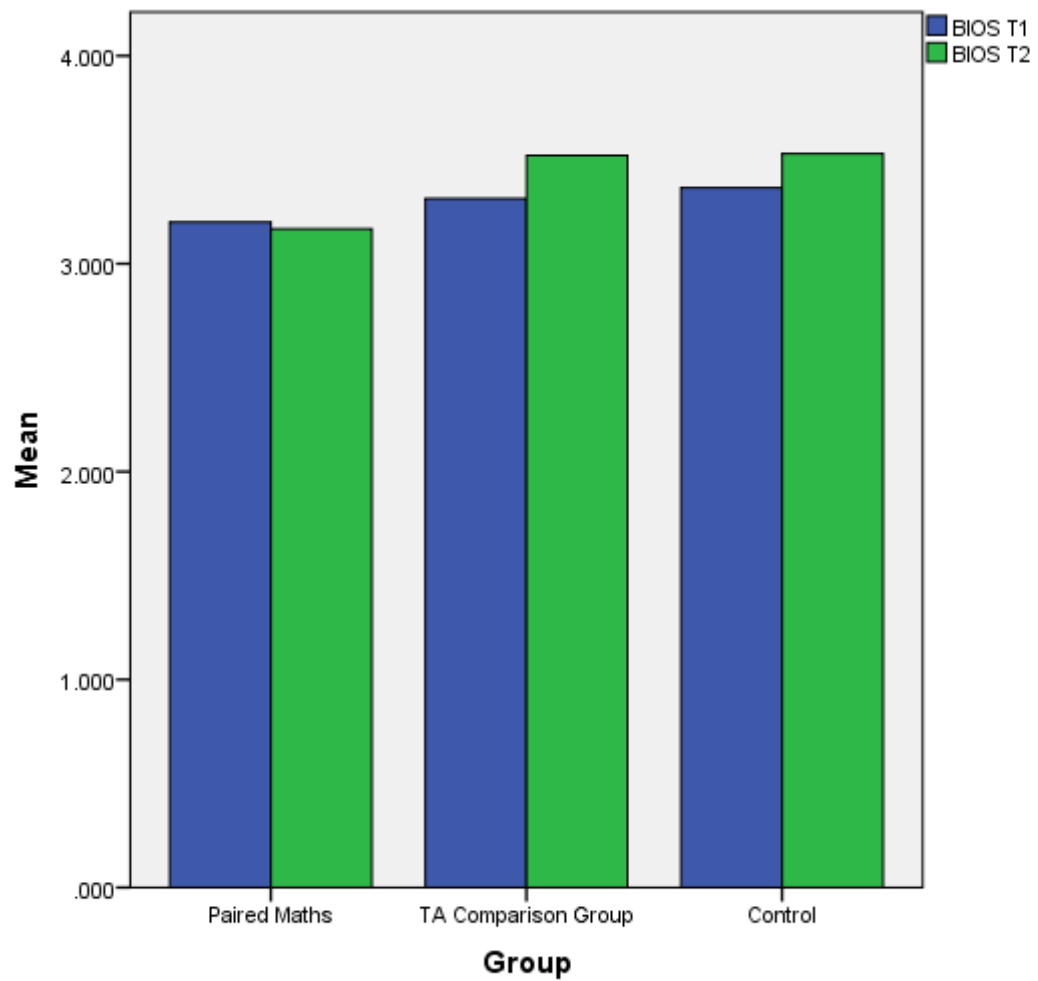


Figure 4.4: A bar chart displaying the mean, by group at pre (T1) and post (T2) testing for BIOS scores for Year 8 Participants

4.3.1.2 Assumption Testing

4.3.1.2.1 Normality

Figure 4.5 illustrates the distribution of BIOS raw scores at T1.

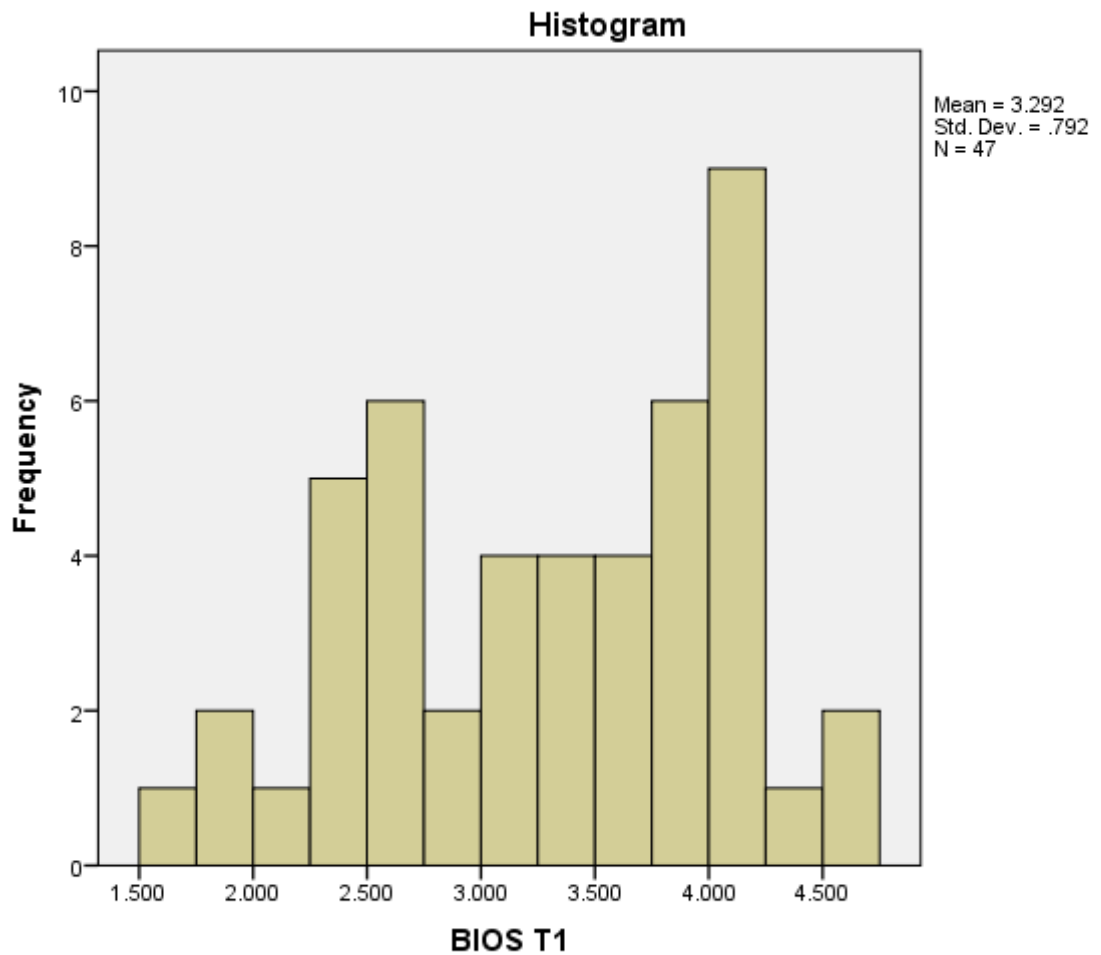


Figure 4.5: Figure showing the distribution of BIOS scores at T1

The Shapiro Wilk statistic ($p > 0.05$) is reported in table 4.7. This was not statistically significant. Data from the BIOS for Year 8 participants at T1 was therefore assumed to be normally distributed.

Shapiro-Wilk		
Statistic	Degree of Freedom	Significant
0.962	47	0.128

Table 4.7: A table displaying the results of the Shapiro-Wilk test for average BIOS scores at T1 for Year 8 participants

4.3.1.2.2 Homogeneity of variance

Levene's Test of Equality Variances (table 4.8) was not statistically significant ($p > 0.05$) when conducted on the DV: scores from BIOS at T2. The groups were therefore considered to have homogeneity of variance.

Levene's Test of Homogeneity of Variance		
Statistic	Degree of Freedom	Significant
1.619	44	.210

Table 4.8: A table displaying the result of Levene's test of equality variances for BIOS scores at T2.

4.3.1.2.3 Linearity

The relationship between the DV and CV, by group, is displayed in figure 4.6. As the relationship at each level is linear (and not curvilinear), (Pallant, 2013), the data was considered appropriate for ANCOVA.

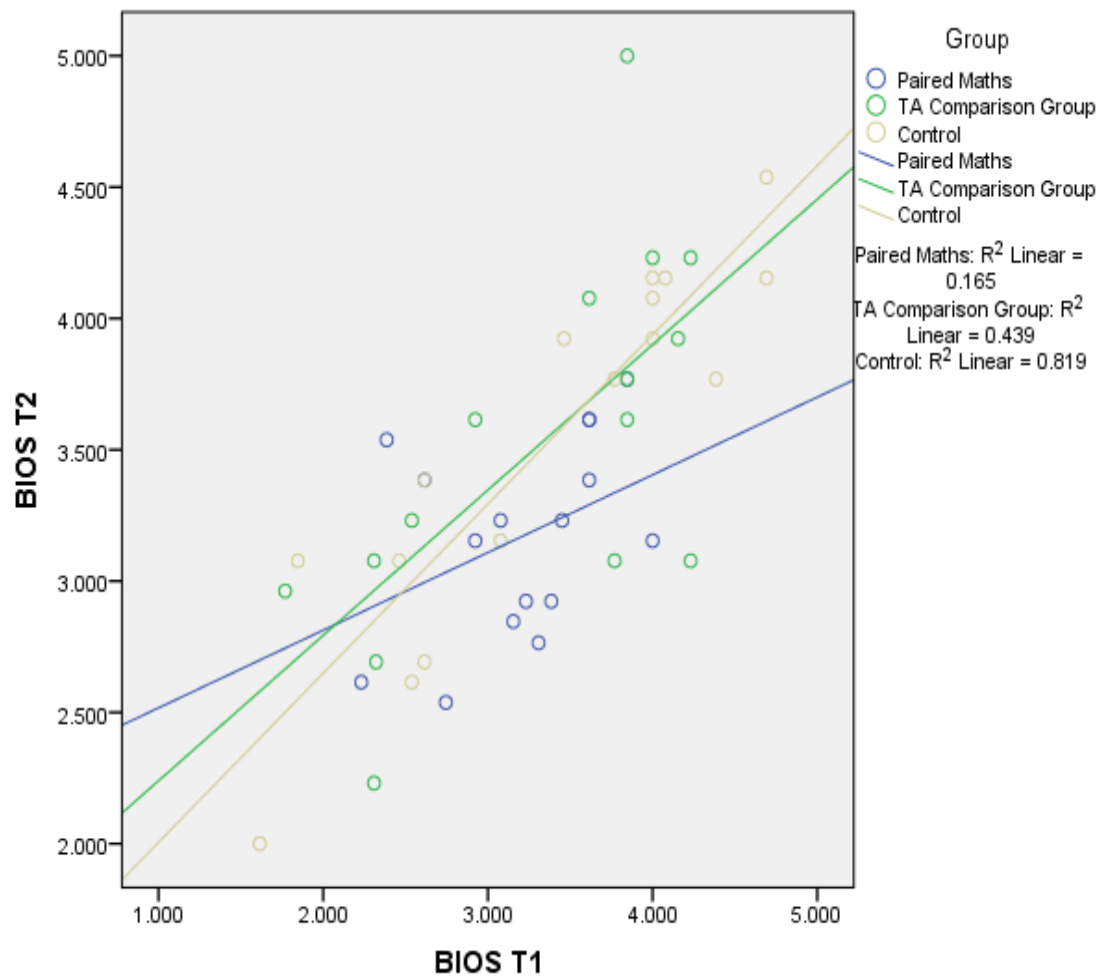


Figure 4.6: A scatter graph displaying the relationship between the DV (T2 BIOS scores) and CV (T1 BIOS scores), by group.

4.3.1.2.4 Homogeneity of regression slopes

Statistical analysis indicated that the test for homogeneity of regression was not statistically significant ($F(2, 47) = 1.097, p = .344$). This satisfies the requirement to perform an ANCOVA.

4.3.1.3 Inferential statistics

An ANCOVA was conducted to assess the differences in BIOS scores at T2 (DV) between groups, after controlling for pre-test differences (CV). The results showed that there was a statistically non-significant effect between groups, $F(2, 47) = 2.347, P = .108$. Output tables and adjusted mean scores when considered against the covariate are displayed in tables 4.9 and 4.10 respectively.

Tests of Between-Subjects Effects

Dependent Variable: BIOS T2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	10.452 ^a	3	3.484	20.172	.000	.585
Intercept	5.988	1	5.988	34.671	.000	.446
BIOS T1	9.107	1	9.107	52.730	.000	.551
Group	.811	2	.405	2.347	.108	.098
Error	7.427	43	.173			
Total	562.260	47				
Corrected Total	17.878	46				

a. R Squared = .585 (Adjusted R Squared = .556)

Table 4.9: A table displaying the results of the ANCOVA assessing the impact of Paired Maths on BIOS scores for Year 8 participants

Group

Dependent Variable: BIOS T2

Group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Paired Maths	3.220 ^a	.104	3.010	3.430
TA Comparison Group	3.509 ^a	.107	3.292	3.725
Control	3.488 ^a	.104	3.278	3.698

a. Covariates appearing in the model are evaluated at the following values: BIOS T1 = 3.29245.

Table 4.10: A table displaying the adjusted means of BIOS Scores in relation to the CV.

4.3.1.4 Summary: BIOS for Year 8s

The descriptive statistics presented showed a small difference between groups, where while the Comparison and Control groups showed an increase in mean BIOS scores at T2, the PM group showed a slight decrease. This is evident after controlling for pre-test differences (table 4.10). However, the difference between groups was not found to be statistically significant.

4.3.2 Burnett Self Scale

The Burnett Self Scale (BSS) involves three subscales: Global Self Esteem, Learning Self Concept and Maths Self Concept, and these will be considered each in turn.

4.3.3 Global Self-Esteem Scale

Table 4.11 displays the means and SDs of the Global Self-Esteem scores by group, at pre (T1) and post (T2) testing. The data shows a small decrease in all groups over time, with the PM group showing a slightly larger decrease: Experimental (PM) (-0.05), Comparison (-1.15) and Control (-.43) over time.

Figure 4.7 presents this data using a bar chart.

Time	Group	Mean	Standard Deviation
Pre Test (T1)	Experimental (Paired Maths) (n=16)	3.867	0.742
	Comparison (n=15)	3.750	0.660
	Control (n=16)	3.727	0.860
	Total (n=47)	3.782	0.747
Post Test (T2)	Experimental (Paired Maths) (n=16)	3.813	0.645
	Comparison (n=15)	3.725	1.005
	Control (n=16)	3.711	0.867
	Total (n=47)	3.750	0.831

Table 4.11: A table displaying the means and standard deviations, by group, at pre (T1) and post (T2) testing for Global Self-Esteem Scores for Year 8 participants

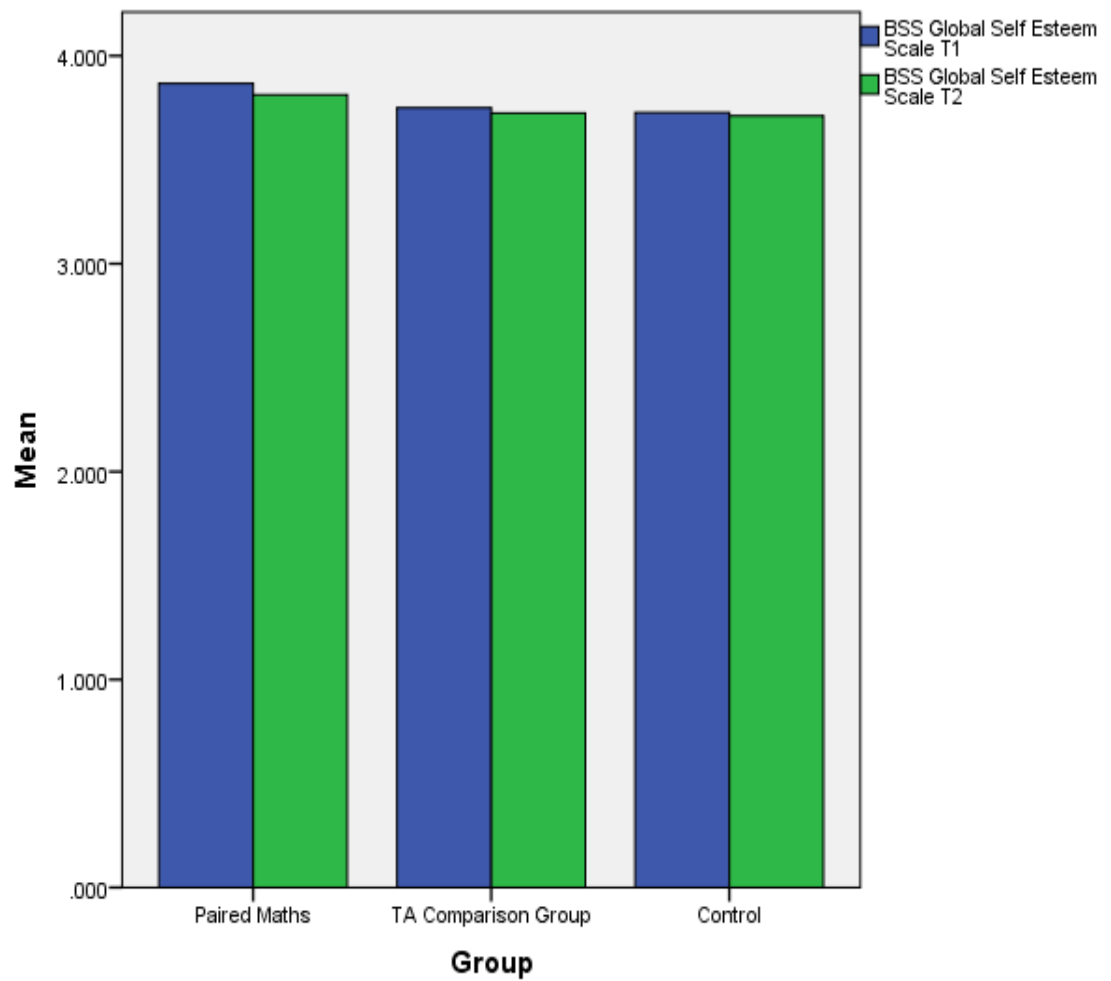


Figure 4.7: A bar chart displaying the mean, by group at pre (T1) and post (T2) testing for BSS Global Self-Esteem Scores for Year 8 Participants

4.3.3.1 Assumption Testing

4.3.3.1.1 Normality

Figure 48. illustrates the distribution of BSS Global Self-Esteem scores at T1.

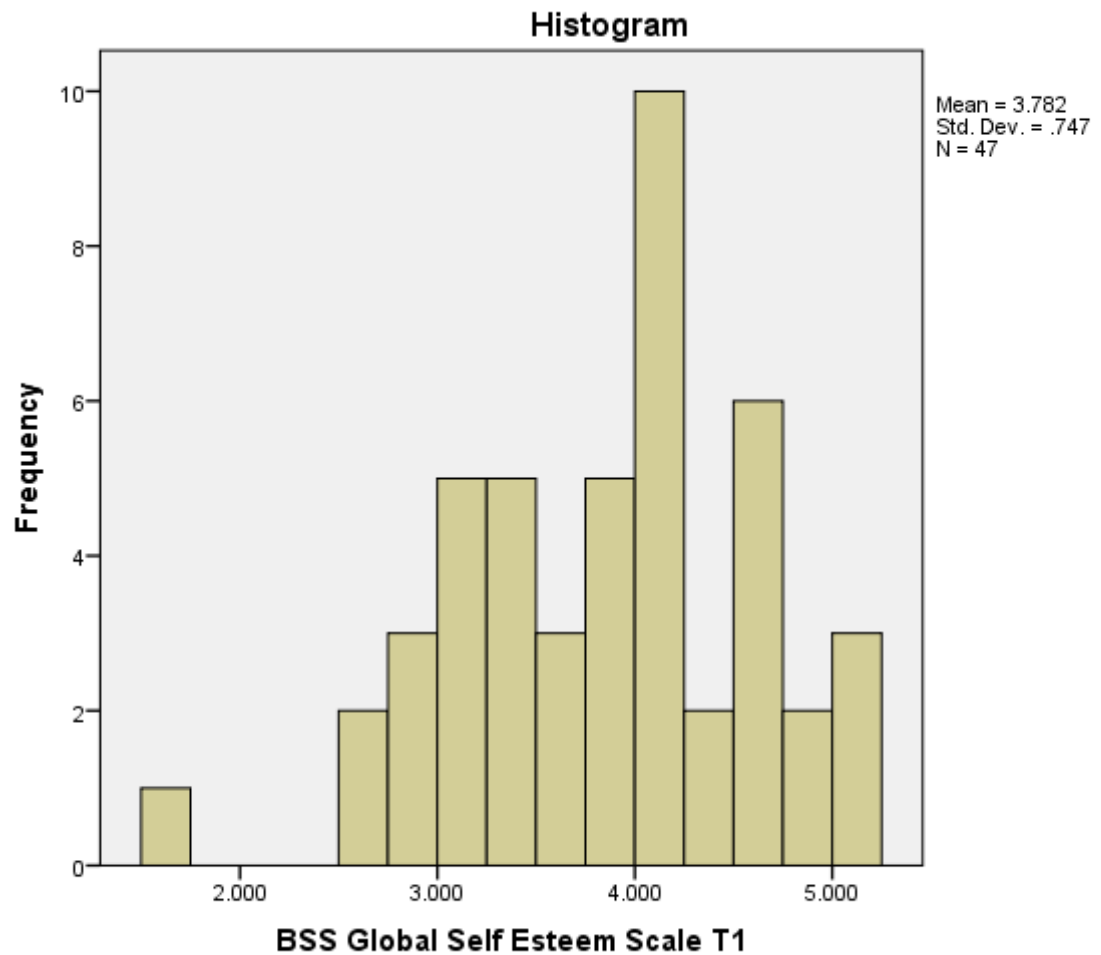


Figure 4.8: Figure showing the distribution of BSS Global Self-Esteem scores at T1

The Shapiro Wilk statistic ($p > 0.05$) is reported in table 4.12. This was not statistically significant. Global self-esteem scores for Year 8s at T1 were therefore assumed to be normally distributed.

Shapiro-Wilk		
Statistic	Degree of Freedom	Significant
0.972	47	0.307

Table 4.12: A table displaying the results of the Shapiro-Wilk test for BSS Global Self Esteem scores at T1.

4.3.3.1.2 Homogeneity of variance

Levene's Test of Equality Variances (table 4.13) was not statistically significant ($p > 0.05$) when conducted on the DV: BSS Global Self-Esteem scores at T2. The groups were therefore considered to have homogeneity of variance.

Levene's Test of Homogeneity of Variance		
Statistic	Degree of Freedom	Significant
0.787	44	.462

Table 4.13: A table displaying the result of Levene's test of equality variances for BSS Global Self-Esteem scores at T2

4.3.3.2 Linearity

The relationship between the DV and CV, by group, is displayed in figure 4.9. As the relationship at each level is linear, the data was considered appropriate for ANCOVA.

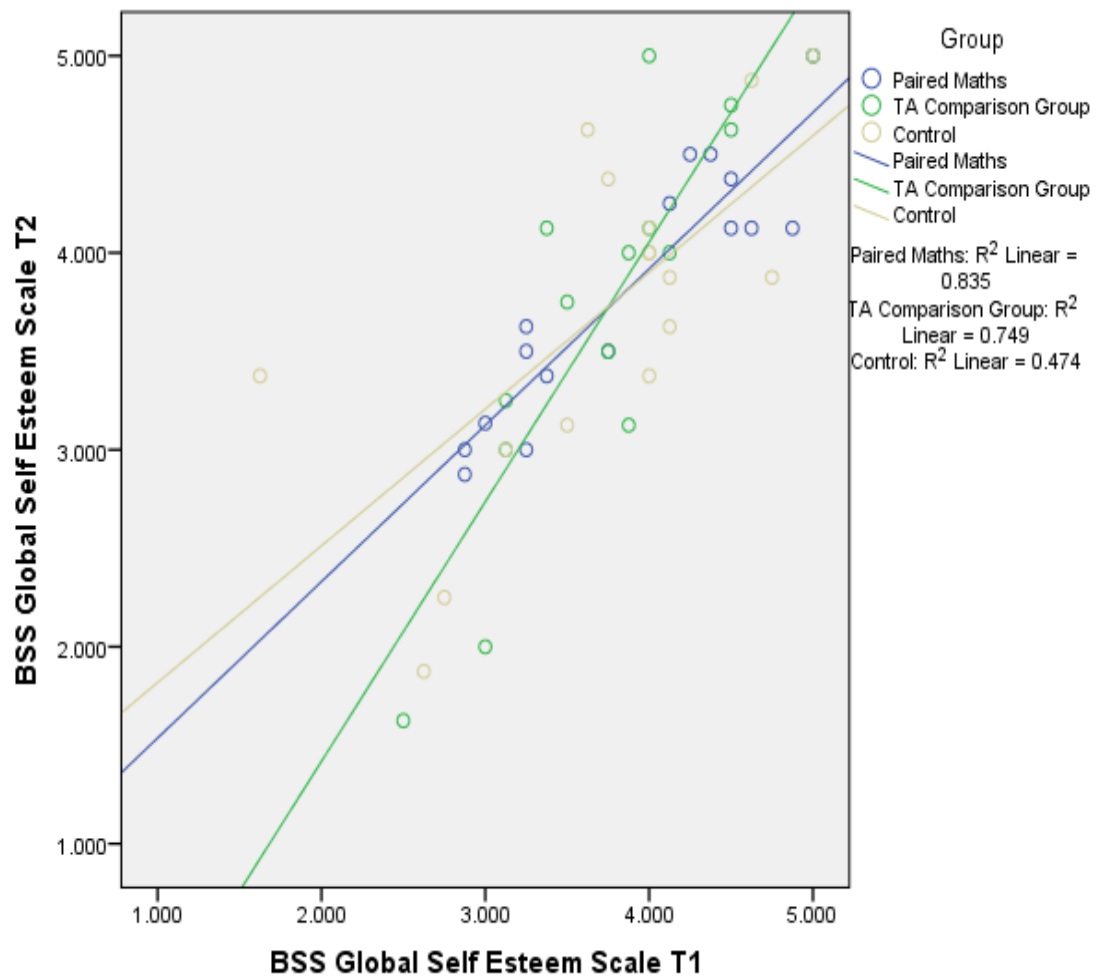


Figure 4.9:A scatter graph displaying the relationship between the DV (BSS Global Self Esteem Scores) and CV (T1 BSS Global Self Esteem Scores) by group for Year 8 Participants.

4.3.3.3 Homogeneity of regression slopes

Statistical analysis indicated that the test for homogeneity of regression was not statistically significant ($F(2, 47) = 3.150, p = 0.06$). This satisfies the requirement to perform an ANCOVA.

4.3.4 Inferential statistics

An ANCOVA was conducted to assess the differences in BSS Global Self Esteem scores at T2 (DV) between groups, after controlling for pre-test differences (CV). The results showed that there was a statistically non-significant effect between groups, $F(2, 30) = .383, P = .625$. Output tables and adjusted mean scores when considered against the covariate are displayed in tables 4.14 and 4.15 respectively.

Tests of Between-Subjects Effects

Dependent Variable: BSS Global Self Esteem Scale T2

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	19.625 ^a	3	6.542	23.224	.000	.618
Intercept	.329	1	.329	1.169	.286	.026
BSSselfT1	19.528	1	19.528	69.325	.000	.617
Group	.004	2	.002	.006	.994	.000
Error	12.112	43	.282			
Total	692.750	47				
Corrected Total	31.738	46				

a. R Squared = .618 (Adjusted R Squared = .592)

Table 4.14: A table displaying the results of the ANCOVA assessing the impact of Paired Maths on BSS Global Self Esteem Scores for Year 8 students in the role of Tutee.

Group

Dependent Variable: BSS Global Self Esteem Scale T2

Group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Paired Maths	3.738 ^a	.133	3.470	4.007
TA Comparison Group	3.753 ^a	.137	3.477	4.029
Control	3.759 ^a	.133	3.492	4.027

a. Covariates appearing in the model are evaluated at the following values: BSS Global Self Esteem Scale T1 = 3.78191.

Table 4.15: A table displaying the adjusted means of BSS Global Self Esteem Scores for Year 8 Participants.

4.3.4.1 Summary: BSS Global Self-Esteem Scale

The descriptive statistics presented highlight that all groups showed a small decrease in BSS Global Self-Esteem scores, with some very small differences between groups in scores at T2 where PM group showed a slightly larger decrease. This was no longer evident after controlling for pre-test differences (table 4.), and any differences between groups were not found to be statistically significant.

4.3.5 BSS Learning Self-Concept Scale

4.3.5.1.1 Descriptive statistics

Table 4.16 displays the means and SDs of the Learning Self Concept Scores by group, at pre (T1) and post (T2) testing. The data shows that while the PM and Comparison groups decreased slightly, (-0.328) and (-0.083) respectively, the Control group increased (+0.047) in BSS Learning Self-Concept scores over time.

Figure 4.10 presents this data using a bar chart.

Time	Group	Mean	Standard Deviation
Pre Test (T1)	Experimental (Paired Maths) (n=16)	3.703	0.672
	Comparison (n=15)	3.850	0.749
	Control (n=16)	3.297	0.843
	Total (n=47)	3.612	0.778
Post Test (T2)	Experimental (Paired Maths) (n=16)	3.375	0.524
	Comparison (n=15)	3.767	0.970
	Control (n=16)	3.344	0.724
	Total (n=47)	3.489	0.764

Table 4.16: A table displaying the means and standard deviations, by group, at pre (T1) and post (T2) testing for Learning Self-Concept Scores for Year 8 participants

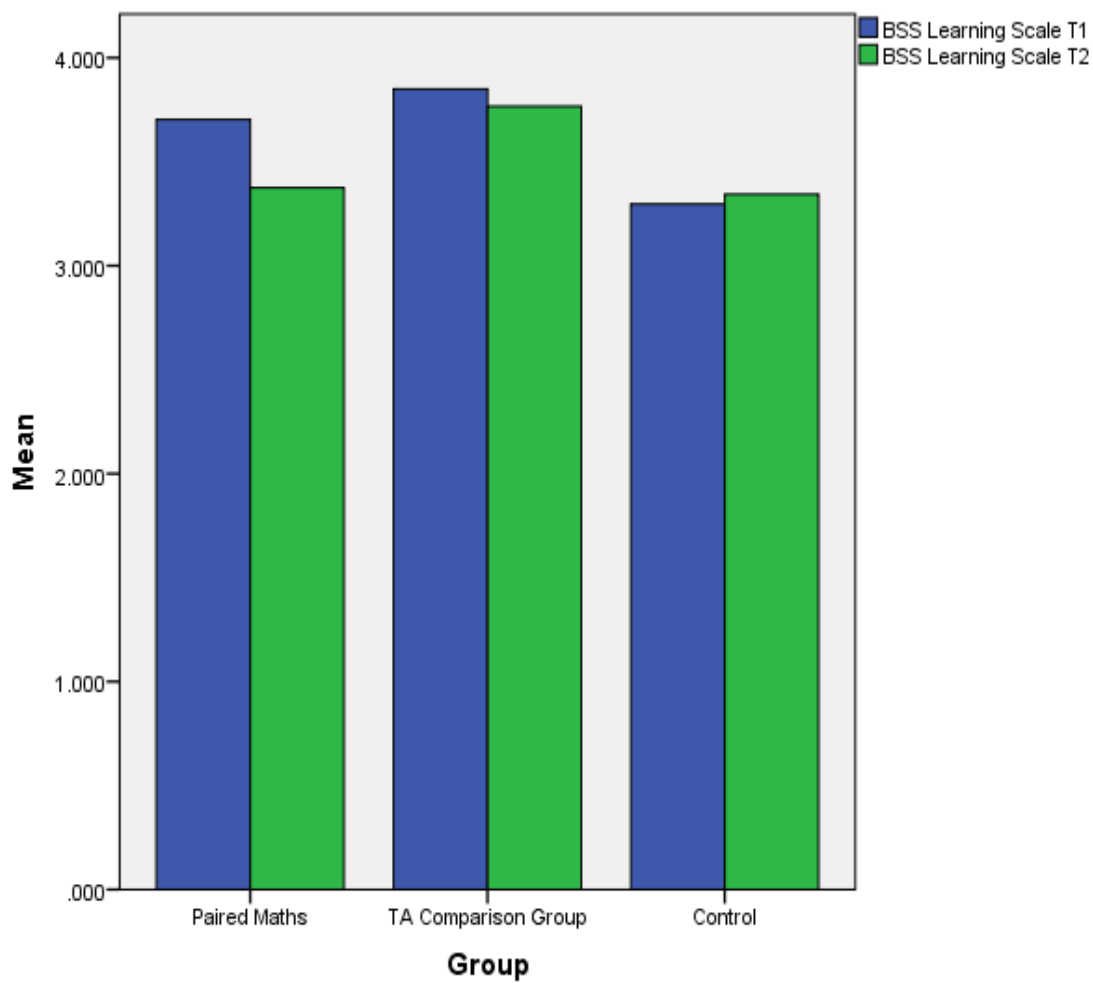


Figure 4.10: A bar chart displaying the mean, by group at pre (T1) and post (T2) testing for BSS Learning Self-Concept Scores for Year 8 Participants

4.3.5.2 Assumption Testing

4.3.5.2.1 Normality

Figure 4.11 illustrates the distribution of BSS Learning Self -Concept Scores at T1.

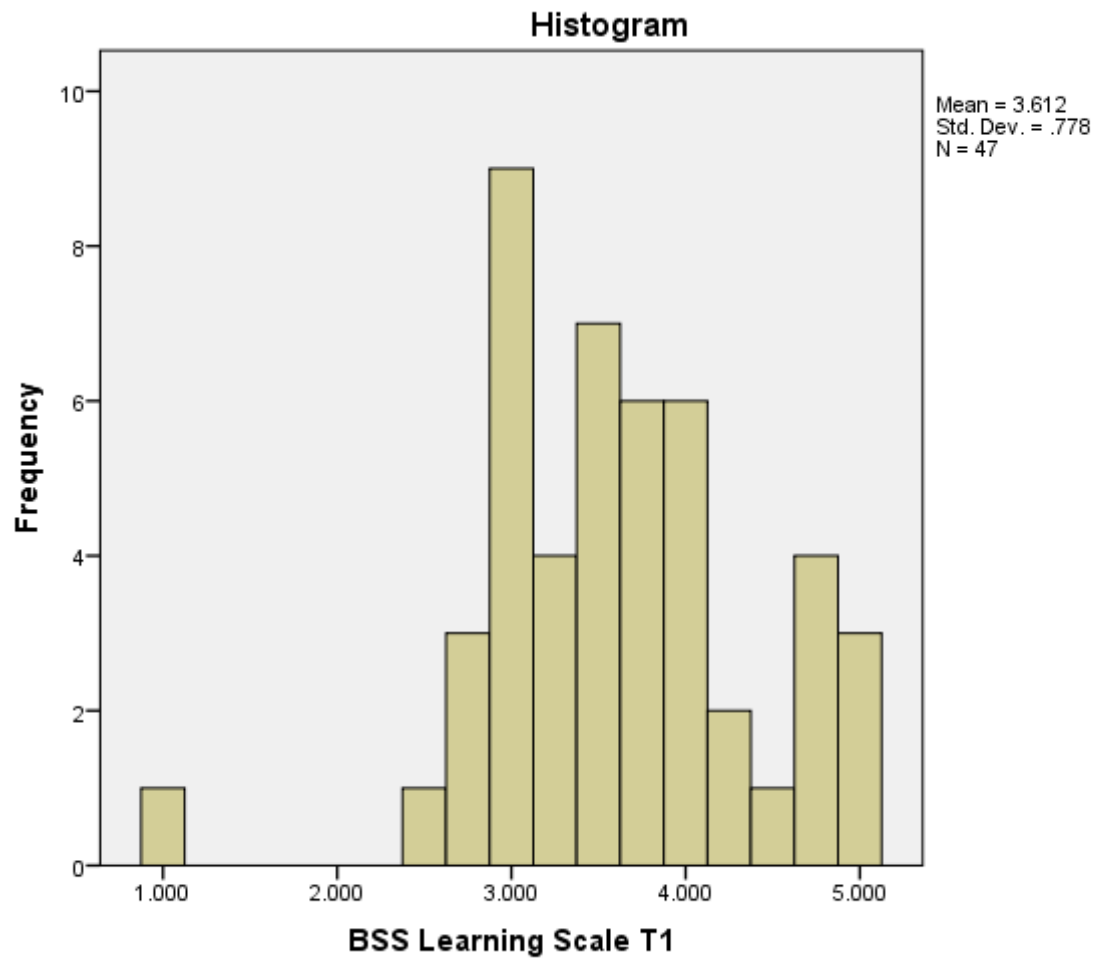


Figure 4.11: Figure showing the distribution of BSS Learning Self-Concept scores at T1

The Shapiro Wilk statistic ($p > 0.05$) is reported in table 4.12. This was statistically significant, and therefore BSS Learning Self-Concept Scores at T1 cannot be assumed to be normally distributed. However, as Vickers (2005) has argued that ANCOVAs should still be used in RCTs even when data cannot be assumed to be normally distributed, the researcher went on to examine whether the other assumptions could be met for an ANCOVA to be carried out.

Shapiro-Wilk		
Statistic	Degree of Freedom	Significant
0.939	47	0.017

Figure 4.12: : A table displaying the results of the Shapiro-Wilk test for BSS Learning Self-Concept scores at T1 For Year 8.

4.3.5.2.2 Homogeneity of variance

Levene's Test of Equality Variances (table 4.17) was not statistically significant ($p > 0.05$) when conducted on the DV: BSS Learning Self-Concept Scores at T2. The groups were therefore considered to have homogeneity of variance.

Levene's Test of Homogeneity of Variance		
Statistic	Degree of Freedom	Significant
2.757	44	0.074

Table 4.17: A table displaying the results of Levene's test of equality variances for BSS Learning Self-Concept Scores at T2.

4.3.5.2.3 Linearity

The relationship between the DV and CV, by group, is displayed in figure 4.13. As the relationship at each level is linear, the data was therefore considered appropriate for ANCOVA.

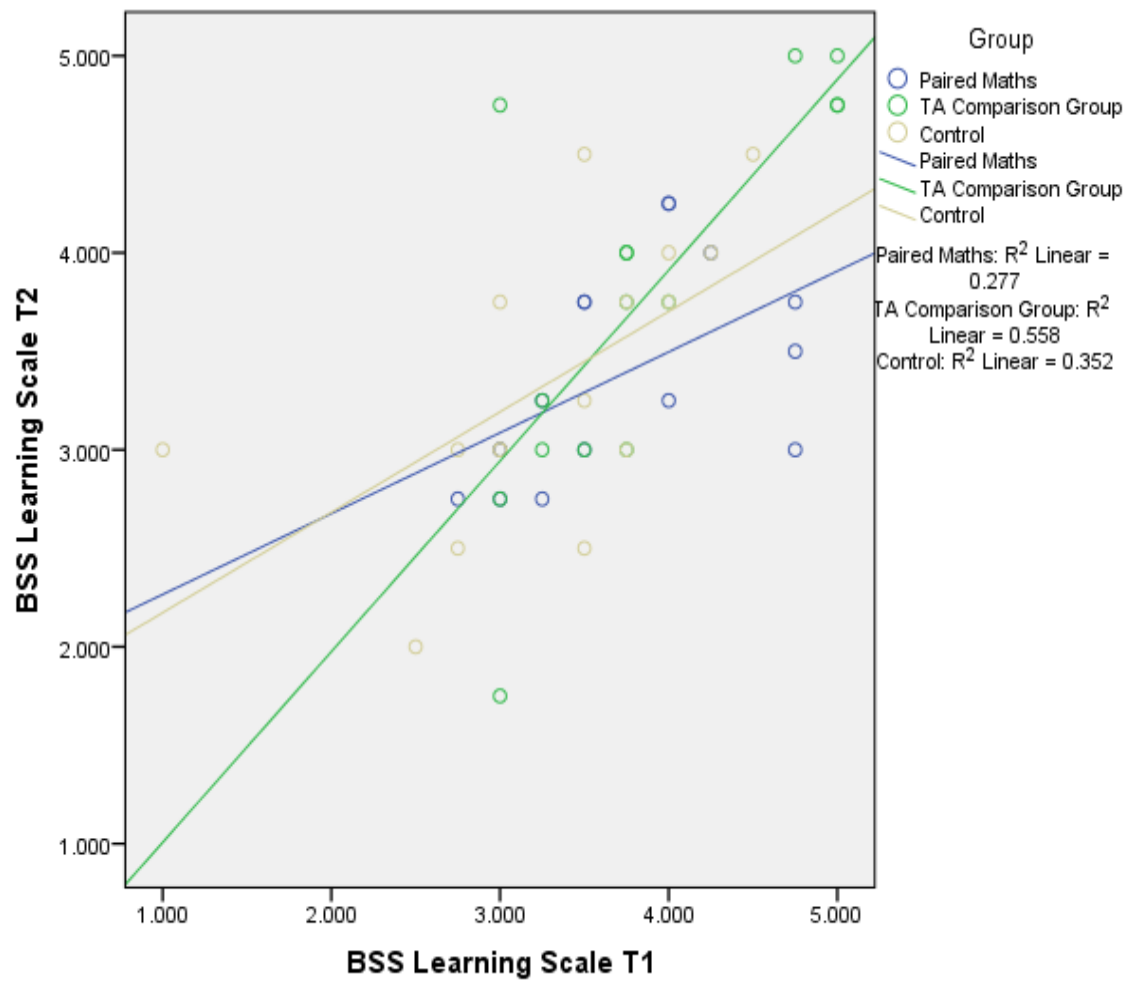


Figure 4.13: A scatter graph displaying the relationship between the DV (BSS Learning Self-Concept Scores) and CV (T1 BSS Learning Self-Concept Scores) by group for Year 8 Participants.

4.3.5.2.4 Homogeneity of regression slopes

Statistical analysis indicated that the test for homogeneity of regression was not statistically significant ($F(2, 47) = 2.032, p = 0.144$). This satisfies the requirement to perform an ANCOVA.

4.3.5.3 Inferential statistics

An ANCOVA was conducted to assess the differences in BSS Learning Self-Concept scores at T2 (DV) between groups, after controlling for pre-test differences (CV). The results showed that there was a statistically non-significant effect between groups, $F(2, 47) = 1.069, P = .352$. Output tables and adjusted mean scores when considered against the covariate are displayed in tables 4.18 and 4.19 respectively.

Tests of Between-Subjects Effects

Dependent Variable: BSS Learning Scale T2

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	11.591 ^a	3	3.864	10.873	.000	.431
Intercept	2.823	1	2.823	7.944	.007	.156
BSSlearningT1	9.889	1	9.889	27.830	.000	.393
Group	.760	2	.380	1.069	.352	.047
Error	15.279	43	.355			
Total	599.125	47				
Corrected Total	26.870	46				

a. R Squared = .431 (Adjusted R Squared = .392)

Table 4.18: A table displaying the results of the ANCOVA assessing the impact of Paired Maths on BSS Learning Self-Concept scores for Year 8 students in the role of Tutee.

Group

Dependent Variable: BSS Learning Scale T2

Group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Paired Maths	3.318 ^a	.149	3.016	3.619
TA Comparison Group	3.618 ^a	.156	3.302	3.933
Control	3.541 ^a	.154	3.231	3.850

a. Covariates appearing in the model are evaluated at the following values: BSS Learning Scale T1 = 3.61170.

Table 4.19: A table displaying the adjusted means of BSS Learning Self Concept Scores for Year 8 Participants.

4.3.5.3.1 Summary: BSS Learning Self-Concept Scale

The descriptive statistics presented showed a small difference between groups, where it appeared the PM and Comparison groups show small decreases while the control group showed a small increase in BSS Learning Self Concept scores at T2. This was no longer evident after controlling for pre-test differences and no difference between groups were found to be statistically significant however.

4.3.6 BSS Mathematics Self-Concept Scale

4.3.6.1 Descriptive statistics

Table 4.20 displays the means and SDs of the BSS Mathematics Self-Concept scores by group, at pre (T1) and post (T2) testing. The data shows that while the comparison and control groups showed a slight increase over time, (+0.167) and (+0.078) respectively, the PM group appears to have shown a slight decrease (-0.031) in BSS Mathematics Self-Concept scores.

Figure 4.14 presents this data using a bar chart.

Time	Group	Mean	Standard Deviation
Pre Test (T1)	Experimental (Paired Maths) (n=16)	3.109	1.106
	Comparison (n=15)	3.016	1.058
	Control (n=16)	2.969	1.091
	Total (n=47)	3.032	1.034
Post Test (T2)	Experimental (Paired Maths) (n=16)	3.078	0.820
	Comparison (n=15)	3.183	1.304
	Control (n=16)	3.047	1.046
	Total (n=47)	3.101	1.045

Table 4.20: A table displaying the means and standard deviations, by group, at pre (T1) and post (T2) testing for Mathematics Self-Concept Scores for Year 8 participants

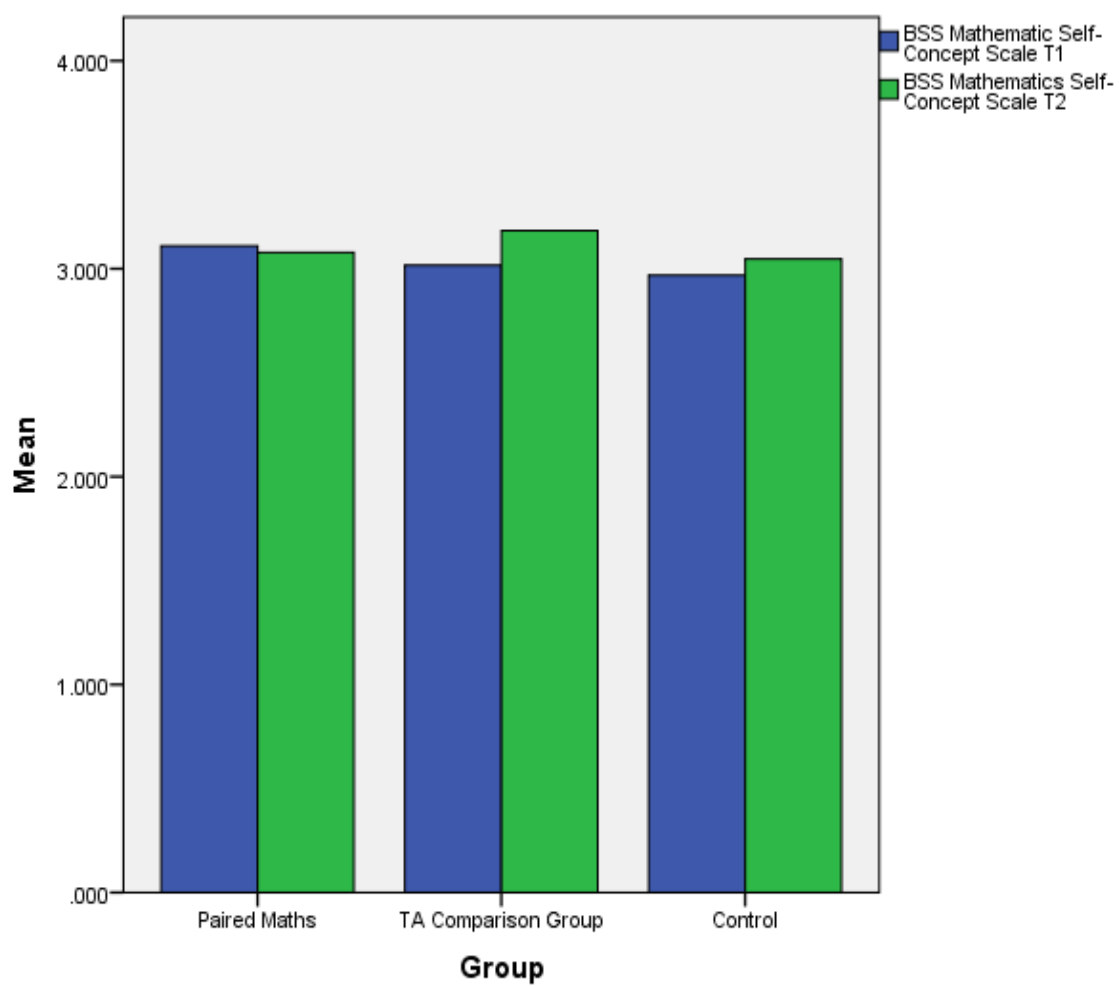


Figure 4.14: A bar chart displaying the mean, by group at pre (T1) and post (T2) testing for BSS Mathematics Self-Concept scores for Year 8 Participants

4.3.6.2 Assumption Testing

4.3.6.2.1 Normality

Figure 4.15 illustrates the distribution of BSS Mathematics Self-Concept Scores at T1.

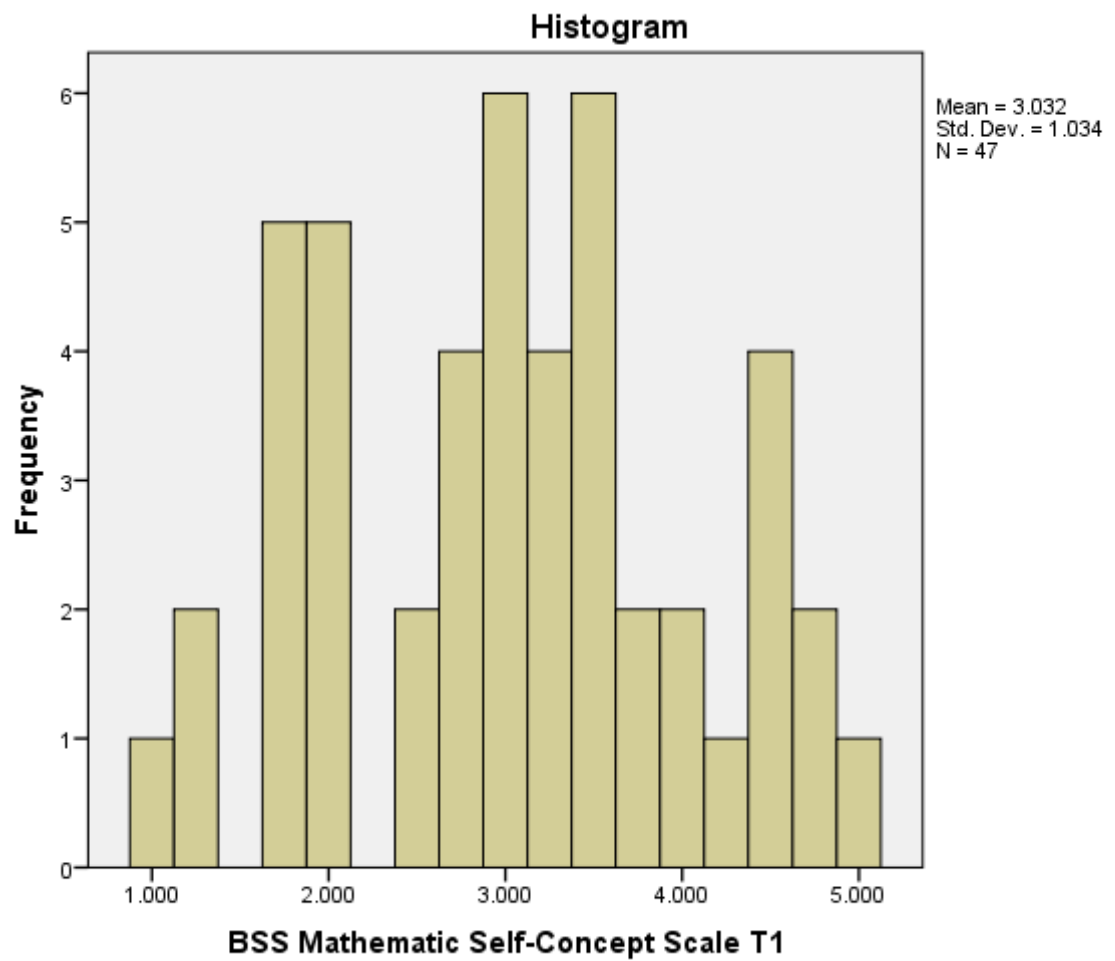


Figure 4.15: Figure showing the distribution of BSS Maths Self-Concept scores at T1

The Shapiro Wilk statistic ($p > 0.05$) is reported in table 4.21. This was not statistically significant. BSS Mathematics Self-Concept scores at T1 were therefore assumed to be normally distributed.

Shapiro-Wilk		
Statistic	Degree of Freedom	Significant
0.968	47	0.226

Table 4.21: A table displaying the results of the Shapiro-Wilk test for BSS Mathematics Self-Concept scores at T1 For Year 8.

.

4.3.6.2.2 Homogeneity of variance

Levene's Test of Equality Variances (table 4.22) was not statistically significant ($p > 0.05$) when conducted on the DV: BSS Mathematics Self-Concept scores at T2. The groups were therefore considered to have homogeneity of variance.

Levene's Test of Homogeneity of Variance		
Statistic	Degree of Freedom	Significant
1.902	44	.161

Table 4.22: A table displaying the results of Levene's test of equality variances for BSS Mathematics Self-Concept scores for Year 8 at T2.

4.3.6.2.3 Linearity

The relationship between the DV and CV, by group, is displayed in figure 4.16. As the relationship at each level is linear, the data was considered appropriate for ANCOVA.

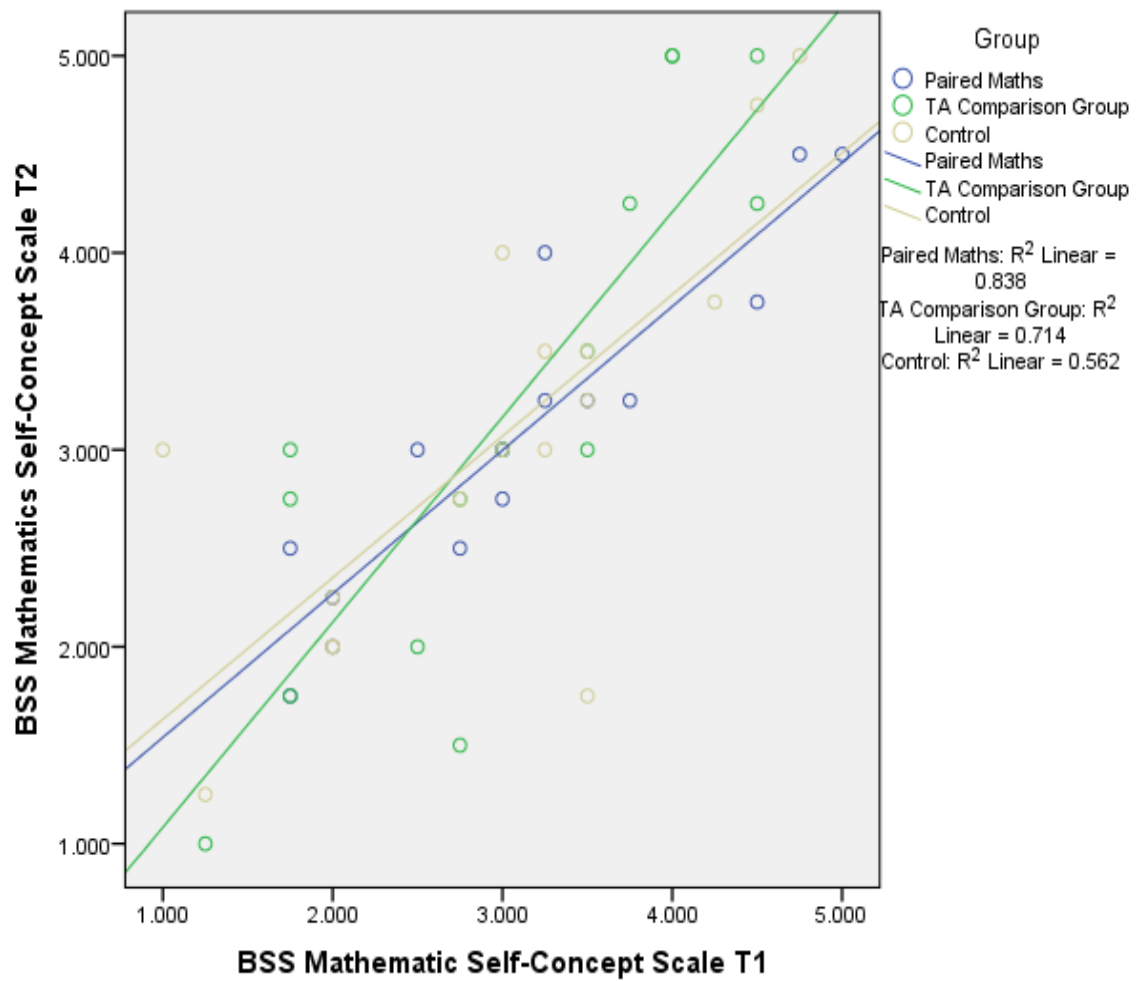


Figure 4.16: A scatter graph displaying the relationship between the DV (BSS Mathematics Self-Concept Scores) and CV (T1 BSS Mathematics Self Concept Scores) by group for Year 8 Participants.

4.3.6.2.4 Homogeneity of regression slopes

Statistical analysis indicated that the test for homogeneity of regression was not statistically significant ($F(2, 47) = 1.416, p = 0.254$). This satisfies the requirement to perform an ANCOVA.

4.3.6.3 Inferential statistics

An ANCOVA was conducted to assess the differences in BSS Mathematics Self-Concept scores at T2 (DV) between groups, after controlling for pre-test differences (CV). The results showed that there was a statistically non-significant effect between groups, $F(2, 47) = .330, P = .721$. Output tables and adjusted mean scores when considered against the covariate are displayed in tables 4.23 and 4.24 respectively.

Tests of Between-Subjects Effects

Dependent Variable: BSS Mathematics Self-Concept Scale T2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	33.549 ^a	3	11.183	28.865	.000	.668
Intercept	1.732	1	1.732	4.470	.040	.094
BSSmathsT1	33.392	1	33.392	86.191	.000	.667
Group	.255	2	.128	.330	.721	.015
Error	16.659	43	.387			
Total	502.188	47				
Corrected Total	50.207	46				

a. R Squared = .668 (Adjusted R Squared = .645)

Table 4.23: A table displaying the results of the ANCOVA assessing the impact of Paired Maths on BSS Mathematics Self-Concept scores for Year 8 students in the role of Tutee.

Group

Dependent Variable: BSS Mathematics Self-Concept Scale T2

Group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Paired Maths	3.014 ^a	.156	2.700	3.328
TA Comparison Group	3.196 ^a	.161	2.872	3.520
Control	3.099 ^a	.156	2.785	3.413

a. Covariates appearing in the model are evaluated at the following values: BSS Mathematic Self-Concept Scale T1 = 3.03191.

Table 4.24: A table displaying the adjusted means of BSS Mathematics Self-Concept Scores for Year 8 Participants

4.3.6.4 Summary: BSS Mathematics Self-Concept Scale

The descriptive statistics presented showed a small difference between groups, where it appeared the Comparison and Control groups showed small increases while the PM group showed a small decrease in BSS Mathematics Self-Concept scores over time. This is evident after controlling for pre-test differences. The difference between groups was, however, not found to be statistically significant.

4.3.7 Summary: Research Question 2

Descriptive statistics indicated some small differences between groups:

- For the BIOS: while the comparison and control groups saw a slight increase in scores, the PM group data showed a slight decrease over time.
- For the BSS:
 - Global Self-Esteem scale: All groups showed a slight decrease in scores, where the PM group decreased slightly more than the control and comparison groups over time.
 - Learning Self-Concept scale: the PM and comparison groups showed a small decreases in scores while the control group showed a small increase over time.
 - While the comparison and control groups saw a small increase in scores, the PM group saw a slight decrease over time.

However, these differences were found not to be significant using inferential statistics.

The findings therefore indicate that one cannot reject the null hypothesis which states:
After completion of a PM intervention, there will be no statistically significant differences in self-esteem between the PM group, comparison group and/or control group for Year 8 children in the role of the Tutee.

4.4 Research Question 3 In a UK secondary school, does a 7 week, cross-age, structured peer-tutor programme around maths (Paired Maths) enhance self-esteem in Year 10 students in the role of Tutors?

The DVs in relation to *Research Question 3* are the pupils self-report scores on the Burnett Self Scale (BSS) and scores from the Behavioural Indicators of Self Esteem (BIOS) completed by their Maths Teachers. The IV is time (T1 and T2).

The hypotheses associated with this research question are outlined below:

Experimental Hypothesis: *Year 10 participants in the role of the tutor in the PM intervention group will have significantly higher gains in self-esteem than those in the control group.*

Null hypothesis 1: *After completion of a PM intervention, there will be no statistically significant differences in self-esteem between the PM group and control group for Year 10 participants.*

4.4.1 Behavioural Indicator of Self Esteem

4.4.1.1 Descriptive statistics

Table 4.25 displays the means and standard deviations (SD) for BIOS scores by group at pre (T1) and post (T2) testing. The data show that the Experimental group (PM) showed an increase on BIOS scores over time ($=0.067$), while the Control group remained the same.

Figure 4.17 presents this data using a bar chart.

Time	Group	Mean	Standard Deviation
Pre Test (T1)	Experimental (Paired Maths) (n=16)	3.967	0.876
	Control (n=15)	3.677	0.822
	Total (n=31)	3.826	0.840
Post Test (T2)	Experimental (Paired Maths) (n=16)	4.034	1.154
	Control (n=15)	3.677	1.201
	Total (n=31)	3.861	1.171

Table 4.25: A table displaying the means and standard deviations, by group, at pre (T1) and post (T2) testing for BIOS scores for Year 10 participants

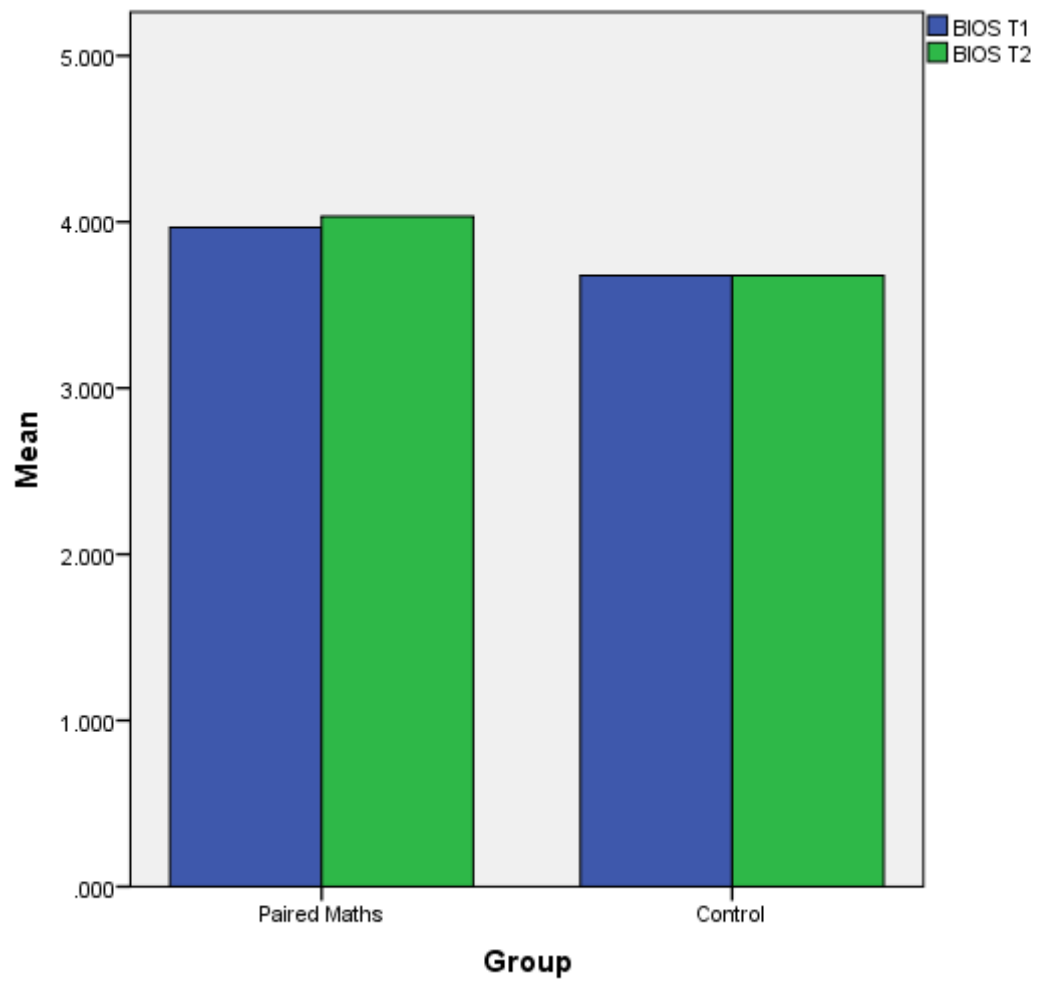


Figure 4.17: A bar chart displaying the mean, by group at pre (T1) and post (T2) testing for BIOS Scores for Year 10 Participants

4.4.1.2 Assumption Testing

4.4.1.2.1 Normality

Figure 4.18 illustrates the distribution of BIOS score at T1 for Year 10 participants.

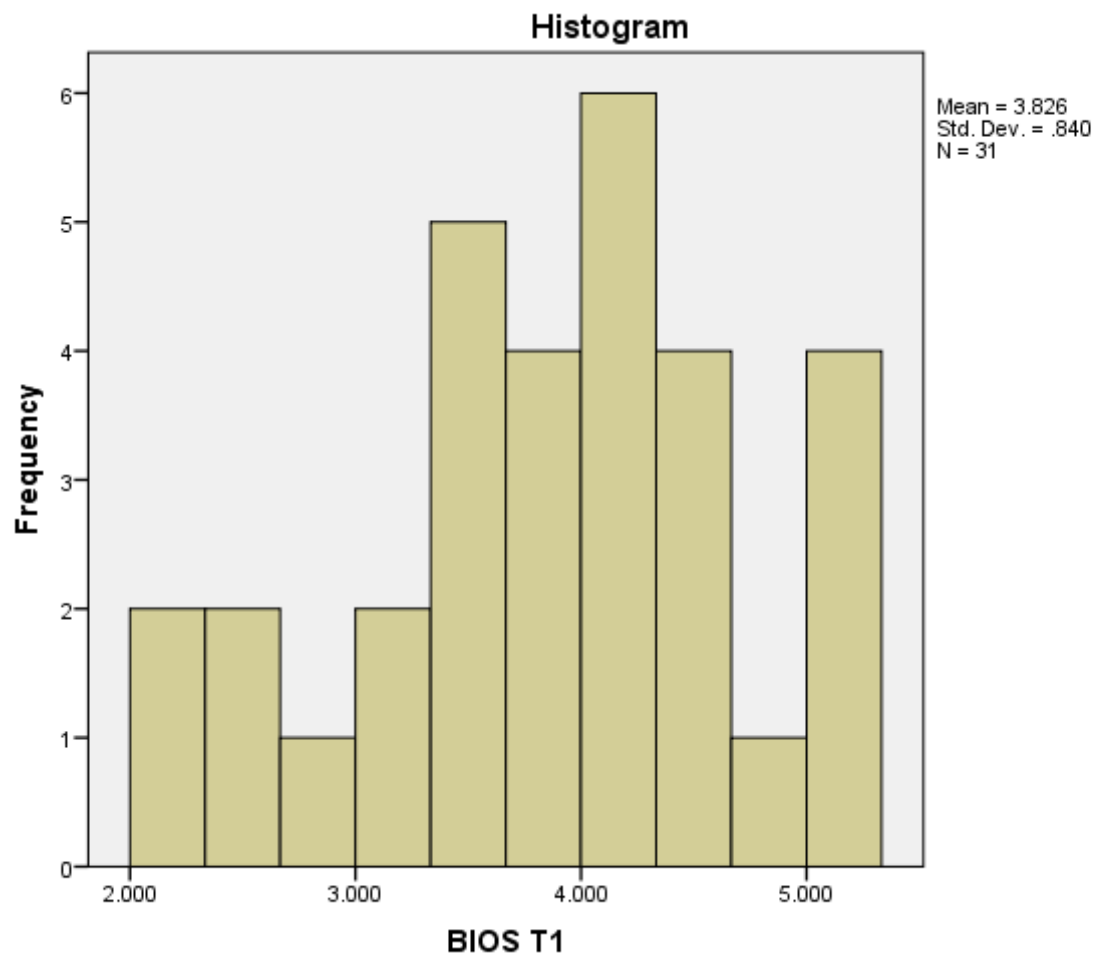


Figure 4.18: Figure showing the distribution of BIOS raw scores at T1 for Year 10 participants

The Shapiro Wilk statistic ($p > 0.05$) is reported in table 4.26. This was not statistically significant. Data from the BIOS for Year 10 participants at T1 was therefore assumed to be normally distributed.

Shapiro-Wilk		
Statistic	Degree of Freedom	Significant
0.945	31	0.116

Table 4.26: A table displaying the results of the Shapiro-Wilk test for average BIOS scores at T1 for Year 10 participants.

4.4.1.2.2 Homogeneity of variance

Levene's Test of Equality Variances (table 4.27) was not statistically significant ($p > 0.05$) when conducted on the DV: scores from BIOS at T2 for Year 10 participants. The groups were therefore considered to have homogeneity of variance.

Levene's Test of Homogeneity of Variance		
Statistic	Degree of Freedom	Significant
.350	29	.559

Table 4.27: A table displaying the result of Levene's test of equality variances for BIOS scores at T2 for Year 10 participants

4.4.1.2.3 Linearity

The relationship between the DV and CV, by group, is displayed in figure 4.19 . As the relationship at each level is linear (and not curvilinear), (Pallant, 2013), the data was considered applicable to ANCOVA.

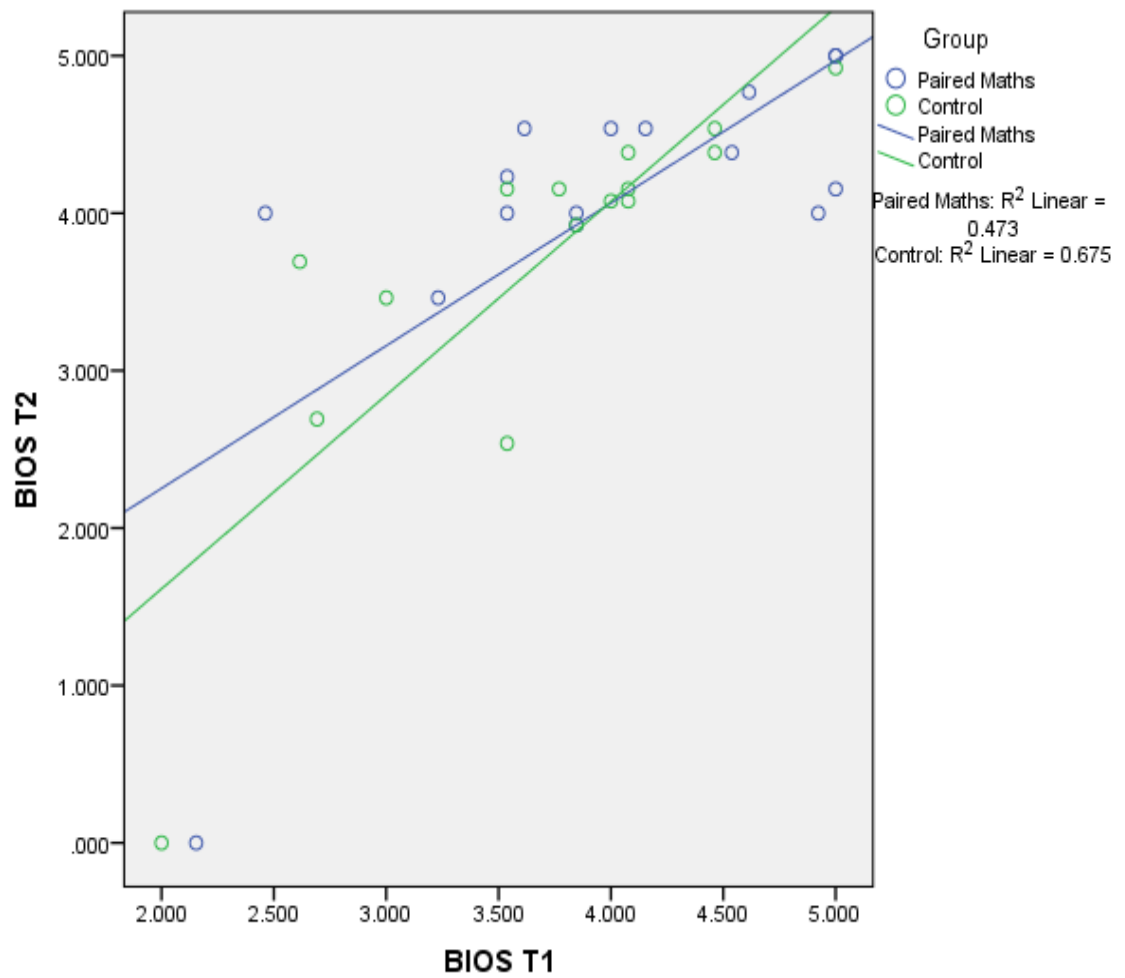


Figure 4.19: A scatter graph displaying the relationship between the DV (T2 BIOS scores) and CV (T1 BIOS scores), by group.

4.4.1.3 Homogeneity of regression slopes

Statistical analysis indicated that the test for homogeneity of regression was not statistically significant ($F(1, 31) = .841, p = .367$). This satisfies the requirement to perform an ANCOVA.

4.4.1.4 Inferential statistics

An ANCOVA was conducted to assess the differences in BIOS scores at T2 (DV) between groups, after controlling for pre-test differences (CV). The results showed that there was a statistically non-significant effect between groups, $F(1, 31) = .034, P = .855$. Output tables and adjusted mean scores when considered against the covariate are displayed in tables 4.28 and 4.29 respectively.

Tests of Between-Subjects Effects

Dependent Variable: BIOS T2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	23.545 ^a	2	11.773	18.733	.000	.572
Intercept	.031	1	.031	.049	.826	.002
BIOS T1	22.560	1	22.560	35.899	.000	.562
Group	.021	1	.021	.034	.855	.001
Error	17.596	28	.628			
Total	503.291	31				
Corrected Total	41.141	30				

a. R Squared = .572 (Adjusted R Squared = .542)

Table 4.28: A table displaying the results of the ANCOVA assessing the impact of Paired Maths on BIOS scores for Year 10 participants in the role of the tutor

Group

Dependent Variable: BIOS T2

Group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Paired Maths	3.887 ^a	.200	3.478	4.296
Control	3.834 ^a	.206	3.411	4.256

a. Covariates appearing in the model are evaluated at the following values:
BIOS T1 = 3.82623.

Table 4.29: A table displaying the adjusted means of BIOS Scores in relation to the CV

4.4.1.5 Summary BIOS

The descriptive statistics presented showed a small difference between groups, favouring PM in BIOS scores at T2. This is evident after controlling for pre-test differences. The difference between groups, however, was not found to be statistically significant.

4.4.2 Burnett Self Scale

Again, as The Burnett Self Scale(BSS) involves three subscales: Global Self-Esteem, Learning Self-Concept and Maths Self-Concept, these will be considered separately in turn.

4.4.3 Global Self-Esteem

4.4.3.1 Descriptive statistics

Table 4.30 displays the means and SDs of the Global Self-Esteem Scores by group, at pre (T1) and post (T2) testing for Year 10 Participants. The data show a very slight increase for the Experimental (PM) group (+0.016) and very slight decreases in the Control group (-0.011) over time.

Figure 4.20 presents this data using a bar chart.

Time	Group	Mean	Standard Deviation
Pre Test (T1)	Experimental (Paired Maths) (n=16)	3.789	.655
	Control (n=15)	3.403	.841
	Total (n=31)	3.602	.763
Post Test (T2)	Experimental (Paired Maths) (n=16)	3.805	.626
	Control (n=15)	3.392	.727
	Total (n=31)	3.605	.698

Table 4.30: A table displaying the means and standard deviations, by group, at pre (T1) and post (T2) testing for Global Self Esteem Scores for Year 10 participants

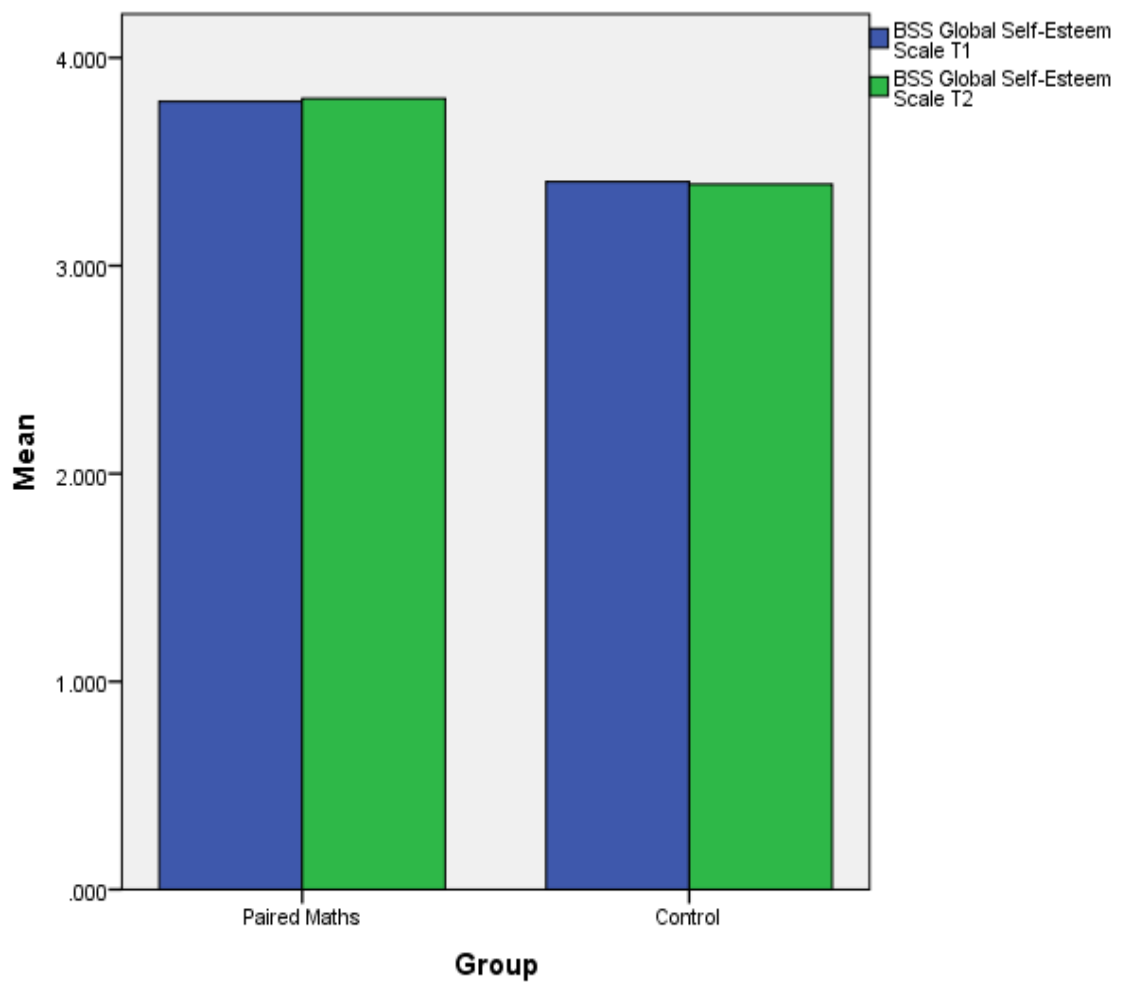


Figure 4.20: A bar chart displaying the mean, by group at pre (T1) and post (T2) testing for BSS Global Self-Esteem Scores for Year 10 Participants

4.4.3.2 Assumption Testing

4.4.3.2.1 Normality

Figure 21 illustrates the distribution of BSS Global Self-Esteem scores at T1 for Year 10 Participants.

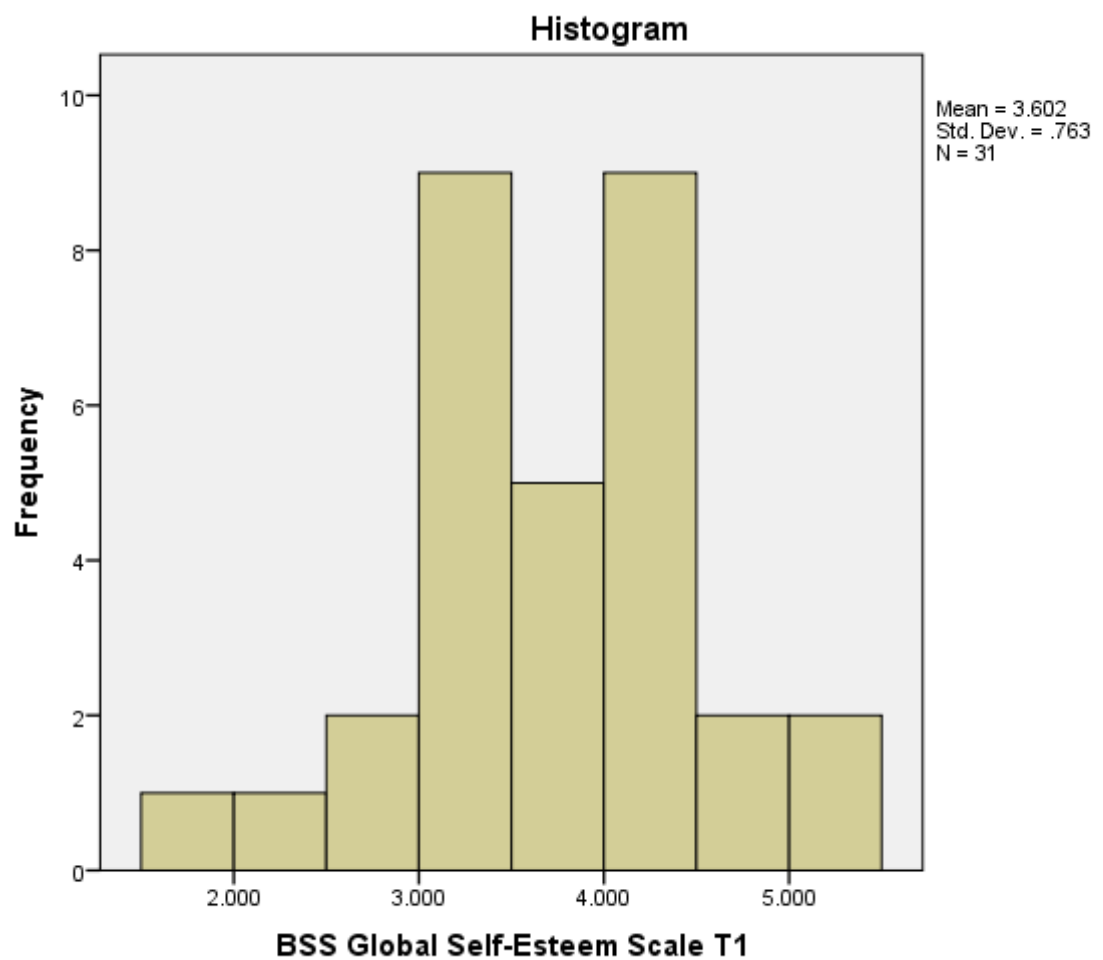


Figure 4.21: Figure showing the distribution of BSS Global Self-Esteem scores at T1 for Year 10 participants

The Shapiro Wilk statistic ($p > 0.05$) is reported in table 4.31. This was not statistically significant. BSS Global Self-Esteem scores at T1 were therefore assumed to be normally distributed.

Shapiro-Wilk		
Statistic	Degree of Freedom	Significant
0.973	31	.617

Table 4.31: A table displaying the results of the Shapiro-Wilk test for BSS Global Self Esteem scores at T1 for Year 10 Participants

4.4.3.2.2 Homogeneity of variance

Levene's Test of Equality Variances (table 4.32) was not statistically significant ($p > 0.05$) when conducted on the DV: BIOS scores at T2 for Year 10 Participants. The groups were therefore considered to have homogeneity of variance.

Levene's Test of Homogeneity of Variance		
Statistic	Degree of Freedom	Significant
1.110	29	.301

Table 4.32: A table displaying the results of Levene's test of equality variances for BSS Global Self-Esteem scores at T2 for Year 10 participants.

4.4.3.2.3 Linearity

The relationship between the DV and CV, by group, is displayed in figure 4.22. As the relationship at each level is linear, the data was considered appropriate for ANCOVA.

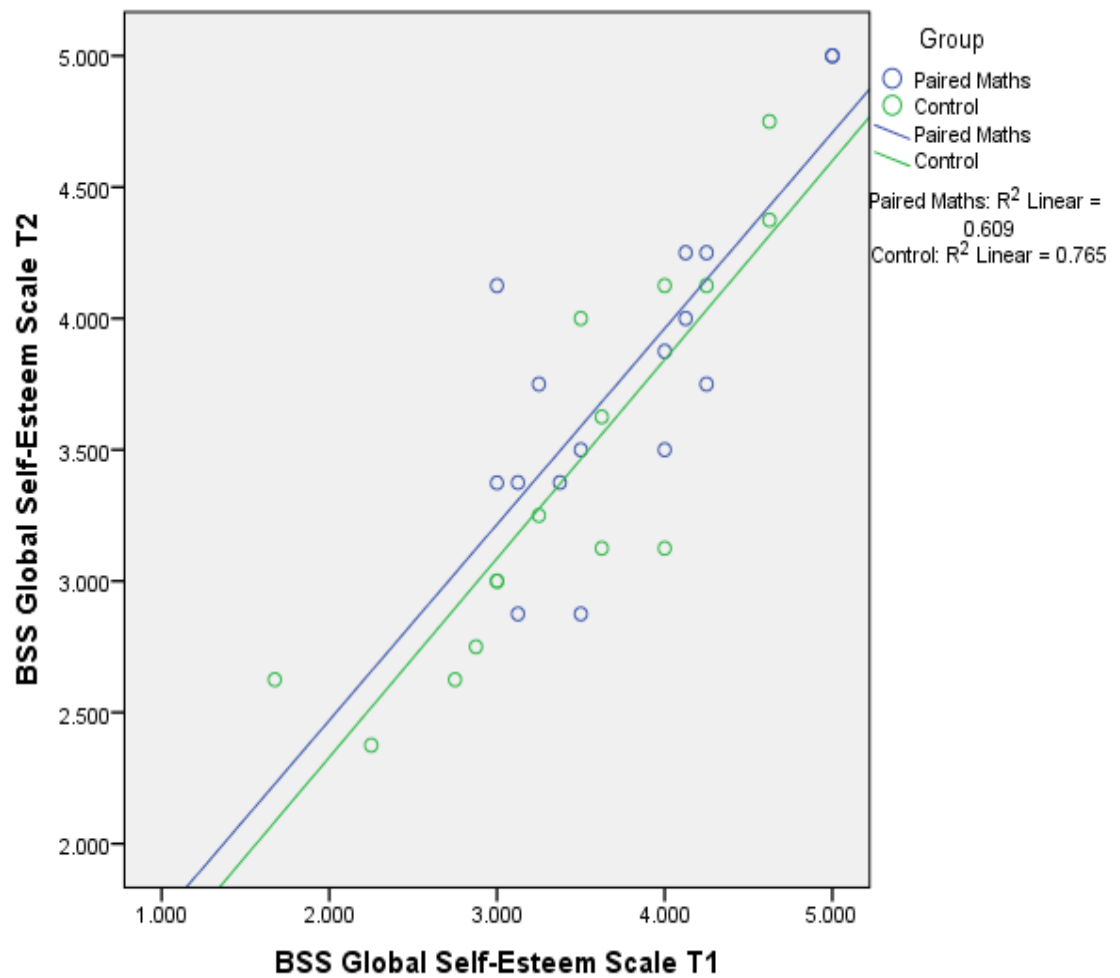


Figure 4.22: A scatter graph displaying the relationship between the DV (T2 BSS Global Self Esteem Scores) and CV (T1 BSS Global Self Esteem Scores) by group for Year 10 Participants.

4.4.3.2.4 Homogeneity of regression slopes

Statistical analysis indicated that the test for homogeneity of regression was not statistically significant ($F(1, 31) = .003, p = .956$). This satisfies the requirement to perform an ANCOVA.

4.4.3.3 Inferential statistics

An ANCOVA was conducted to assess the differences in BSS Global Self-Esteem scores at T2 (DV) between groups, after controlling for pre-test differences (CV). The results showed that there was a statistically non-significant effect between groups, $F(1, 31) = .756, P = .392$. Output tables and adjusted mean scores when considered against the covariate are displayed in tables 4.33 and 4.34 respectively.

Tests of Between-Subjects Effects

Dependent Variable: BSS Global Self-Esteem Scale T2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	10.561 ^a	2	5.281	36.634	.000	.724
Intercept	.965	1	.965	6.698	.015	.193
BSSGlobalFT1	9.241	1	9.241	64.108	.000	.696
Group	.109	1	.109	.756	.392	.026
Error	4.036	28	.144			
Total	417.431	31				
Corrected Total	14.597	30				

a. R Squared = .724 (Adjusted R Squared = .704)

Table 4.33: A table displaying the results of the ANCOVA assessing the impact of Paired Maths on BSS Global Self-Esteem Scores for Year 10 students in the role of Tutor.

Group

Dependent Variable: BSS Global Self Esteem Scale T2

Group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Paired Maths	3.738 ^a	.133	3.470	4.007
TA Comparison Group	3.753 ^a	.137	3.477	4.029
Control	3.759 ^a	.133	3.492	4.027

a. Covariates appearing in the model are evaluated at the following values: BSS Global Self Esteem Scale T1 = 3.78191.

Table 4.34: A table displaying the adjusted means of BSS Global Self-Esteem Scores for Year 10 Participants.

4.4.3.4 Summary: BSS Global Self-Esteem Scale

The descriptive statistics presented highlight a small difference over time between groups, favouring year 10s in the PM condition for Global Self-Esteem scores. This was also evident after controlling for pre-test differences. Differences between groups were not found to be statistically significant however.

4.4.4 BSS Learning Self-Concept Scale

Table 4.35 displays the means and SDs of the Learning Self- Concept Scores by group, at pre (T1) and post (T2) testing for Year 10 participants. The data show a small difference between groups over time where while the control group showed a very slight increase (+0.082) the PM group showed a small decrease (-0.109) in BSS learning Self-Concept scores.

Figure 4.23 presents this data using a bar chart.

Time	Group	Mean	Standard Deviation
Pre Test (T1)	Experimental (Paired Maths) (n=16)	4.000	.508
	Control (n=15)	3.733	.608
	Total (n=31)	3.871	.566
Post Test (T2)	Experimental (Paired Maths) (n=16)	3.891	.593
	Control (n=15)	3.815	.612
	Total (n=31)	3.733	.636

Table 4.35: A table displaying the means and standard deviations, by group, at pre (T1) and post (T2) testing for Learning Self-Concept Scores for Year 10 participants

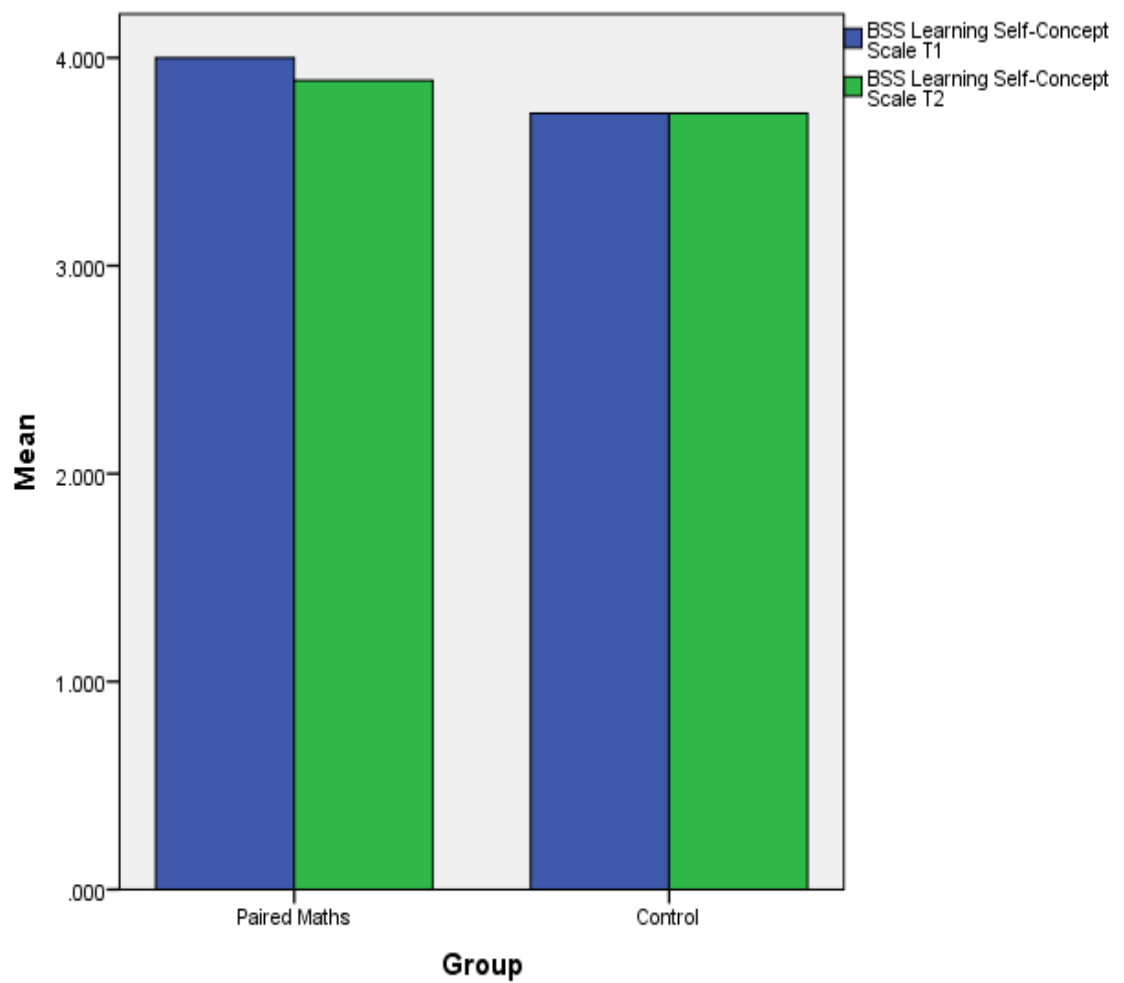


Figure 4.23: A bar chart displaying the mean, by group at pre (T1) and post (T2) testing for BSS Learning Self-Concept Scores for Year 10 Participants

4.4.4.1 Assumption Testing

4.4.4.1.1 Normality

Figure 4.24 illustrates the distribution of BSS Learning Self-Concept Scores at T1.

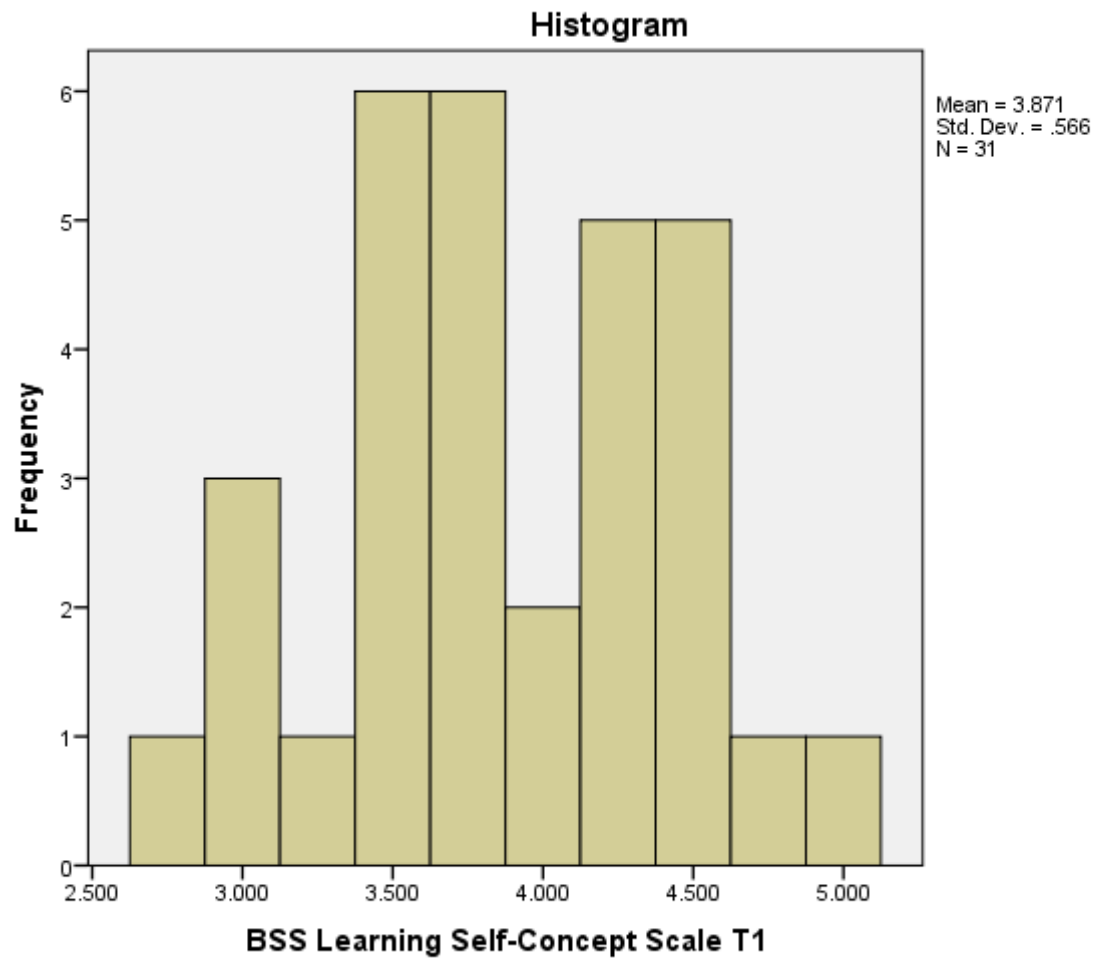


Figure 4.24: Figure showing the distribution of BSS Learning Self-Concept scores at T1 for year 10 participants

The Shapiro Wilk statistic ($p > 0.05$) is reported in table 4.36. This was not statistically significant. BSS Learning Self-Concept Scores at T1 for Year 10 participants were therefore assumed to be normally distributed.

Shapiro-Wilk		
Statistic	Degree of Freedom	Significant
.963	31	0.359

Table 4.36: A table displaying the results of the Shapiro-Wilk test for BSS Learning Self-Concept scores at T1 For Year 10.

4.4.4.1.2 Homogeneity of variance

Levene's Test of Equality Variances (table 4.37) was not statistically significant ($p > 0.05$) when conducted on the DV: BSS Learning Self-Concept Scores at T2 for Year 10 Participants. The groups were therefore considered to have homogeneity of variance.

Levene's Test of Homogeneity of Variance		
Statistic	Degree of Freedom	Significant
.491	29	0.489

Table 4.37: A table displaying the results of Levene's test of equality variances for BSS Learning Self-Concept Scores at T2 for Year 10s.

4.4.4.1.3 Linearity

The relationship between the DV and CV, by group, is displayed in figure 4.25. As the relationship at each level is linear, the data was considered appropriate for ANCOVA.

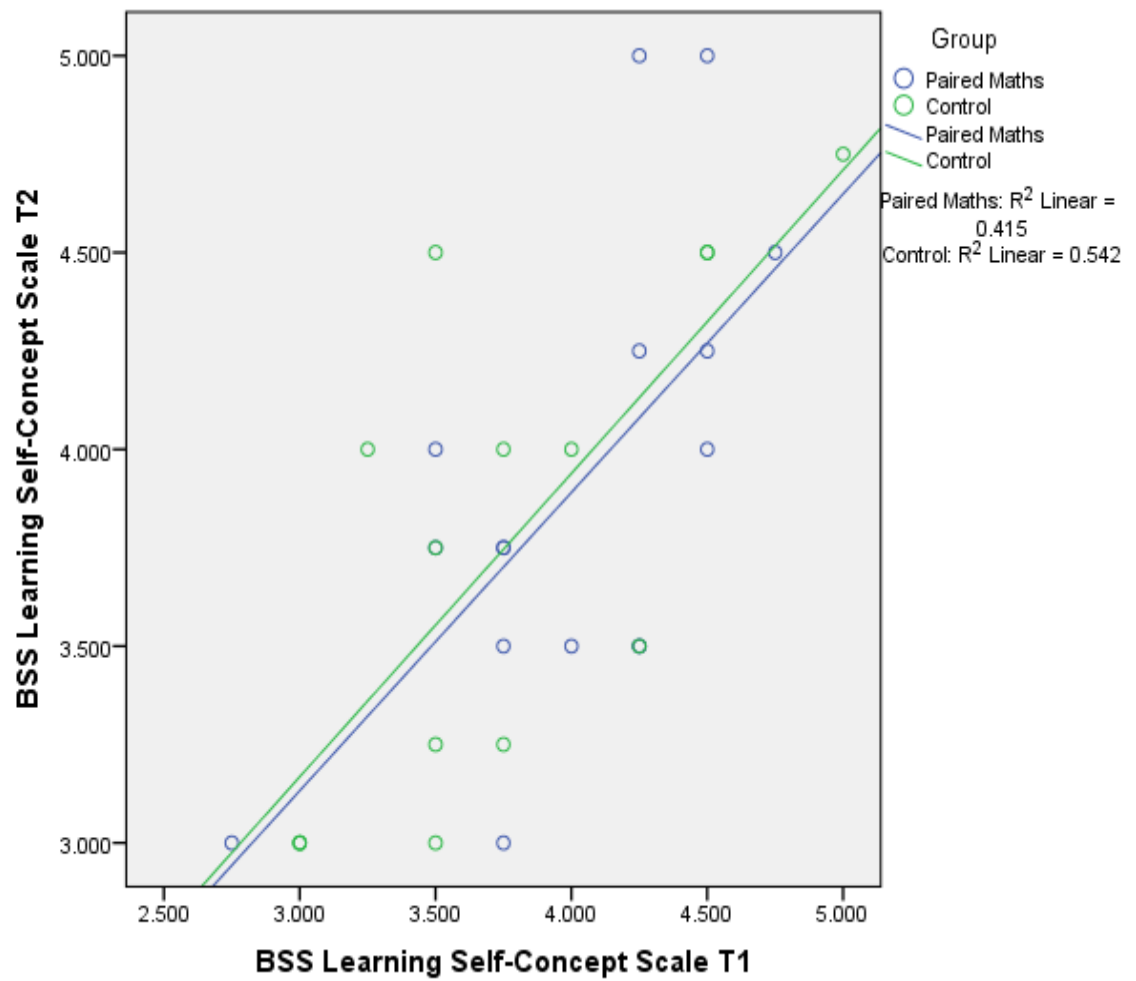


Figure 4.25: A scatter graph displaying the relationship between the DV (T2 BSS Learning Self-Concept Scores) and CV (T1 BSS Learning Self-Concept Scores) by group for Year 10 Participants

4.4.4.1.4 Homogeneity of regression slopes

Statistical analysis indicated that the test for homogeneity of regression was not statistically significant ($F(1, 31) = .002, p = 0.967$). This satisfies the requirement to perform an ANCOVA.

4.4.4.2 Inferential statistics

An ANCOVA was conducted to assess the differences in BSS Learning Self-Concept scores at T2 (DV) between groups, after controlling for pre-test differences (CV). The results showed that there was a statistically non-significant effect between groups, $F(1, 31) = .078, P = .782$. Output tables and adjusted mean scores when considered against the covariate are displayed in tables 4.38 and 4.39 respectively.

Tests of Between-Subjects Effects

Dependent Variable: BSS Learning Self-Concept Scale T2

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	5.498 ^a	2	2.749	13.393	.000	.489
Intercept	.432	1	.432	2.103	.158	.070
BSSlearningT1	5.307	1	5.307	25.853	.000	.480
Group	.016	1	.016	.078	.782	.003
Error	5.748	28	.205			
Total	462.313	31				
Corrected Total	11.246	30				

a. R Squared = .489 (Adjusted R Squared = .452)

Table 4.38: A table displaying the results of the ANCOVA assessing the impact of Paired Maths on BSS Learning Self-Concept scores for Year 10 students in the role of Tutor.

Group

Dependent Variable: BSS Learning Self-Concept Scale T2

Group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Paired Maths	3.792 ^a	.115	3.556	4.027
Control	3.839 ^a	.119	3.595	4.082

a. Covariates appearing in the model are evaluated at the following values:
BSS Learning Self-Concept Scale T1 = 3.87097.

Table 4.39: A table displaying the adjusted means of BSS Learning Self-Concept Scores for Year 10 Participants.

4.4.4.3 Summary: BSS Learning Self-Concept Scale

The descriptive statistics presented showed a small difference between groups, where it appeared the PM group showed a small decrease while the control group showed a very slight increase in BSS Learning Self-Concept scores over time. This is evident after controlling for pre-test differences. The difference between groups, however, was not found to be statistically significant.

4.4.5 BSS Mathematics Self-Concept Scale

4.4.5.1 Descriptives

Table 4.40 displays the means and SDs of the BSS Mathematics Self-Concept scores by group, at pre (T1) and post (T2) testing for Year 10 participants. The data shows small differences between groups in favour of the PM condition, where while the PM group appears to have shown a slight increase (+0.125) , the Control group has shown a decrease (-0.331) in BSS Mathematics Self-Concept scores over time.

Figure 4.26 presents this data using a bar chart.

Time	Group	Mean	Standard Deviation
Pre Test (T1)	Experimental (Paired Maths) (n=16)	3.859	.724
	Control (n=15)	3.617	.687
	Total (n=31)	3.742	.706
Post Test (T2)	Experimental (Paired Maths) (n=16)	3.984	.536
	Control (n=15)	3.286	1.006
	Total (n=31)	3.646	.861

Table 4.40: A table displaying the means and standard deviations, by group, at pre (T1) and post (T2) testing for Mathematics Self-Concept Scores for Year 10 participants

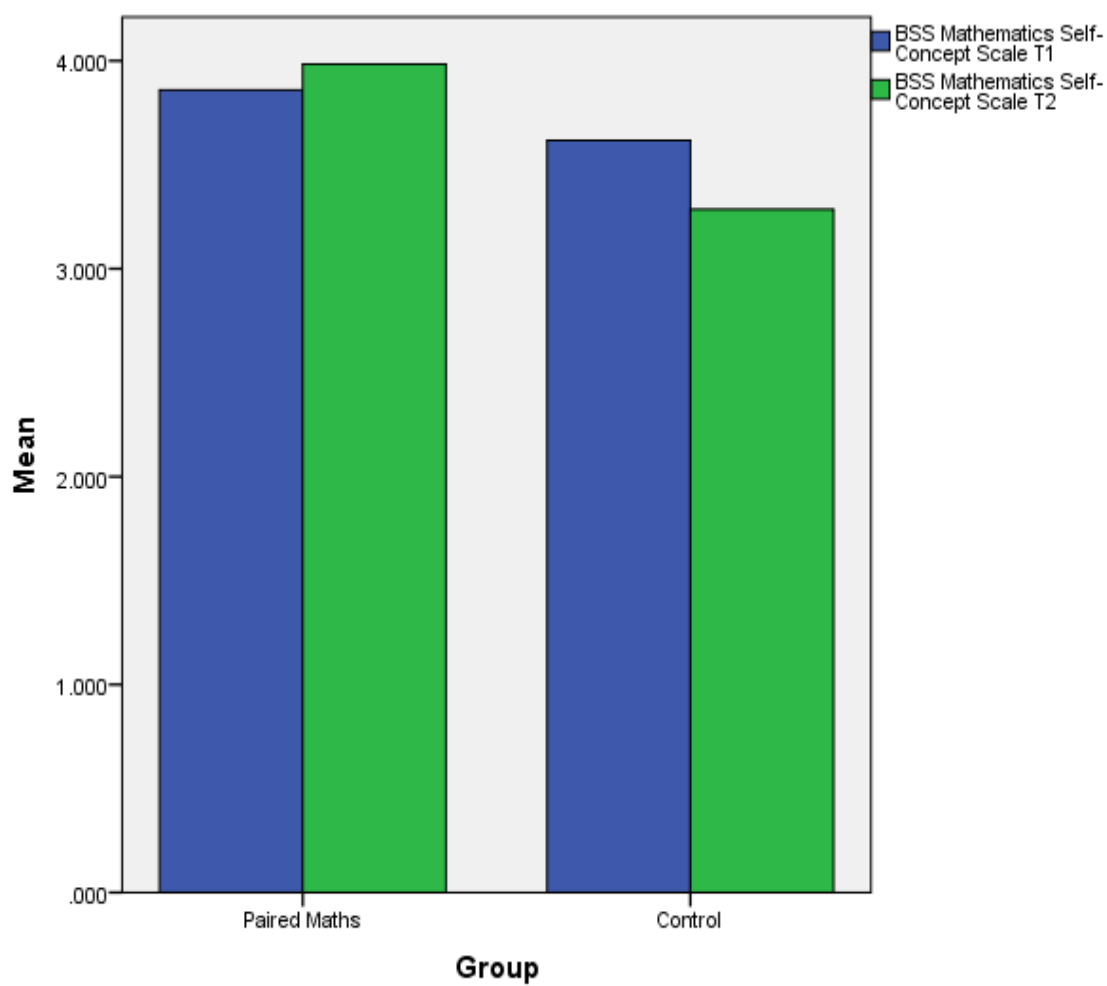


Figure 4.26: A bar chart displaying the mean, by group at pre (T1) and post (T2) testing for BSS Mathematics Self Concept scores for Year 10 Participants

4.4.5.2 Assumption Testing

4.4.5.2.1 Normality

Figure 4.27 illustrates the distribution of BSS Mathematics Self-Concept Scores at T1 for Year 10 participants.

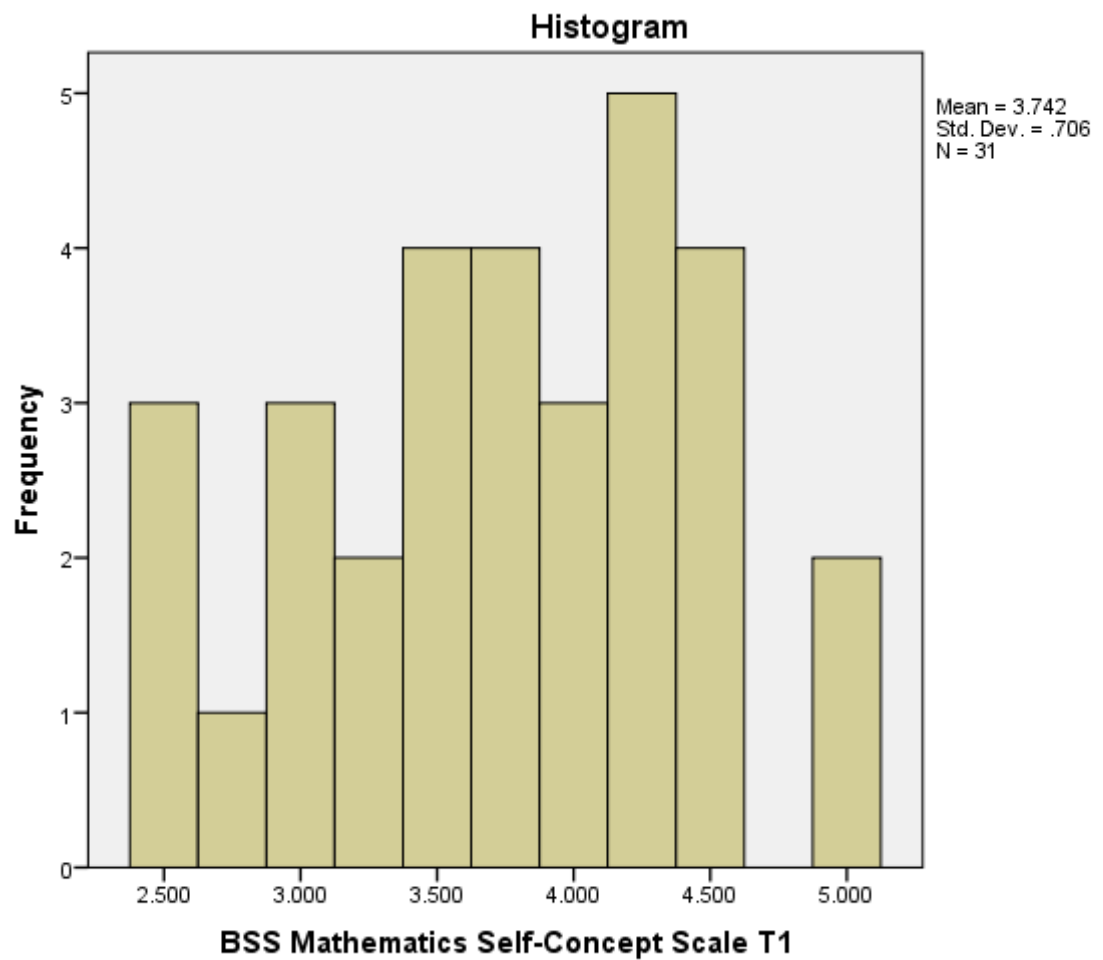


Figure 4.27: Figure showing the distribution of BSS Maths Self-Concept scores at T1 for year 10 participants

The Shapiro Wilk statistic ($p > 0.05$) is reported in table 4.41. This was not statistically significant. BSS Mathematics Self-Concept scores at T1 for Year 10s were therefore assumed to be normally distributed.

Shapiro-Wilk		
Statistic	Degree of Freedom	Significant
.959	31	.269

Table 4.41: A table displaying the results of the Shapiro-Wilk test for BSS Mathematics Self-Concept scores at T1 for Year 10.

4.4.5.2.2 Homogeneity of variance

Levene's Test of Equality Variances (table 4.42) was statistically significant ($p > 0.05$) when conducted on the DV: BSS Mathematics Self-Concept scores at T2. The groups were therefore considered to *not* have homogeneity of variance. The decision was consequently taken to use alternative non-parametric tests.

Levene's Test of Homogeneity of Variance		
Statistic	Degree of Freedom	Significant
7.742	29	0.009

Table 4.42: A table displaying the results of Levene's test of equality variances for BSS Maths Self-Concept scores for Year 10 at T2.

4.4.5.3 Inferential statistics

4.4.5.3.1 Mann Whitney U Test

The results of the Mann Whitney U Tests showed that there was no significant difference (where $p < 0.05$) between groups in the scores at T1 ($U=91.5$, $z=-.1.134$, $p=0.264$., two-tailed) . However, there was significant difference in scores at T2 between the Paired Maths and Control group ($U=69.5$, $z= -2.35$, $p= 0.045$, two-tailed).

4.4.5.3.2 Wilcoxon Signed Ranks Test

The results of the Wilcoxon Signed Ranks Tests showed that there were no significant differences between T1 and T2 for the Paired Maths group ($z = .917$, $p = .359$, two-tailed), however, for the Control group, T2 scores were found to be significantly lower than at T1 ($z = -2.422$, $p = .015$, two-tailed).

4.4.5.4 Summary: BSS Mathematics Self-Concept Scale.

Descriptive statistics showed that while the PM group showed a very small increase in scores, scores for the control group decreased. Non-parametric testing did show a significant difference between the PM group and Control groups at T2. However, further analysis showed that this difference was only due to a significant decrease in scores *within* the control group, and no significant differences were found between T1 and T2 for the PM group.

4.4.6 Summary: Research Question 3

Analysis of the descriptive statistics showed the following:

- For the BIOS Teacher-report measure: While the PM group showed a small increase in scores, scores for the control group remained the same.
- For the BSS Self-report measure:
 - Global Self-Esteem Scale: While the PM group showed a very small increase in scores, scores for the control group showed a very small decrease
 - Learning Self-Concept Scale: While the PM showed a small decrease in scores, the control group's scores remained the same
 - BSS Math Self-concept: While the PM group showed a very small increase in scores, scores for the control group decreased

Although inferential statistics showed that there were no significant differences between groups on the BIOS or the Global Self-Esteem or Learning-Self subscales of the BSS measures of self-esteem, non-parametric testing did show a significant difference between the PM group and Control groups at T2. However, further analysis showed that this difference was only due to a significant decrease in scores within the

control group, and no significant differences were found between T1 and T2 for the PM group.

Therefore, overall, results suggest one cannot reject the null hypothesis which states:
After completion of a Paired Maths intervention, there will be no statistically significant differences in self-esteem between the Paired Maths group, a comparison intervention group and/or control group for Year 10 tutors.

4.5 Summary of Results Section

Table 4.43 presents a summary of the results for each research question. Chapter five reviews these findings in reference to the literature and methodology reviewed in chapters two and three.

Question	Measure	Construct	Null hypothesis rejected?
1. In a UK secondary school, does a 7 week, cross-age, structured peer-tutor programme (Paired Maths) enhance maths attainment in Year 8 students in the role of Tutees?	Basic Number Screening Test	Mathematics Attainment	No
2. In a UK secondary school, does a 7 week, cross-age, structured peer-tutor programme around maths (Paired Maths) enhance self-esteem in Year 8 students in the role of Tutees?	Behavioural Indicator of Self -Esteem	Teacher- observations	No
	Burnett Self Scale	Global-Self Esteem	No
		Learning Self-Concept	No
		Maths Self-Concept	No
3. In a UK secondary school, does a 7 week, cross-age, structured peer-tutor programme around maths (Paired Maths) enhance self-esteem in Year 10 students in the role of Tutees?	Behavioural Indicator of Self -Esteem	Teacher- observations	No
	Burnett Self Scale	Global-Self Esteem	No
		Learning Self-Concept	No
		Maths Self-Concept	Only <i>Control</i> participants significantly decreased over time

Table 4.43: A table summarising the results by question, by measure, against the null hypotheses.

5 Discussion

5.1 Introduction

The current research employed two RCTs to assess the impact of Paired Maths (PM) - a cross-age structured peer-tutoring intervention focused on mathematics- with secondary aged students. The first RCT looked at the impact of the intervention on mathematics attainment and self-esteem for year 8 children in the role of the *tutee*, whereas the second RCT looked at the impact of the same intervention on self-esteem for year 10 children in the role of the *tutor*. This chapter begins by giving an overview of findings for each research question and considers how findings relate to the existing literature. Possible limitations of the methodology are then considered, before possible explanations for findings are explored. Future research directions and implications for Educational Psychology practice are also discussed. Overall conclusions, including unique contributions, bring the chapter to a close.

5.2 Research Question 1: Impact on maths attainment for Year 8 tutees

Research Question one aimed to investigate the efficacy of a cross-age peer-tutoring intervention (PM) in increasing mathematics attainment in year 8 children in the role of the tutee, in comparison to a group supported by a TA and a further control group of children who did not receive any additional mathematics intervention at all.

The stated hypothesis was that there would be a statistically significant difference between groups, over time, favouring the children in the PM Group. Analysis of the descriptive statistics, however, showed that mathematics attainment increased slightly across all groups, with perhaps the control participants making marginally greater improvements compared to the intervention groups. Furthermore, inferential statistics showed that there were no statistically significant differences between groups. That is, the null hypothesis – that after completion of the PM intervention there would be no statistically significant differences -could not be rejected. Findings do not indicate that PM is effective in increasing mathematics attainment for secondary school students in the role of the tutee.

Although not statistically significant, and whilst one cannot say that it is not due to chance, it is interesting that the control group showed a greater increase than the intervention groups. The author considers, perhaps, that this could have been because the control group had an overall lower starting point, so perhaps they had more room for improvement. This is explored further in section 5.6.

5.2.1 Comparison of current findings to existing literature

When considering the four studies included in the systematic literature review (Johnson and Bailey, 1974; Sharpley, Irvine and Sharpley, 1983; Beirne-Smith, 1991; Tymms et al. 2011) all researchers reported that students in the role of the tutee involved in a cross-age peer-tutoring intervention targeting mathematics experienced significant improvements in mathematic achievement when compared to control participants. The findings from this study research, therefore, do not support the existing experimental research.

Of course, none of the above cited studies involved secondary aged participants. However, when reviewing experimental research around general peer-tutoring interventions, Leung (2015) found that studies which included students from secondary schools actually displayed the largest effect sizes compared to children in other age groups (university/ college level, primary/elementary school and kindergarten/nursery). Results from the present study are clearly not in line with Leung's conclusions.

Leung's (2015) meta-analysis did not solely include cross age peer-tutoring, and did not focus only with mathematics. Nevertheless, Gerston and colleagues(2009) meta-analysis *did* only consider peer-tutoring techniques involving mathematics, and they also inferred that cross-age tutoring appeared to work more effectively on increasing maths attainment in tutees than same-age peer-tutoring, and any other peer assisted learning approaches. Again, the findings from the present study are not in line with this conclusion. The review from Gerston et al (2009), however, only included children considered to have learning disabilities. In addition, it was highlighted by authors that there were relatively fewer studies that used *cross-age* peer-tutoring and argued that further research was needed to increase the reliability of findings (Gerston et al. 2009).

Burden (2015) highlights that research which fails to reject the null hypothesis is usually less likely to be published in peer-reviewed research journals. Therefore, the researcher considered it was important to also consider results (with caution) in light of any recent unpublished studies that seemed particularly pertinent to the current research.

The Durham Shared Maths Project, results published online by the Education Endowment fund in 2015, involved a large scale study in UK primary schools considering the effectiveness of structured cross-age peer-tutoring in mathematics. Contrary to other published research, no significant differences were found between the PM intervention and control groups for tutees on maths attainment measures, and it was suggested that cross-age peer-tutoring programmes, in this instance, had no additional impact on tutees above the maths instruction being received in the normal classroom. The current research supports these findings.

Another recent piece of unpublished research of particular significance is Posselthwaite, Semmes and Clarke (submitted for publication, 2015). When considering the same PM framework as employed in the current research -and in *the same local authority*, albeit with primary aged students and times table knowledge- it was found that both the tutee participants in the PM intervention group *and participants in a comparison group* showed significantly improved knowledge of the 3, 4 and 5 times tables, while the control group did not. Therefore, although in the current research the PM intervention did not, apparently, lead to a significant difference in maths attainment compared to the control condition, findings do support research from Posselthwaite, Semmens and Clarke (submitted for publication, 2015) in that there were no significant differences between the PM intervention group and the *comparison group*.

Overall, the researcher considers that the findings for the impact of maths attainment in tutees do not support the current published research around cross-age peer-tutoring interventions around mathematics, but this may not necessarily be the case for research reviewing cross-age peer-tutoring in mathematics overall. In addition, to the researcher's knowledge, this was the first truly *experimental* piece of research of a

cross-age peer-tutoring intervention around mathematics with *Secondary* aged pupils. Other possible explanations for these findings are explored in section 5.7.

5.3 Research Question 2: Impact on Self-esteem for Year 8 Tutees

Research Question Two aimed to investigate the efficacy of PM in increasing self-esteem in year 8 children in the role of the tutee, in comparison to a group supported by a TA and children who did not receive any additional maths intervention at all.

The hypothesis stated that there would be a statistically significant difference between groups, and time, favouring the children in the PM Group. Inferential statistics, however, showed that there were no *significant* differences between groups on either the BIOS or the BSS measures of self-esteem. That is, the null hypothesis - that after completion of the PM intervention there is no statistically significant differences in self-esteem between the PM group, a comparison group and/or control group- could not be rejected.

Findings do not indicate that PM is effective in increasing self-esteem for secondary school students in the role of the tutee. Further, although not statistically significant (so one cannot say that it is not due to chance alone) it is interesting that the descriptive statistics throughout the measures of self-esteem generally show a larger decrease in the PM group compared to both the control and comparison group. This *could possibly* indicate that PM could actually have a negative impact on self-esteem for secondary school students in the role of the tutee (though, again, this should be viewed with considerable caution). Possible explanations around these findings are explored further in section 5.7.

5.3.1 Comparison of current findings to existing literature

None of the experimental studies specifically investigating the literature around cross-age peer-tutoring around maths included in the systematic literature review (Johnson and Bailey, 1974; Sharpley, Irvine and Sharpley, 1983; Beirne-Smith, 1991; Tymms et al. 2011), investigated how the intervention impacted on the self-esteem of the tutee. The researcher will therefore consider how findings relate to research more generally around peer-tutoring interventions. In their meta-analysis, Ginsburg-Block, Rohrbeck

and Fantuzzo (2006) focussed particularly on social, self-concept (or self-esteem) and behavioural outcomes for peer-assisted learning interventions. They found that individuals who are involved in peer assisted learning interventions see significantly larger increases in self-esteem outcomes than children who are not, and that these are "moderate" positive effects. The current findings do not support these conclusions. However, this review did not only consider peer-tutoring, but also a number of different peer assisted learning techniques. Therefore, it could be possible that the *other* peer-learning approaches, such as co-operative learning, accounted for the positive findings around self-esteem, and that actually specifically cross-age peer-tutoring does not see these positive outcomes.

The current research findings do, however, support earlier reviews regarding the impact of *peer-tutoring* on self-esteem. For example, Cohen and colleagues in their review from 1982 found that studies showed only a very small effect on self-esteem in peer-tutoring interventions for tutees. In addition Rohrbeck and colleagues (2003), when considering very specifically *mathematics focused* peer-tutoring, found no significant effects of the intervention for tutees (e.g. Bar-Eli & Raviv, 1982; Sharpley, Irvine and Sharpley, 1983).

However, results do not support findings from the recent (yet to be published) study in the same local authority as the present research. Posslethwaite, Semmens and Clarke (submitted for publication, 2015) found that tutees in their PM intervention group did show an increase in their self-esteem. This finding was considered particularly pertinent by the authors, because the group who received TA support did not see increases in self-esteem, but they *did* also see significant increases in their level of maths knowledge. It was put forward this could mean that it was unique factors associated with the nature of the PM intervention itself that could have caused the increased self esteem-irrespective of academic achievement. The current findings, where the PM group neither increased their maths attainment *nor* their self-esteem is not in line with these claims. However, again this piece of research was quite small in scale and did not include secondary aged pupils, so it may not be a valid comparison.

Findings are quite mixed in the existing literature, and none of the experimental studies reviewed in the systematic review examined the impact on self-esteem for children in the role of the tutee. Therefore, overall, the researcher considers that the findings in the current research *could* support *some* of the existing literature. They *could* support the literature suggesting that a PM intervention *does not* have a significantly positive impact on self-esteem for children in the role of the tutee. However, again, it is difficult to make truly valid comparisons, as no other *experimental* research involving a cross-age peer-tutoring intervention focused on mathematics has considered the impact on self-esteem for the tutee, or for secondary aged participants. Possible explanations around why these findings may have occurred in the current research are explored later in section 5.7.

5.4 Research Question 3: Impacts on Self-Esteem for Year 10 Tutors

Research Question Three aimed to investigate the efficacy of PM in increasing self-esteem in year 10 children in the role of the tutor, in comparison to a control group of children who did not receive any additional maths intervention.

The hypothesis here was that there would be a statistically significant difference between groups, over time, favouring the children in the PM Group in comparison to a control group. Although inferential statistics showed that there were no *significant* differences between groups on the BIOS or the Global Self-Esteem or Learning-Self subscales of the BSS measures of self-esteem, non-parametric testing did show a significant difference between the PM group and Control groups for the self-report BSS measure of Maths Self-Concept. However, further analysis showed that this difference was only due to a significant decrease in scores within the control group, and no significant differences were found between the pre- and post- testing for the PM group.

Therefore, overall, findings *do not support* the conclusion that a cross-age peer-tutoring intervention around mathematics in secondary school students is effective in increasing self-esteem for children in the role of the tutor.

5.4.1 Comparison of current findings to existing literature

Only one of the experimental studies included in the systematic literature review considered the impact of cross-age mathematics peer-tutoring on the self-esteem of the tutor. Sharpley, Irvine and Sharpley, (1983) using the Coopersmith Self-Esteem Inventory (Coopersmith, 1967), found no significant gains in self-esteem for tutors when compared to the control group. The current findings are therefore in line with this research.

However, the current research does not support the findings from a more recent meta-analysis of studies from Ginsburg-Block, Rohrbeck and Fantuzzo (2006). They found that any individuals who are involved in peer assisted learning interventions (in any role) see significantly larger increases in self-esteem outcomes (though they term it *self-concept*) than children who are not, and that these are "moderate" positive effects. However, again it is important to note that this meta-analysis did consider a number of different peer assisted learning techniques, across a number of different subjects, so results may be confounded by other elements of these other interventions.

Finally, when considering the most recent research into PM interventions in primary schools within the same local authority as the current research, Posselthwaite, Semmens and Clarke (submitted for publication 2015) reported that tutors did show a significant increase across all of the same self-esteem measures as used in the current research. This finding was not replicated in the current research with children in secondary schools, although, as highlighted in section 2, the research from Posselthwaite, Semmens and Clarke did not include a control group for tutor participants.

5.5 Overall Summary of Findings

Howell (2013) argues that when data analysis does not lead the researcher to reject the null hypothesis, any attempts to make interpretations are challenging. The researcher also considers that, particularly since this research was exploratory in nature and there were number of implementation issues (considered below), further caution should be exercised when considering findings within this research.

Therefore it is not possible to conclude that a cross-age peer-tutoring intervention (PM) is either effective or ineffective in increasing maths attainment (here in children in the role of the tutee) or self-esteem (in children in the role of the tutee or tutor).

However, one may tentatively consider that findings *could* suggest that the guidance within the "Teaching and Learning Toolkit" from the Sutton Trust (Higgins et al, 2013) - that Peer-tutoring can be considered to be a high impact for low cost intervention- *may not* be completely accurate, at least for a *cross-age* peer-tutoring intervention around *mathematics* with *secondary* aged pupils.

5.6 Methodological Limitations

In an attempt to try add more clarity around the results, the researcher first critically considers the methodology of the current research, before further exploring any other possible explanations for the findings presented (Section 5.7).

5.6.1 Measures

Before considering the limitations of each measure individually, the researcher reflects on some limitations that may have been apparent across all measures.

The first possible limitation is around how the measures were administered as a group, rather than on an individual basis. Although students were encouraged to complete assessments on their own in silence, participants may still have influenced each other's performance in some way. At T2 in School 3 for example, one pupil presented as extremely distractible and made some noises throughout the numeracy test that appeared to distract some of the other pupils. This may well have influenced the individual participant's and others responses on the test.

The second possible limitation surrounding the year 8 data is that- due to the presentation of the assessments- there may have been some order effects. This is because for all year 8 participants, the mathematics attainment measure was given directly before the self-esteem self-report measure, on the same day. The researcher considers the experience of the mathematics test may have influenced responses to the second self-esteem.

Finally, in an attempt to increase the validity and reliability of the assessment procedures, the researcher administered all of the measures. Although standardised administration procedures were used in groups, and the researcher was blind to the allocation of groups at T1, the researcher was aware of some of the allocation of groups at T2 (due to the treatment fidelity checks carried out). Thus, there may have been a slight risk of (unconscious) researcher bias (Mertens, 2014). Though overall, this was not considered a great threat to the validity of the research.

In future research, using individual administration of measures, across time, possibly with an independent researcher may reduce these threats.

While it was a priority for the researcher to choose measures that would be both reliable and valid, potential limitations relating to instruments employed remain.

5.6.1.1 Measuring mathematics attainment:

Though the Basic Number Screening Test (BNST) was chosen as a standardised assessment of maths attainment with care, the researcher considers the following limitations to its use within this research:

Firstly, some participants initially scored quite highly at T1, with one participant actually achieving the maximum score. Although the researcher selected a measure which, according to the selection criteria given to the SENCOs in each school, *should* have provided enough challenge to allow any improvement to be captured, without too much experience of failure during the assessment process, it appears that there may have been a ceiling effect here. Ceiling effects increase the risk of a Type II error, where the null hypothesis is wrongly accepted (Mertens, 2014). It may have been that participants did experience more gains in maths attainment, but that these were not captured by the BNST. Indeed, as mentioned earlier, this could have been one possible explanation to why the control group showed a slightly higher increase in their BNST scores compared to the PM and comparison groups, as they scored lower on average at pre-test and therefore may have had more room for improvement.

Second, a limitation which the researcher considers could be quite significant within this research, is the possibility that the items and areas of mathematics included in the

BNST did not measure the specific mathematics skills that were improved during the PM intervention. Indeed, in their research using a cross-age peer-tutoring intervention, Johnson and Bailey (1974) highlighted that a sub-analysis of their results showed only significant improvements were made for the specific skills that had been taught within the tutoring. Further, in the Durham Shared Maths Project (Lloyd et al., 2015a) although the children did not make any significant gains on the standardised measures of Maths attainment used, qualitative information gained from Teacher views indicated the tutees may have consolidated existing skills and maths knowledge and improved their reasoning and problem solving abilities (Lloyd et al., 2015a).

Indeed, a related further limitation of the assessment of mathematics attainment is that only one measure was used. As done so in the Durham Shared Maths Project (Lloyd et al., 2015a), Bryman (2006) advocates the use of both quantitative and qualitative methods in order to try and triangulate results and increase the chances of fully capturing a construct. However, in a post-positivist epistemology (as adopted in this research), qualitative methodologies do not typically feature. This is considered further in section 5.8.

5.6.1.2 Measuring self-esteem

5.6.1.2.1 Burnett Self Scale

Many have advocated the use of self-report assessments with children in research, as they arguably allow access to a crucial and unique perspective (Sturges, Rodger and Ozanne, 2002). However, there are limitations to such assessments, which may have been evident in this research.

First, particularly for the lower ability year 8 students, there may have been some issues around the comprehension and understanding of statements – or of the response scale involved. Although the researcher and a member of staff in each school gave support with reading if needed, and the authors of the measure advocate that it is of a simplicity suitable for the participants included, this still may have occurred. In addition, some of the children may have had difficulty in actually conceptualising and reporting on the ‘self’ in general.

Second, the researcher considers that there could have been some level of participant bias. Although participants were reassured that there were no 'right or wrong answers', that their answers would remain anonymous and that they should answer as honestly as possible, pupils may still have tried to give a 'correct' answer, or the most socially desirable answer to suit the researcher rather than an answer that reflected their true thoughts.

However, the researcher considered these possible limitations during the design of the research, and for this reason included a further measure of self-esteem involving a different perspective, in order to provide triangulation and increase reliability.

5.6.1.2.2 Behavioural Indicators of Self-Esteem

Although the BIOS was included in the research in attempt to triangulate and therefore address some of the possible limitations of the BSS measure of self-esteem, this measure also has some potential limitations:

First, is the potential issue around how motivation may have affected the accuracy of responses on the questionnaires. Although every teacher who filled out the questionnaires was spoken to by the researcher (and they seemed to value the research) one could argue that they may not have been sufficiently motivated to fill out each questionnaire accurately- particularly as each teacher often had quite a large volume to complete.

Further, the researcher considers that there was some threat of bias from the teachers when completing the BIOS. Though blind to allocations of groups at T1, due to their daily interactions with the children, the teachers may well have been aware of allocations of groups by T2.

5.6.2 Internal validity

As reflected earlier in section 3, the use of the RCT methodology helps to guard against many threats to internal validity (Robson, 2011). However, there are still many threats that can remain, particularly when the RCT occurs within a 'real world' context. How these threats were considered is detailed in Chapter 3, however the key threats to

internal validity considered by the researcher to be present in this research and their implications are highlighted below:

Testing: There is little evidence to suggest that testing threatened the validity of the findings for the BNST. However, as the same forms were used for T1 and T2 for both the BIOS and BSS measures of self-esteem, there may have been some practice effects. However, a 12 week time period existed between T1 and T2 testing, and as there are no 'right or wrong' answers, this threat was considered to be low.

Instrumentation: As standardised procedures were carefully adhered to by the researcher for the administration of the BNST and the BIOS, the threat of instrumentation for these measures was considered low. However as discussed above, it is acknowledged that there was a higher level of threat for the BIOS measure, as although specific guidance and supervision was given, it was not the researcher that completed this measure. Implications for this on the research are unknown, and a recognised limitation.

Compensatory rivalry: Clear communication and scripting of assessment administration were used to try and limit this threat. At T2 however, in one school, at the beginning of the assessment procedures measures for the year 8 pupils the researcher heard a pupil from the control condition say: "I know what she's trying to do, I'm not bad at maths just because I wasn't involved in the stupid group." This may well have impacted upon the responses of pupils across all conditions. The direction and size of the possible impact is unknown.

Compensatory equalisation of treatments: The purposes and procedures of the research were clearly outlined to all stakeholders directly involved in the research. It was hoped that this limited stakeholder inclination to equalise treatments. The treatment fidelity checks within the comparison TA groups did not suggest that there was compensatory equalisation of treatments for year 8 participants, and all 3 TAs were considered to follow the guidelines for support. However, it is considered *possible* that within the children's usual maths lessons, as their maths teachers may have become aware of the allocation of groups, there could have been some level of bias in teaching/ approach.

5.6.3 Treatment integrity

Table 5.1 and 5.2 display the percentage of sessions observed that used each of the key elements for PM (both the more general characteristics of the sessions and the extent the specific structured PM approach was followed). In sum, the researcher considered that out of the 9 sets of tutor-tutee pairings observed, only 2 pairs appeared to be adequately following the PM process in their interactions with their tutee. The implications of this are further considered when exploring possible explanations for findings in section 5.7.

General Element of PM Intervention	Present
Facilitator available if needed?	100%
Intervention Lasted at least 15 minutes?	100%
Additional Maths apparatus available?	100%

Table 5.1: Table displaying the results for the Treatment Integrity observation, general features.

Element of PM Process	Present?
Did they read the question together?	77%
Did the tutor clarify the tutees understanding of the question/ explain key terms?	44%
Did the tutor encourage the use of apparatus?	22%
Did the tutor praise the tutee?	44%
Did the tutor encourage 'thinking out loud'?	22%
Did the tutor check the accuracy of the answer?	100%
Did the tutor model how to get to the correct answer if incorrect?	77%
Was the tutee asked to 'sum up' how they had arrived at their answer?	22%
Did the tutor relate the question to real life?	22%
Did the tutor relate the question to other maths problems?	22%
Did the tutor help the tutee choose the activity?	75%

Table 5.2: Table displaying the results for the Treatment Integrity observation, features of PM Process within pairs

Research which occurs within the post-positivist paradigm seeks to *specifically* define what occurred within different treatments (Mertens, 2014). The researcher considers that there are two key limitations in this research around how 'specific' the treatments in each group can be described:

First, due to practicalities, there were only a limited number of treatment observations carried out for each group (3 in each school for the PM group and 2 in each school for the comparison group). In addition, within the PM groups, only 9 out of a total of 18 pairs of tutees and tutors were closely observed within the sessions. This researcher considers that this only allows a level of *speculation* on treatment integrity across the 7 week intervention period.

Second, only the researcher carried out the treatment observations. Although the researcher attempted to remain as objective as possible, a degree of bias may exist.

5.6.4 Ecological validity

The three schools involved in the research were of varying sizes, locations and standards according to most recent Ofsted inspections (see Chapter 3). However, the sample size was still relatively small and came from within one city in one local authority, which is a typically white-British rural area of the UK. Therefore, it is recognised that the findings of this study may not be generalisable to other contexts- particularly more ethnically and socially diverse populations. Though, Jones and Morgan (2001) consider that, given the limited resources allocated to education research in general -and to mathematics education in particular- small-scale research may be inevitable. Further, as also advocated by Tooley and Darby (1998), it is argued where research is undertaken on a smaller-scale, authors should not make unwarranted generalisations, but suggest how their findings may develop future directions of larger-scale research (Jones and Morgan, 2001). This is the current researcher's aim.

Although the researcher gave the schools criteria for selecting participants, they were not themselves directly involved in selecting the participants. A screening process was also not employed (based on ethical considerations outlined in Chapter 3).The

researcher, therefore, cannot be sure that certain children were not excluded from being selected based on some other criteria used by the school (even though it was discussed with schools how it was important this did not occur). Therefore, it is possible that the sample included was not a true representation of the students struggling with mathematics and who may have benefitted from the intervention. Indeed, more evidence for this possibility comes from the unexpected ceiling effects encountered on the BNST. This threatens the ecological validity of the research.

Although a comparison *and* a control group was included for the RCT examining the impact of the PM intervention for tutees, only a control group was included in the RCT looking at the impact on self-esteem for tutors. Therefore, there was a risk of a Hawthorne effect (explored further in Section 5.7).

Finally, though, the researcher considers their attempt to provide a detailed and thorough account of the procedures involved within the current research increases its ecological validity.

5.7 Further Possible Explanations

Within this research most findings indicate that the null hypotheses cannot be rejected. However, it has been emphasised that this does not mean that one can simply conclude that an intervention is not useful or effective (Burden, 2015), and other possible tentative explanations should be explored. Therefore, in this section, the researcher considers possible explanations that could account for the findings. These explanations were formed by considering the theory, research, and practice already explored throughout, but are by no means an exhaustive list. It is also important, particularly taking into account the methodological limitations already explored, that these are considered *possible explanations*, and not *firm conclusions*.

The researcher first presents possible explanations applicable to all of the research questions, followed by possible explanations associated with each question.

5.7.1 Possible explanations relevant to all research questions

The first possible explanation considered is that cross-age peer-tutoring interventions, (here PM) are simply not effective for secondary populations. The majority of existing research was undertaken in Primary settings, and none, to the author's knowledge with a valid control group. It is possible that there are certain aspects of the secondary age group which mean it is not suitable, or it is only suitable when specific features and environmental characteristics are present (implications of this are discussed later when considering future avenues of research in section 5.10).

However, the second possible explanation considered was that perhaps the PM intervention here, was not actually PM as it was intended. Indeed, of the 9 pairs of students observed during the fidelity checks, the researcher considered that only 2 of the tutors were following the complete PM structure. As highlighted in section 2, the existing research seems to suggest that interventions that are more structured are more beneficial. For example, the reviews from Cohen and colleagues (1982) and Leung (2015) both showed larger effect sizes for academic achievement for structured peer-tutoring interventions compared to unstructured tutoring. Ginsburg-Block, Rohrbeck, and Fantuzzo (2006) found a similar pattern when they reviewed outcomes relating to self-concept (or self-esteem), social skills and behaviour, where tutoring that involved more structured roles led to more positive outcomes. Why specific parts of the PM structure may have been particularly key to the results are further explored later considering each research question in turn.

Another possibility considered was that the intervention may have been too short. Indeed, Mertens (2014) suggests that when researchers cannot refute the null hypothesis, the duration of the intervention is a variable that should be considered as a possible explanation. However, the existing peer-tutoring research has suggested that the intervention should not be any longer than 10 weeks to be most effective (Leung, 2015). Though, while the intention was for the Paired maths intervention to happen 3 times a week for 7 weeks (21 sessions in total) in one school, for example, the sessions only took place 15 times. This was due to occurrences such as staff absence, special school assemblies and trips. As research has shown that the

frequency of the intervention per week does matter, where at least 3 times a week is most effective (Leung, Marsh, & Craven, 2005), perhaps the intervention occurred over a long enough period of time, but in some cases did not occur frequently enough each week.

Another possible explanation considered was that the PM intervention may increase mathematics attainment, but more indirectly through improvement of other skills or behaviours conducive to maths learning not measured in this research. If assessments were carried out a longer period of time after the intervention, some positive delayed effects may have been seen. Indeed, peer-tutoring has been found to improve on task-time behaviour in class, for both tutees and tutors (Ginsburg-Block, Rohrbeck and Fantuzzo 2006; Ginsburg-Block & Fantuzzo, 1997), increased attendance rates, and have positive effects on retention rates (e.g. (House & Wohlt, 1991; Cohen et.al, 1982; Ginsburg-Block, Rohrbeck and Fantuzzo 2006), all of which could arguably increase mathematics attainment. Further, a number of studies have shown that peer-tutoring can, for both tutees and tutors, have a positive effect on attitudes towards both the subject matter being focussed on and other subject areas (Cohen et al.,1982; Brush, 1997; Fogarty & Wang, 1982; Ponzio & Peterson, 1999) . These effects on attitudes could possibly have longer term impacts. Indeed in the Durham Shared Maths Project (Lloyd et al. 2015a), they found when interviewing teachers after the intervention, they reported wider perceived benefits from the intervention - such as improvements in confidence in maths, approaches to problem-solving and social skills. Further, authors thought that these benefits may *in time* help support improvements in learning, and also transfer to other lessons.

A final possible explanation, applicable to all of the research questions, is that the tutors may have needed more time to practice and get more fluent in using the PM framework for the intervention to be effective. Indeed, In Haring and Eaton's (1978) theory around learning- the Instructional Hierarchy- it is proposed that there is a high level of cognitive effort and resources involved in the initial acquisition of a new skill. One can only move up the hierarchy through to levels of fluency and maintenance in that skill after a certain level of experience. Before reaching these fluency and

maintenance levels, there are also bound to be some mistakes and difficulties. However, as each pair observed in the treatment integrity checks was only observed once, and therefore were not observed *over time*, there is no direct evidence to suggest this.

5.7.2 Possible explanations of findings for Research Question 1: Mathematics attainment for Tutees

A number of the key explanations around the findings for mathematics attainment were explored when considering the BNST measure earlier in the chapter. The further possible explanations here centre around the possibility that some of the key elements to successful mathematics instruction and intervention, or learning in general (suggested by the research and theory reviewed in Chapter two) were not present within the Paired Maths intervention in this research.

Although a bank of mathematics activities ranging in difficulty were offered, it was the choice of the participants to decide which activities to attempt. The first possible explanation is that the activities attempted during the intervention were not at the correct difficulty level for the tutee. Indeed, the review from Askew and Colleagues (1997) identified that ensuring a pupil is sufficiently challenged enough is key to mathematics learning. Reflecting on Vygotsky's socio-cultural theory of development, perhaps the activities were not pitched within the zone of proximal development (Vygotsky, 1978). It is argued that for learning to take place, an activity needs to be pitched at a difficulty level within this zone, where a more skilled learner and less skilled learner can cooperate together to construct a shared and meaningful understanding.

Further, in her review of research into numeracy interventions, Dowker (2009) concluded that interventions which have been individualised to the learner by identifying their areas and levels of difficulty, through detailed assessment of a child's learning needs, work best. As tutors were not given any prior information around their tutee's mathematics skills, it is possible that there was a mismatch between the mathematics areas and difficulty explored in the activities and the tutee's maths learning needs.

Perhaps, going forward, tutors need to be given more specific assessment information around their tutee's mathematics skills in order to select activities that are both challenging enough (within their zone of proximal development) and address the most pertinent area of mathematics skills.

Another possible explanation is that mathematics learning did not occur for the tutees, because the quality of the mathematical conversations needed for this were not present. Within the treatment observations, the discussions around the maths being used by the tutor were minimal. The tutee was rarely encouraged to think out loud, to summarize or explain how they arrived at their answers or to reflect on how this could relate to other areas of learning. Piaget and those who follow his doctrine, assert that learning takes place as an outcome of interactions which allow cognitive perspectives to be explored. The resulting "sociocognitive conflict" that comes from this is said to lead individuals to 'elaborate' upon and "reconstruct" their own existing knowledge and thinking processes, and allows for connections between new and prior perspectives to be integrated or added to (Foot & Howe, 1998; Topping & Ehly, 1998; Slavin, 1996; O'Donnell & O'Kelly, 1994). Perhaps here then, there was not enough discussion to allow for this 'sociocognitive' conflict to occur in order for learning to take place. Indeed, this explanation is also supported by Askew and colleagues (1997) who identified that methods which involve purposeful maths discussion -involving pupils' own descriptions of their methods and challenges them to think through explaining, listening, and problem solving- are the most effective for maths learning.

One possible explanation around why these discussions did not take place within the PM intervention group (additional to the explanation explored earlier around the tutor's confidence in following the PM Process) is that perhaps the activities involved did not lend themselves to these discussions. Indeed, a supervising TA expressed to the researcher after the research that she felt many of the activities involved calculations that were too 'closed', or it was obvious what maths was necessary to complete the activity independently. She also expressed that when questions involved the need for more interpretation around what mathematics was needed, she observed more discussion. This could provide some anecdotal evidence for this hypothesis.

5.7.3 Possible explanations for Research Question 2: Impact on self-esteem for tutees

The first possible explanation is one that the researcher considers may seem the most obvious. The literature reviewed in section 2 around why we may see increased self-esteem following an intervention, suggests that it is the associated increased *academic* achievements from an intervention which leads to significant increases in self-esteem (e.g. Bachman and O'Malley, 1977; Byrne, 1986; Helmke and van Aken, 1995, Maruyama et al., 1981, Pottebaum et al., 1986, Skaalvik and Hagtveg, 1990). Therefore, as no statistically significant differences were seen between the PM group, the comparison and control groups in mathematics attainment, it may not be surprising that there were no statistically significant differences found between groups on either measures of self-esteem.

However, Posselthwaite, Semmens and Clarke (submitted for publication 2015) in their recent research in the same Local Authority with primary aged children, found that even though both their comparison group and PM group significantly increased their times table knowledge, only children in the PM intervention showed significant increases in measures of self-esteem. They therefore suggested that there may be something unique to the PM process- irrespective of mathematics attainment- which increases self-esteem. As highlighted in Chapter 2, research has shown that praise which is explicitly linked to effort and approach/strategy has been found to be very effective in terms of promoting self-esteem and motivation (Kamins & Dweck, 1999; Muller & Dweck, 1998). Giving tutee's praise is a key element of the PM process and this was emphasised both in the training given to tutees and features as part of the PM diagram. However, only two of the tutors observed within treatment integrity checks were seen praise their tutees. Therefore it could be that the element of PM key to increasing self-esteem, praise, was not present in this study. This could explain why no statistically significant increases in self-esteem were captured here for tutees in the PM intervention. Indeed, support for this possible explanation comes from the study from Johnson and Bailey (1974) included in the systematic literature review. When carrying out systematic observations of the tutor-tutee interactions, they considered that amount of praise given by the tutors mediated the impact on arithmetic learning.

Finally, as mentioned earlier, although not statistically significant (and therefore should be considered with high levels of caution) some of the results *could* suggest that the PM intervention actually had a *negative* impact on self-esteem in tutees. One possible explanation for this comes from some reflection around Role Theory (Thomas & Biddle, 1929). According to Role Theory, the way an individual behaves is at least partially determined by the role they are given to play in a situation (Thomas & Biddle, 1929). When an individual assumes a role, that individual will be likely to also take on the attitudes and self-perceptions of that role as well as the behaviours, and that this can actually 'spillover' into other situations (Turner, 2002). The researcher considers it a possibility that, because tutors did not provide much positive feedback or praise to tutees, tutees may have viewed themselves more negatively and possibly incompetent in the role of the tutee, which could have led to the findings that self-esteem *may* have decreased more generally for tutees. Though again, the difference seen in the PM group, though interesting, was not found to be statistically significant when compared to the other groups, and so caution *must* be taken when considering this tentative explanation.

5.7.4 Possible explanations for Research Question 3: Impact on self-esteem for tutors

One possible explanation around why tutors did not see any positive gains in self-esteem again comes from some reflection around Role Theory (Thomas & Biddle, 1929). The researcher considers it a possibility that, because of the lack of interactions the tutors had with their tutees, or perhaps an awareness that their support was not successful in helping tutees, tutors did not see themselves more positively or more competent in the role of the tutor. Considering the 'spillover' element of Role Theory, this could be why no increase in self-esteem more generally were seen.

This could also potentially explain the finding that the control participants significantly decreased on the Maths Self-Concept score. By not being selected to take the role of a more competent maths learner to help less competent maths learners, perhaps this could have had a negative impact on their view of their own maths competency. However, it was explicitly explained that both groups had been selected to become a

tutor at some point in time, and the researcher considers that this finding is possibly more likely to be explained by participants trying to please the researcher-a Hawthorne Effect may have occurred.

5.8 Reflections on Real World Research and the Post-Positivist paradigm

Conducting research in the real world, for the educational researcher, involves research within schools, and these are considered to be 'open systems' (Robson, 2011). Real world research is therefore susceptible to influences beyond a researcher's control, as highlighted within the methodological limitations and possible explanations put forward for within the current research. The researcher now considers some key reflections that have arisen during this research on operating within a post-positivist paradigm within real world research.

Firstly, some have questioned whether, in an attempt to be as rigorous as possible in their measurements, post-positivist researchers actually blind themselves to important aspects of the interventions they are trying to study (Haslum, 2007). For example, in the current research, all schools involved reported to the researcher that they had really valued the PM intervention and that the children had really enjoyed themselves. Two of the schools were also considering implementing the intervention with other age groups in their school. The current design did not capture this.

Secondly, although the use of an RCT can control for many threats to internal validity, there are still a number of ethical and pragmatic considerations and also just general variations that occur due to the social context of the research. Many of these were realised and were highlighted earlier in the chapter. High levels of precision and control have not been achievable- an important tenet of the post-positivist paradigm.

Finally, Robinson (1993) argues that there is an important distinction between the espoused theory around an intervention (what does literature say PM is?) and it's theory in action (what does PM actually look like in practice?). The researcher considers that while a fair amount is known around why cross-age mathematics peer-tutoring should work *in theory* for secondary aged pupils, very little is known about whether that can be realised *in practice*. This is reflected in the variability and issues

seen in the treatment fidelity of the PM interventions within the current research. Therefore, to attempt to 'specify' or 'prescribe' a PM intervention for secondary aged children (which is the aim of the post-positivist stance) possibly seems inappropriate as there is such limited research existing which explore the realities of implementing it with this population.

In view of these reflections, the researcher considers that by adopting a constructivist epistemological standpoint instead, where reality exists as it is socially constructed (Mertens, 2014), this could address some of the downfalls experienced within the post-positivist paradigm. This is further explored in the suggested directions for future research to follow.

5.9 Implications for Directions for Future Research

There are several research avenues that the researcher considers would be useful to explore to build on the lessons learned and questions raised by the study. Some of these avenues could be explored using methodologies that exist within the current post-positivist paradigm:

First, the study could be replicated maintaining the same experimental rigour, while altering parts of its design, in an attempt to address some of the contextual and implementation issues experienced in the current research (particularly treatment fidelity). Some additions could include:

- A longer period of training and practice could be offered to allow the tutors to become more fluent and confident in the PM process, before the tutoring began. Some refresher training sessions could also be offered mid way through the intervention period, to ensure continued treatment fidelity throughout.
- The supervising TA in the PM intervention group could also receive more specific training around how to support the tutors and encourage them to follow each stage of the PM process.
- A pilot study could be incorporated, to both look at whether the chosen mathematics activities provide enough scope for the discussion needed in the

PM framework and to see if the chosen mathematics measure is sensitive enough and check for any floor/ ceiling effects.

- Some more specific criteria could be set to help teachers select the pupils to be included in the research. Though, as has been highlighted previously, the researcher did not consider it to be ethically appropriate to involve a screening test in their own research (as they did not wish to test children without the intention to include them in an intervention), some may argue that children undergo tests routinely in schools and that using a screening test may select the pupils most suitable for the intervention.

Second, as this was a relatively small scale study, replicating the current study with a larger sample could increase the reliability of results and reduce the risk of a type II error.

Third, there is the suggestion from other theory and research that a cross-age mathematics peer-tutoring intervention could have a positive impact on other skills or behaviours not measured in this research, which could in turn lead to later positive gains in mathematics attainment. Further research could include the additional measurement of these skills. Further, a third point of measurement for mathematics attainment, a period of time after the intervention has taken place, could be incorporated to try and capture some of these potential delayed effects.

Fourth, further triangulation from multiple sources of data and measures could be included to more accurately, and with more validity, measure mathematics attainment.

Fifth, comparison could be made between groups of tutors receiving different levels training and support, possibly incorporating the idea of giving more detailed assessment data around the tutee's maths attainment. This could help to explore what training best supports tutors, and how much of this support is required before it has an impact on enhancing maths attainment/ self-esteem?

Sixth, a comparison group for the tutors (as well as a control group) could be incorporated, though the nature of this comparison would need careful consideration, beyond the scope of this discussion.

Seventh, a comparison could be made between groups using different types of mathematics activities, to see if there are particular activities that lend themselves more to the successful application of the PM process- perhaps the discussion element in particular.

However, in view of the critical reflections around operating within post-positivism, and it seems that more knowledge is needed around the critical factors needed in the *implementation* of a cross-age mathematics peer-tutoring intervention such as PM with secondary aged pupils, the researcher considers that Action Research could be a valuable approach. Kurt Lewin was the first to introduce the term "Action Research" in 1944. In his 1946 paper: "Action Research and Minority Problems" Lewin described action research as "a comparative research on the conditions and effects of various forms of social action, and research leading to social action." Put more simply, perhaps, Denscombe (2014) describes that Action Research is employed when the aim is to solve a particular problem and to produce guidelines for best practice; to improve the quality of the action in the situation. This involves the researcher, alongside those within the social context of the situation, to engage in 'self-reflective spiral of cycles' of: observing, reflecting, planning, acting and evaluating (Elliott, 1991). This leads to theory around an intervention or situation that is generated and validated through the reflection of the actual practice that is occurring by the individuals who are actively involved in it- rather than just being independently applied. Here, then, an action research project could involve the researcher engaging in these cycles with a secondary school in the implementation of a PM project. Within each 'spiral' learning- from the individuals who are engaging in the intervention- what elements do or do not ensure its successful implementation.

Finally the researcher considers that the current study has been both ethically and ecologically beneficial, as they have attempted to address an important real-world problem- mathematics difficulties. As highlighted earlier, there is quite staggering

research highlighting the long-term detrimental impact of poor mathematics/ numeracy skills, where this can be both irrespective of literacy skills, for both men and women and impact upon a number of key life outcomes including levels of employment, offending and depression (Parsons & Bynner, 2006; Crawford & Cribb, 2013). Therefore, the researcher feels that more research around mathematics intervention in general (possibly particularly around secondary) should be undertaken. If not cross-age peer-tutoring, what other interventions or approaches can help tackle poor mathematics?

5.10 Implications for Educational Psychology Practice

As no *firm* conclusions can be drawn from this research (since findings only meant that the researcher could not reject the Null Hypotheses) the following professional implications presented *must* be considered with caution.

The current research reinforces the importance of evaluating educational interventions and also provides a model for how this can be undertaken. It could also encourage Educational Psychologists (EPs) to use pre- and post- standardised assessments to monitor progress.

However, adequate training and the potential barriers to continuing fidelity of interventions in real life practice should also be a key factor to consider when EPs make recommendations for interventions, not just their empirical research evidence base. These factors are also considered key when contemplating the recommendation of cross-age peer-tutoring interventions in secondary schools.

Indeed, findings indicate that there are further questions to be explored around how/ if PM can implemented successfully in the UK secondary school before it should be recommended as an approach to schools by EPs, particularly considering it's *possible* negative impact upon self-esteem for tutees. This study, therefore, also highlights that EPs cannot simply assume that an intervention which has been shown to be effective with the primary aged population can be applied with the same success within the secondary aged population.

The final key implication for practice is to highlight the importance for EPs to raise the profile of mathematics/ numeracy in their casework. Indeed, while support around poor literacy skills is very common in EP casework, it is vital not to overlook poor mathematics/ numeracy, particularly considering the potentially long-term detrimental impacts (Gersten, Jordan & Flojo, 2005; Parsons & Bynner, 2006; Crawford & Cribb, 2013). Further, while it is key that difficulties and intervention are identified as early as possible, the researcher also considers it important for EPs to encourage practitioners not to 'give up', when children reach secondary age. What is being done to support these children that 'slip through the net' of early intervention at primary school?

5.11 Summary of Discussion

Possible explanations that may account for the findings presented have been explored in light of the literature reviewed (Chapter two) and methodology adopted (Chapter three).

Overall, although the research and theory reviewed throughout this account could suggest that a cross-age peer tutoring intervention in maths (PM) could be beneficial for both children in the role of the tutee and also the tutor in secondary schools, the researcher considers this cannot be inferred from the current research findings. The significant issues around treatment fidelity are recognised as a significant limitation of the research, and suggestions for future research mainly centre around what might aid the successful implementation of a cross-age mathematics peer-tutoring intervention in secondary schools. The researcher has aimed to carefully consider, and provide a detailed account of an ecologically valid design, which they consider to have significantly increased replicability and external validity of the research.

6 Conclusion

6.1 Unique Contribution of This Research

One unique contribution of this research is its attempt to quantitatively analyse the impact of a cross-age peer-tutoring mathematics intervention (PM) for secondary aged pupils, a population where there is a very limited existing body of evidence, particularly around mathematics intervention. The previous research around cross-age peer-tutoring in mathematics has also mostly occurred within the USA, this exploratory study occurred within the UK school context.

Further, to the author's knowledge, this is the first study to use an RCT to evaluate the impact of a cross-age peer-tutoring intervention in maths within the secondary population. RCTs are seen as the *gold standard* of experimental approaches and are ranked highly in the 'hierarchy of evidence' (Burns, Klingbeil, Ysseldyke & Peterson-Brown, 2012; Petticrew & Roberts, 2006).

This study also sought to make a unique contribution in trying to quantitatively examine the affective outcome of self-esteem of a cross-age peer- tutoring mathematics intervention for both the tutor *and* the tutee- something previous literature has suggested might exist, but which lacks rigorous empirical support. This seems particularly important given the well-documented potential negative effects of low-self esteem for both children and adults across many areas of life (Baumeister, Campbell, Krueger and Vohs, 2003), and indeed there has been a call for more research around school-based interventions that promote both academic *and* socioemotional outcomes for students, particularly self-esteem (Durlak, 1995).

A final unique contribution of the study has been to the researcher herself as a research practitioner and applied psychologist. Designing, implementing and reporting on the research has proved a challenging experience in which some of the difficulties of conducting real world research have been encountered. The importance of engaging effectively with stakeholders has been a key learning point along the research journey. It is hoped that the skills and learning gained through the doctoral research will

continue to be influential throughout the researcher's future practice within the EP profession.

6.2 Final Conclusion

This study aimed to examine the impact of a cross-age mathematics peer-tutoring intervention, Paired Maths (PM), for secondary aged pupils in the role of the tutee *and* the tutor. Although no firm conclusions can be drawn from the research with regard to the efficacy of PM in this context, because of the issues highlighted throughout the discussion, the researcher considers that previous literature and theory reviewed suggest that cross-age peer-tutoring in mathematics *could* possibly have a positive impact for secondary-age children. However more research is required to examine what may constitute its successful implementation within this secondary population. The researcher further considers that while RCTs are widely considered the 'gold standard' in research (MacDonald, 1996), when less is known about the 'theory in practice' of an intervention (Robinson, 1993) using methodological approaches such as Action Research may allow for consideration of the real-life practical barriers to the successful implementation- something that is vitally important within Educational Psychology practice.

7 References

- Anderson, C.A., Lindsay, J.J. and Bushman, B.J. (1999). Research in psychological laboratory: Truth or triviality? *Current Directions in Psychological Science*, 8, 3-9.
- Annis, L. F. (1983). The processes and effects of peer tutoring. *Human Learning: Journal of Practical Research & Applications*, 2(1), 39-47.
- Askew, M., Brown, M., Rhodes, V., William, D., & Johnson, D. (1997). *Effective Teachers of Numeracy: Report of a study carried out for the Teacher Training Agency*. London: King's College, University of London.
- Axworthy, D. (2005). *Turning data into information that improves teaching and learning: The WA experience*. Accessed on 18th January 2016 at: http://research.acer.edu.au/cgi/viewcontent.cgi?article=1017&context=research_conference_2005
- Bachman, J. G., & O'Malley, P. M. (1977). Self-esteem in young men: A longitudinal analysis of the impact of educational and occupational attainment. *Journal of Personality and Social Psychology*, 35(6), 365.
- Bandura, A. (1997). Editorial. *American Journal of Health Promotion*, 12(1), 8-10.
- Bar-Eli, N., & Ravia, A. (1982). Underachievers as tutors. *Journal of Educational Research*, 75:139-143.
- Baumeister, R.F., Campbell, J.D., Krueger, J.I. and Vohs, K.D. (2003). 'Does high self-esteem cause better performance, interpersonal success, happiness or healthier lifestyles?' *Psychological Science in the Public Interest*, 4 (1), 1-44
- Beirne-Smith, M. (1991). Peer-tutoring in arithmetic for children with learning disabilities. *Exceptional Children*, 57, 330-337.
- Benware, C. A., & Deci, E. L. (1984). Quality of learning with an active versus passive motivational set. *American Educational Research Journal*, 21(4), 755-765.
- Bland, M., and Harris, G. (1989). Peer tutoring. *School Science Review*, 71, 142-144.
- Bong, M. (2001). Between-and within-domain relations of academic motivation among middle and high school students: Self-efficacy, task value, and achievement goals. *Journal of educational psychology*, 93(1), 23.
- Bowman-Perrott, L., Davis, H., Vannest, K., Williams, L., Greenwood, C., & Parker, R. (2013). Academic benefits of peer tutoring: A meta-analytic review of single-case research. *School Psychology Review*, 42(1), 39-55.

- Brace, N., Kemp, R., & Snelgar, R. (2012). *SPSS for Psychologists (5th Edition)*. Palgrave MacMillan: Basingstoke.
- British Psychological Society (BPS), (2009). *Code of Ethics and Conduct*. Leicester: BPS.
- British Psychological Society (BPS), (2014). *Code of Human Research Ethics*. Leicester: BPS.
- Britz, M. W., Dixon, J., & McLaughlin, T. F. (1989). The effects of peer tutoring on mathematics performance: A recent review. *Journal of Special Education*, 13, 17-33.
- Brush, T. A. (1997). The effects on student achievement and attitudes when using integrated learning systems with cooperative pairs. *Educational Technology Research and Development*, 45, 51-64.
- Bryman, A. (2006). Integrating quantitative and qualitative research: how is it done? *Qualitative research*, 6, (1), 97-113.
- Burden, R. (2015). All that glitters is not gold: Why randomized control trials (RCTs) are of limited value in judging the effectiveness of literacy interventions. *Educational & Child Psychology*, 32, (1), 11- 20.
- Burnett, P. C. (1998) Measuring behavioural indicators of self-esteem in the classroom, *The Journal of Humanistic Education and Development*, 37, 2, 107-116.
- Burnett, P.C. (1994) Self-concept and self-esteem in elementary school children. *Psychology in the Schools*, 11, 164-171.
- Burns, M., Klingbeil, D., Ysseldyke, J., & Peterson-Brown, S. (2012). Trends in Methodological Rigor in Intervention Research Published in School Psychology Journals. *Psychology in the Schools*, 49, (9), 843-852.
- Burr, T. (2008). *Skills for Life: progress in improving adult literacy and numeracy*. Stationary Office: London.
- Button, E.J., Sonuga-Barke, E.J.S., Davies, J., & Thompson, M. (1996). A prospective study of self-esteem in the prediction of eating problems in adolescent schoolgirls: Questionnaire findings. *British Journal of Clinical Psychology*, 35, 193–203.
- Byrne, B. M. (1984) The general academic self-concept nomological network: A review of construct-validation research. *Review of Educational Research*, 54, 427–456
- Byrne, D. (1998). *Complexity Theory and the Social Sciences*. London: Routledge.
- Byrne, B.M. & Shavelson R.J. (1986) On the structure of the adolescent self concept. *Journal of Educational Psychology*, 78, 474–48
- Chen, H., & Rossi, P. (1987). The theory-driven approach to validity. *Evaluation and program planning*, 10, (1), 95-103.

Children Act 2004 (2004). London, HMSO.

Cockcroft, W. H. (1982). Report of the Committee of Inquiry into the Teaching of Mathematics in Schools. London: HMSO

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, 237-248.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research Methods in Education. Seventh Edition*. Routledge: London.

Cook, T., & Campbell, D. (1979). *Quasi-experimentation: design & analysis is-sues for field settings*. Chicago: Rand McNally College Publishing.

Coopersmith, S. (1967). *The antecedents of self-esteem*. San Francisco: W. H. Freeman & Co.

Crawford, C., & Cribb, J. (2013). Reading and maths skills at age 10 and earnings in later life: a brief analysis using the British Cohort Study. *Centre for Analysis of Youth Transitions: Institute of Education, University of London*.

Dancey, C., & Reidy, J. (2014). *Statistics Without Maths for Psychology. Sixth edition*. Pearson: Essex

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, 18(1), 23-45.

Denscombe, M. (2014). *The good research guide: for small-scale social research projects*. McGraw-Hill Education (UK).

Department for Education (2012) *Literacy and numeracy catch-up strategies*. London: Department for Education.

Department for Education (2013) *The National Curriculum in England: Key Stages 1 and 2 framework document*. London: Department for Education

Department for Education (2014) *Mental health and behaviour in schools: Departmental advice for school staff*. London: Department for Education

Department for Education and Employment.(1999) *The National Numeracy Strategy*. London: Department for Education and Employment publications

Department for Education and Skills (2001) *Guidance to Support Pupils with Dyslexia and Dyscalculia*. London: DfES

- deWinter, P. & Cahusac, P. (2014) *Starting out in Statistics: An Introduction for Students of Human Health, Disease, and Psychology*. London: Wiley-Blackwell
- Donaldson, M. (1978) *Children's Minds*. London, Fontana.
- Dowker, A. D. (2004). *Children with Difficulties in Mathematics: What Works?* London: Department for Education and Skills.
- Dowker, A. (2009). *What Works for Children with Mathematical Difficulties? The effectiveness of intervention schemes*. The National Strategies Primary. Department for Children Schools and Families.
- Durlak, J. A. (1995). *School-based prevention programs for children and adolescents* (Vol. 34). Sage Publications.
- Eagly, A. h., & Wood, W. (1994). Using research to plan future research. in h. cooper & L. V. hedges (Eds.), *The handbook of research synthesis* , 485–500. New York: Russell Sage foundation
- Ehri, L. C., Dreyer, L. G., Flugman, B., & Gross, A. (2007). Reading Rescue: An effective tutoring intervention model for language-minority students who are struggling readers in first grade. *American Educational Research Journal*, 44(2), 414-448.
- Elliot, J. (1991). *Action research for educational change*. McGraw-Hill Education (UK).
- Every Child a Chance Trust (2008, with Colin Hudson, Daniel Price and Jean Gross). *The Long Term Costs of Numeracy Difficulties*. London: Every Child a Chance Trust.
- 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: 272-281.
- 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: 331-339.
- Fantuzzo, J. W., Riggio, R. E., Connely, S., & Dimeff, L. A. (1989). Effects of reciprocal peer tutoring on academic achievement and psychological adjustment: A component analysis. *Journal of Educational Psychology*, 81: 173-177.
- Fielding, N. (2009). Going out on a limb: Postmodernism and multiple method research. *Current Sociology*, 57, 427-447
- Fogarty, J. L., & Wang, M. C. (1982). An investigation of the cross-age peer tutoring process: Some implications for instructional design and motivation. *The Elementary School Journal*, 82: 451-469.

- Foot, H., & Howe, C. (1998). The psychoeducational basis of peer-assisted learning. In K. Topping & S. Ehly (Eds.), *Peer-assisted Learning*, Erlbaum, Mahwah, NJ, 27-43.
- Forsyth, D. R., & Kerr, N. A. (1999). Are adaptive illusions adaptive. In *Poster presented at the annual meeting of the American Psychological Association, Boston, MA*.
- French, S.A., Leffert, N., Story, M., Neumark-Sztainer, D., Hannan, P., & Benson, P.L. (2001). Adolescent binge/purge and weight loss behaviors: Associations with developmental assets. *Journal of Adolescent Health, 28*, 211–221
- Fuchs, L. S., Fuchs, D., Bentz, J., Phillips, N. B., & Hamlett, 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.
- Furnham, A., & Cheng, H. (2000). Lay theories of happiness. *Journal of Happiness Studies, 1*, 227–246.
- Furnham, A., & Cheng, H. (2003). Personality, Self-Esteem, and Demographic Predictions of Happiness and Depression. *Personality and Individual Differences, 34*, 921- 942.
- Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research, 79*(3), 1202-1242.
- Gersten, R., Clarke, B., & Mazzocco, M.M.M. (2007). Historical and contemporary perspectives on mathematical learning disabilities. In D. B. Berch & M.M.M. Mazzocco (Eds.), *Why is math so hard for some children? The nature and origins of mathematics learning difficulties and disabilities* (pp. 7-27). Baltimore: Paul H. Brookes.
- Gersten, R., Jordan, N. C. & Flojo, J. R. (2005). Early identification and intervention for student with mathematics difficulties. *Journal of Special Education, 33*, 18-28.
- Ginsburg-Block, M. D & Fantuzzo, J. (1997). Reciprocal peer tutoring: An analysis of "teacher" and "student" interactions as a function of training and experience. *School Psychology Quarterly, 12*, 134.
- 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 Educational Psychology, 90*: 560-569.
- Ginsburg-Block, M. D., Rohrbeck, C. 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*, 732.

Gough, D. & Elbourne, D (2002) Systematic research synthesis to inform policy, practice and democratic debate. *Social Policy and Society* 1: 225-236

Greer, R. D., & Polirstok, S. R. (1982). Collateral gains and short-term maintenance in reading and on task responses by inner-city adolescents as a function of their use of social reinforcement while tutoring. *Journal of Applied Behavior Analysis*, 15(1), 123-139

Gross, J. (2007) Supporting children with gaps in their mathematical understanding: The impact of the National Numeracy Strategy on children who find mathematics difficult. *Educational and Child Psychology*, 24, 146-154

Gutkin, T. B. & Reynolds, C. R. (2009). *The Handbook of School Psychology (4th Ed)*. London: Wiley Publishers.

Hansford, B. C., & Hattie, J. A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research*, 52(1), 123-142.

Haring, N., & Eaton, M. (1978). Systematic instructional procedures: An instructional hierarchy. *The fourth R: Research in the classroom*, 23-40.

Haslum, M. (2007). What kind of evidence do we need for evaluating therapeutic interventions? *Dyslexia*, 13, 234-239.

Higgins J., & Green, S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0. The Cochrane Collaboration.

Heller, L. R., & Fantuzzo, J. W. (1993). Reciprocal peer tutoring and parent partnership: Does parent involvement make a difference? *School Psychology Review*, 22: 517-534.

Helmke, A., & van Aken, M. A. (1995). The causal ordering of academic achievement and self-concept of ability during elementary school: A longitudinal study. *Journal of Educational Psychology*, 87(4), 624.

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

House, J. D., & Wohlt, V. (1991). Effect of tutoring on voluntary school withdrawal of academically underprepared minority students. *Journal of School Psychology*, 29(2), 135-142.

Howell, D. (2013). *Statistical Methods for Psychology*. Wadsworth: Cengage Learning.

Jianghong L., PhD, RN (2004). Childhood externalizing behavior: Theory and Implications, *Journal of Child Adolescent Psychiatry Nurse*. 17(3), 93-103.

- Jenkins, J. R., & Jenkins, L. M. (1982). *Cross-age and peer tutoring*. Reston, VA: Council for Exceptional Children.
- Johnson, M. and Bailey, J. S. (1974) Cross-age tutoring: fifth graders as arithmetic tutors for kindergarten children. *Journal of Applied Behavior Analysis*, 7, 223-232.
- Jones, K., & Morgan, C. (2001). Research in mathematics education: Some issues and some emerging influences. In C. Morgan & K. Jones (Eds.), *Research in Mathematics Education Volume 3: Papers of the British Society for Research into Learning Mathematics* (pp. 1-20). London: British Society for Research into Learning Mathematics.
- Kamins, M. L., & Dweck, C. S. (1999). Person versus process praise and criticism: implications for contingent self-worth and coping. *Developmental psychology*, 35(3), 835.
- Koestner, R., Ryan, R. M., Bernieri, F., & Holt, K. (1984). Setting limits on children's behavior: The differential effects of controlling vs. informational styles on intrinsic motivation and creativity. *Journal of personality*, 52(3), 233-248.
- Lent, R. W., Brown, S. D., & Larkin, K. C. (1986). Self-efficacy in the prediction of academic performance and perceived career options. *Journal of counseling psychology*, 33(3), 265.
- Leung, K. C. (2015). Preliminary empirical model of crucial determinants of best practice for peer tutoring on academic achievement. *Journal of Educational Psychology*, 107(2), 558.
- Leung, K. C., Marsh, H. W., & Craven, R. G. (2009). *Are peer tutoring programs effective in promoting academic achievement and self-concept in educational settings: A meta-analytical review*. SELF Research Centre, University of Western Sydney, Australia.
- Lewin, K. (1944). The dynamics of group action. *Educational leadership*, 1(4), 195-200.
- Lloyd, C., Morris, S., Edovald, T., Skipp, A., Kiss, Z., & Haywood, S. (2015a) *Durham Shared Maths Project Evaluation Report and Executive Summary*. London: Education Endowment Fund.
- MacDonald, G. (1996). Evaluating the effectiveness of social interventions. In A. Oakley and H. Roberts (eds). *Evaluating social interventions: A Report of Two Workshops Funded by the Economic and Social Research Council*. Ilford: Barnardo's.
- Maheady, L., & Gard, J. (2010). Classwide peer tutoring: Practice, theory, research, and personal narrative. *Intervention in School and Clinic*, 46, 71-78.

- Marsh, H. W., & Yeung, A. S. (1998). Longitudinal structural equation models of academic self-concept and achievement: Gender differences in the development of math and English constructs. *American Educational Research Journal*, 35(4), 705-738.
- Maruyama, G. M., Rubin, R. A., & Kingsbury, G. G. (1981). Self-esteem and educational achievement: Independent constructs with a common cause? *Journal of Personality and Social Psychology*, 40(5), 962.
- McCarthy, J. (2008) *Paired Maths Initiative Evaluation Report*. Dublin: Junior Certificate School Programme Support Service
- Mertens, D. (2014). *Research and evaluation in education and psychology*. Sage. London.
- Mertens, D., & Wilson, A. (2012). *Program Evaluation Theory and Practice: A Comprehensive Guide*. Guilford Press. London
- Miller, G.M. and Chapman, J.P. (2001). Misunderstanding Analysis of Covariance. *Journal of Abnormal Psychology*, 110, 1, 40-48.
- Mintz, R., & Muller, D. (1977). Academic achievement as a function of specific and global measures of self-concept. *The Journal of psychology*, 97(1), 53-57.
- Mueller, C. M., & Dweck, C. S. (1998). Praise for intelligence can undermine children's motivation and performance. *Journal of personality and social psychology*, 75(1), 33.
- Muir, T. (2008). Principles of practice and teacher actions: Influences on effective teaching of numeracy. *Mathematics Education Research Journal*, 20(3), 78-101.
- National Numeracy (2015) Numeracy hidden behind literacy. Retrieved from <http://www.nationalnumeracy.org.uk/numeracy-hidden-behind-literacy/index.html>
- National STEM Learning Centre (2015) SMILE Wealth of Worksheets. Retrieved from: <https://www.stem.org.uk/elibrary/resource/25755/smile-wealth-of-worksheets>
- Nazzari, A. (2002). Peer tutoring and at-risk students: An exploratory study. *Action in Teacher Education*, 24(1), 68-80.
- O'Donnell, A. M., & O'Kelly, J. (1994). Learning from peers: Beyond the rhetoric of positive results. *Educational Psychology Review*, 6(4), 321-349.
- Ojose, B. (2008). Applying Piaget's theory of cognitive development to mathematics instruction. *The Mathematics Educator*, 18(1).26-10.

- O'Mara, A. J., Marsh, H. W., Craven, R. G., & Debus, R. L. (2006). Do self-concept interventions make a difference? A synergistic blend of construct validation and meta-analysis. *Educational Psychologist*, 41(3), 181-206.
- Oakley, A. (2002). Social science and evidence-based everything: the case of education. *Educational Review*, 54, 3, 277-286.
- Ofsted. (2005). *The National literacy and numeracy strategies and the primary curriculum*. Crown Copyright: Ofsted Publications Centre
- Ofsted and Department for Education (2014) Ofsted annual report 2013/2014: Schools. Available at: www.ofsted.gov.uk/resources/140178 (Accessed 12.12.14)
- Pallant, J. (2013). *SPSS Survival Manual (5th Edition)*. McGraw Hill: Maiden-head.
- Parsons, S. & Bynner, J. (2006). *Does numeracy matter more?* London: National Research and Development Centre for Adult Literacy and Numeracy.
- Petticrew, M. & Roberts, H. (2006). *Systematic Reviews in the Social Sciences: a practical guide*, Blackwell, Oxford
- Piaget, J., & Inhelder, B. (1962). *The Psychology of the Child*. New York: Basic Books
- PISA (2012). Results in Focus What 15-year-olds know and what they can do with what they know. 2014-12-03]. [http://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview, pdf](http://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf).
- Polirstok, S. R., & Greer, R. D. (1986). A replication of collateral effects and a component analysis of a successful tutoring package for inner-city adolescents. *Education and Treatment of Children*, 101-121.
- Ponzio, R. C., & Peterson, K. D. (1999). Adolescents as effective instructors of child science: Participant perceptions. *Journal of Research and Development in Education*, 33: 36-46.
- Posselthwaite, R., Semmens, C., Clarke, V. (2015) Paired Maths: a study exploring the effectiveness of a cross-age peer tutor approach to improve times-table knowledge and increase self-esteem. *Submitted for Publication*.
- Pottebaum, S. M., Keith, T. Z., & Ehly, S. W. (1986). Is there a causal relation between self-concept and academic achievement?. *The Journal of Educational Research*, 79(3), 140-144
- Reckrut, M. (1994). Peer and cross-age tutoring: The lessons of research. *Journal of Reading*, 37: 356-362.

- Reynolds, D., & Muijs, D. (1999). The effective teaching of mathematics: A review of research. *School Leadership & Management*, 19(3), 273-289.
- Ritchie, J., & Lewis, J. (2003). *Qualitative Research Practice: A guide for Social Science Students and Researchers*. Sage Publications. London.
- Robinson, V. (1993). *Problem-Based Methodology. Research for the improve-ment of practice*. Pergamon Press: Oxford.
- Robinson, N.S., Garber, J., & Hilsman, R. (1995). Cognitions and stress: Direct and moderating effects on depressive versus externalizing symptoms during the junior high school transition. *Journal of Abnormal Psychology*, 104, 453–463
- Robson, C. (2011). *Real World Research (Third Edition)*. Chichester: John Wiley & Sons.
- 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, 240-257.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Roth, A. and Fonagy, P. (1996). *What Works For Whom: A Critical Review Of Psychotherapy Research*. London: Guilford.
- Scheirer, M. A., & Kraut, R. E. (1979). Increasing educational achievement via self concept change. *Review of educational research*, 131-149.
- Schlosser, R. W., & O'Neil-Pirozzi, T. (2006). Problem formulation in evidence-based practice and systematic reviews. *Contemporary Issues in Communication Sciences and Disorders*, 33, 5-10.
- Scruggs, T. E., & Mastropieri, M. A. (1998). Tutoring and students with special needs. *Peer-assisted learning*, 165-182.
- Sharpley, A. M., Irvine, J. W., & Sharpley, C. F. (1983). An examination of the effectiveness of cross-age tutoring program in mathematics for elementary school children. *American Educational Research Journal*, 20: 103-111.
- Shuell, T. J. (1996). Teaching and learning in a classroom context. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 726-764). NY: Simon & Schuster Macmillan.
- Skaalvik, E. M. (1997). Self-enhancing and self-defeating ego orientation: Relations with task and avoidance orientation, achievement, self-perceptions, and anxiety. *Journal of educational psychology*, 89(1), 71.

- Skaalvik, S., & Skaalvik, E. M. (2005). Self-concept, motivational orientation, and help-seeking behavior in mathematics: A study of adults returning to high school. *Social Psychology of Education*, 8(3), 285-302.
- Skaalvik, E. M., & Hagtvet, K. A. (1990). Academic achievement and self-concept: An analysis of causal predominance in a developmental perspective. *Journal of Personality and Social Psychology*, 58(2), 292.
- Skinner, E. A., Wellborn, J. G., & Connell, J. P. (1990). What it takes to do well in school and whether I've got it: A process model of perceived control and children's engagement and achievement in school. *Journal of educational psychology*, 82(1), 22.
- Slavin, R. E. (1996). Research for the future: Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology*, 21, 43-69.
- Slavin, R.E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, 31, 7, 15-21.
- Sturges, J., Rodger, S., & Ozanne, A. (2002). A review of the use of self-report assessment with young children. *The British Journal of Occupational Therapy*, 65(3), 108-116.
- Swanson, H. L., & Hoskyn, M. (1998). Experimental intervention research on students with learning disabilities: A meta-analysis of treatment outcomes. *Review of Educational Research*, 68, 277-321
- Thomas, E. & Biddle, B. (1979) The Nature and History of Role Theory, in B. Biddle & E. Thomas (eds) *Role Theory: Concepts and Research*. Huntingdon, NY: Robert E. Krieger.
- Tooley, J. and Darby, D. (1998). *Educational research - a critique*. London, Office for Standards in Education.
- Topping, K. (1988). *The Peer Tutoring Handbook: Promoting Co-Operative Learning*. Brookline Books, PO Box 1046, Cambridge.
- Topping, K. J. (2005). Trends in peer learning. *Educational psychology*, 25(6), 631-645.
- Topping, K., & Ehly, S. (1998). Introduction to peer-assisted learning. In K. Topping & S. Ehly (Eds.), *Peer-assisted living*, Erlbaum. Mahwah, NJ, pp. 1-23.
- Topping, K. J., Kearney, M., McGee, E., & Pugh, J. (2004). Tutoring in mathematics: A generic method. *Mentoring & Tutoring: Partnership in Learning*, 12(3), 353-370.
- Topping, K. J., Miller, D., Murray, P., Henderson, S., Fortuna, C., & Conlin, N. (2011). Outcomes in a randomised controlled trial of mathematics tutoring. *Educational Research*, 53(1), 51-63.

- Trzesniewski, K.H., Donnellan, M.B., Robins, R.W., Moffitt, T.E., & Caspi, A. (2002, February). *Do juvenile delinquents have high or low self-esteem?* Paper presented at the annual meeting of the Society for Personality and Social Psychology, Savannah, GA.
- Trzesniewski, K. H., Donnellan, M. B., Moffitt, T. E., Robins, R. W., Poulton, R., & Caspi, A. (2006). Low self-esteem during adolescence predicts poor health, criminal behavior, and limited economic prospects during adulthood. *Developmental psychology*, 42(2), 381.
- Turner, R. H. (2002). Role theory. In J. H. Turner (Ed.), *Handbook of sociological theory*, Kluwer Academic/Plenum Publishers, New York, pp. 233-254.
- Tymms, P., Merrell, C., Thurston, A., Andor, J., Topping, K., & Miller, D. (2011). Improving attainment across a whole district: school reform through peer tutoring in a randomized controlled trial. *School effectiveness and school improvement*, 22(3), 265-289.
- The University of York Institute for Effective Education (2015) Peer Learning and Paired Maths. Retrieved from <https://www.york.ac.uk/iee/uk-research/teaching-and-learning/peer-paired-maths/>
- Wilmot, E. (2015). Measurement of teacher knowledge: Unearthing hidden differences among different categories of mathematics teachers in Ghana. *Inter-national Journal of Learning and Development UK*, 3, (2), 39-55.
- Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*, 39(2), 111-133.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge, MA: MIT Press.
- Vickers, A. J. (2005). Parametric versus non-parametric statistics in the analysis of randomized trials with non-normally distributed data. *BMC medical research methodology*, 5(1), 1.

8 Appendices

8.1 A review of the selected literature, by paper, forming the systematic literature review.

Author(s) and Date Published	Location of Study	School Setting	Age of Sample (and any important characteristics)	Sample Size (after attrition)	Design and nature of control	Training for Tutors	Length and frequency of Intervention	Tutoring methods/ resources	Outcomes Investigated	Measures
Johnson and Bailey (1974)	US	Public Elementary school	Tutees: Kindergarten Tutors: 5th graders	Tutees: 5 Tutors: 5 Control Tutee: 5	Quasi-experimental design (by class) Pre/Post Control: only for tutees- instruction as usual	3 x 30 minute session	26x 20 minute sessions over 7.5 weeks (daily)	Matrix of 'skills' in order given to tutors. Flashcards used as a test in each session. Non-mathematical 'play' after	Tutee's Arithmetic skills around "using" numbers 0-100 Tutor interactions	Skills-based arithmetic test constructed and administered by the experimenter Tutor Observation Scale (Niedermeyer, 1970).
Sharpley, Irvine and Sharpley (1983)	US	Public Elementary School	Tutee: 2nd and 3rd graders Tutor: 5th and 6th graders	Tutee: 76 Tutors: 76 Control Tutee: 95 Control for tutor: 63	Quasi-experimental. Pre/post Detail on Control Group (for tutees) Not available	Training given, but not detailed.	30 min/day, 4 days/wk over 5 wks INSTEAD of regular teaching	Not detailed	"Operations" (4 basic operations) and "other mathematics" estimation, geometry, fractions, time Self Esteem (tutors only)	Yardsticks Criterion Referenced Tests in Mathematics (1975) Coopersmith Self-Esteem Inventory (Coopersmith, 196

Author(s) and Date Published	Location of Study	School Setting	Age of Sample (and any important characteristics)	Sample Size (after attrition)	Design and nature of control	Training for Tutors	Length and frequency of Intervention	Tutoring methods/ resources	Outcomes Investigated	Measures
Beirne-Smith (1991)	US	Primary self-contained learning disabilities classrooms	Tutees: 6-10 years (average 8.7) "identified as "learning disabled" Tutors: grades 3-6	Tutees: 20 Tutors: 20 Control: 10	RCT: Tutees randomly allocated to two treatment groups and control group. Tutors randomly allocated to two treatment groups (no control for tutors) Pre/Post	2 x 45minute session	30 minutes (no detail on frequency) for 4 weeks	Tutors trained use of contingent reinforcement, task and error correction procedures. Tutors for Method A three-task, counting-on procedure Tutors for Method B :two-task, rote-memorization procedure	Tutee's knowledge of Single Digit Addition facts	Test around single-digit addition facts constructed by experimenter
Tymms,Merrell ,Thurston, Andor. Topping and Miller 2011	UK (Scotl and)	Primary Schools	Tutee: age 8 (Primary 4 in Scotland) Tutor: age 8 or age 10 (Primary 6 in Scotland) (2 cohorts)	Tutee: 334 Tutor:447 Comparison for tutees: 85	A factorial clustered randomized controlled design: schools Allocated to cross- or same-age tutoring; light or intensive tutoring; maths an/or reading	Training videos given to tutors, but no other specified	Intensive group: 3x per light group: 1 x per week Each session: 30 min 15 weeks in	Tutor and tutee dialogue structured by "Dialog Maths"	Mathematics, reading, vocabulary, nonverbal ability, and attitudes to mathematics and reading.	Performance Indicators in Primary Schools [PIPS

8.2 Paired Maths Diagram

Paired Maths $+$ $-$ $\times$ $\div$ $=$

Doing maths in a pair

Understanding the question

READ - Read the question out loud together

IDENTIFY - Make sure you both know what the question is asking you to do

LISTEN - Let your tutee tell you how they will solve the question

Finding an answer

QUESTION - Ask questions to help the tutee understand the question

THINK OUT LOUD - Give your tutee time to think.

Get them to think out loud

Finishing the question

Check - Do you have the right answer?

Yes - **SUM IT UP** - Get the tutee to explain how they got the answer

No - Go back to the start

LINK IT UP - Could the method we used help us to do a similar problem?

**Remember to tell your tutee
how well they are doing**

Applied Psychology Service

Helpful Hints and Tips!

1 Understanding the Question

Read: Your tutee might be having trouble reading a word question. If so, read it for them.

Identify: Make sure that your tutee understands what the question is asking. Here are a few pointers to help you:

- *To help this happen encourage them to use fingers, counters, cubes, sticks or any other objects to show the reality of the question*
- *Or have them draw dots, a picture, a list, table, diagram, graph or map*
- *Useful things to help might include a number line, a multiplication matrix, and a place value chart*
- *With your tutee's permission, mark their written working out with lines, arrows, colours, or numbering to help them*
- *Have your tutee think of what they have learned before or questions they have solved before, relevant to the current question*
- *Work through a similar but simpler question*
- *How can this question be related to people, places, events and experiences in the life of the tutee? (or those of someone they know or have seen on television)*
- *Make up a similar question using the pupil's own name*
- *Try to use everyday language.*

Listen: Give your tutee time to think about the question and then ask them to explain how they might solve it. *Do not just jump in to fix what you assume the problem is.*

2. Finding an Answer to the Question

Question: Ask helpful and intelligent questions that give clues, to stimulate and guide your tutees thinking. Challenge things that you think are wrong. *However, do not say “that’s wrong!” – ask another question to give a clue. Ask “why?”*

Try to avoid:

- *Closed questions which require only a yes or no answer;*
- *Questions which just rely on memory;*
- *Questions which contain the answer;*
- *The question “Did you understand that?”*
- *Answering your own questions;*
- *Indicating the “difficulty” of any step.*

Examples of questions you may want to ask are:

“What kind of question is this?”

“What are we trying to find out here?”

“Can you state the question in different words or a different way?”

“What important information do we already have?”

“Can we break the question into parts or steps?”

“How did you arrive at that?”

“Does that make sense?”

“Where was the last place you knew you were right?”

“Where do you think you might have gone wrong?”

“What kind of mistake do you think you might have made?”

Praise: Give your tutee praise and encouragement very often, even for a very small success with a single step in solving a question. Keep their confidence high.

Think out loud: Give your tutee some thinking time, before expecting an answer. Encourage them to tell you what they are thinking all the time. Then you will find out where and how they are going wrong. Remember tutors also need time to think! If you are not sure, say so. You are not supposed to know everything.

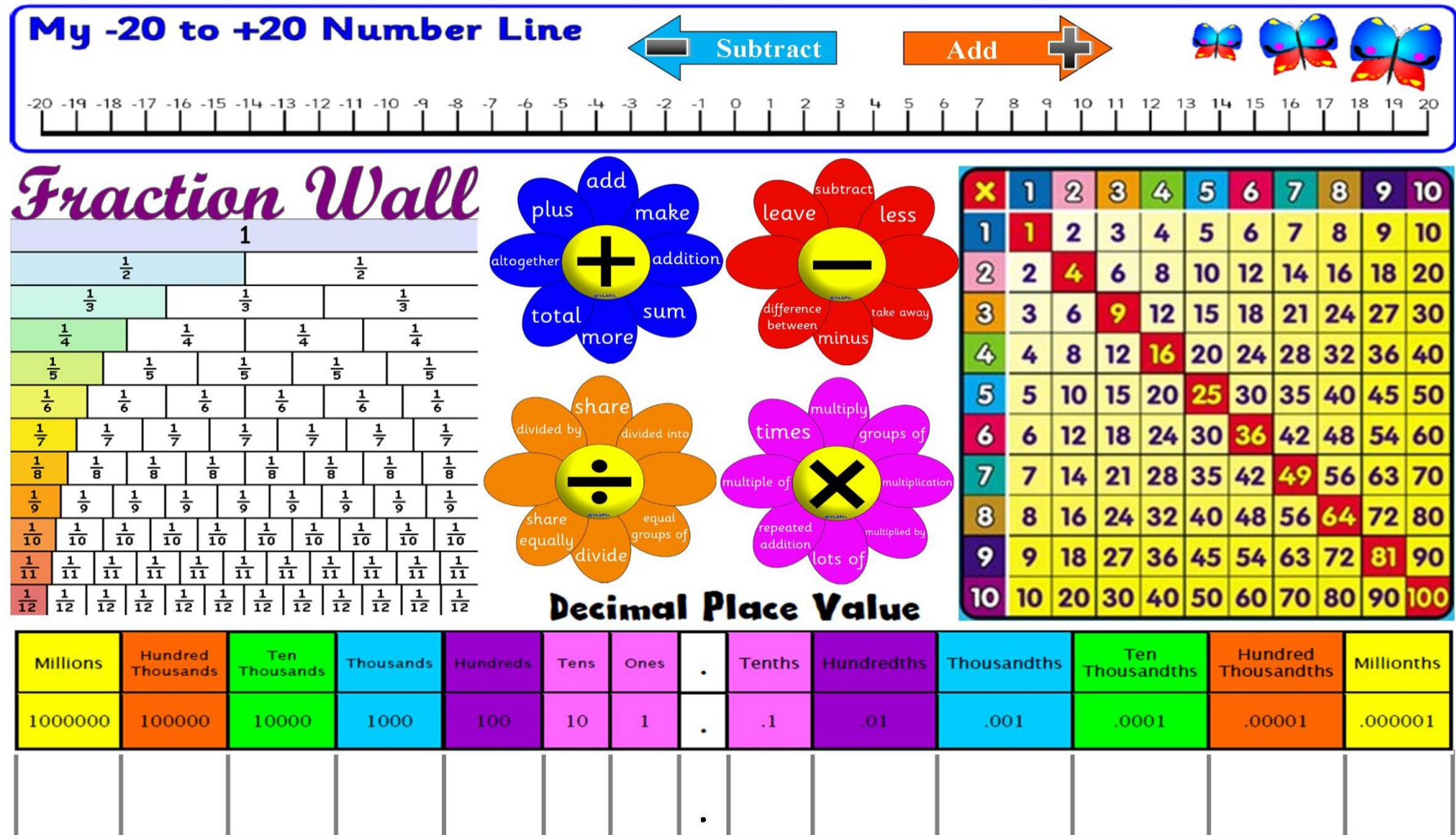
3. Finishing the question :what have they done and how does it link to things they have done in the past?

Check: Check that your tutee eventually gets the right answer. But remember there is probably more than one “right” way to solve the question. *If the answer is wrong read the question over and try again. Only if all else fails show your tutee how you would do it (while you think out loud).*

Sum-it-up: Have your tutee summarise the key steps in doing the maths question. Point out any errors or gaps, then summarise the key steps to yourself.

Link-it-up: The tutor and tutee should talk about how the learning might be used to do another similar question (generalize it to another maths question) or how the learning might be useful in a wider context.

8.4 Maths aid (original is A3 sized)



Brackets (), Others ^{2 3} √, Divide ÷, Multiply x, Add +, Subtract -

8.5 Paired Maths Training Presentation



Paired maths



Aims of this project

- ▶ This project will look at the effects of a paired maths intervention on improving:
 - Maths skills
 - Self-esteem



What is paired maths?

- ▶ Paired Maths is a method of learning.
- ▶ It involves discussion between two pupils (tutor and tutee) to solve the maths question.
- ▶ It focuses on pairs of pupils working together and solving maths questions in three main steps:
 1. Understanding the question
 2. Finding an answer to the question
 3. Finishing the question by asking what have they done and how it links to things they have done in the past



Tutors and tutees

- ▶ Tutors– Year 10
- ▶ Tutees–Year8



The role of the tutor

- ▶ It is the tutor's job to provide support to the tutee.
- ▶ The tutor will help the tutee to attempt the maths question by using a structured approach.
- ▶ This will involve asking the tutee questions.



The role of the tutee

- ▶ It is the job of the tutee to do the actual working in order to answer the maths question.



Paired Maths Diagram



Model the approach



Understanding the question

Read

- ▶ Read the question out loud together.
- ▶ Your tutee might be having trouble reading a word question. If so, read it for them.



Understanding the question

Listen

- ▶ Give your tutee time to think about the question and then ask them to explain how they might solve it.



Finding an answer

Question

- ▶ Ask helpful and intelligent questions that give clues and guide your tutees thinking.
- ▶ Challenge things that you think are wrong.
- ▶ However, do not say "that's wrong!" – ask another question to give a clue. Ask "why?"



Finding an answer

- ▶ **Praise & Encourage** – give your tutee praise and encouragement very often, even for a very small success.
- ▶ **Think out loud** – give your tutee some thinking time, before expecting an answer.
 - Encourage them to tell you what they are thinking all the time.
 - Then you will find out where and how they are going wrong.
 - Remember tutors also need time to think!
 - If you are not sure, say so.
 - You are not supposed to know everything.



Try to avoid

- ▶ Questions which require only a yes or no answer;
- ▶ Questions which just rely on memory;
- ▶ Questions which contain the answer;
- ▶ The question "Did you understand that?";
- ▶ Answering your own questions;
- ▶ Indicating the "difficulty" of any step.



Finishing the question

- ▶ **Check** – check that your tutee eventually gets the right answer.
 - But remember there is probably more than one "right" way to solve the question.
- ▶ If the answer is wrong read the question over and try again. Only if all else fails show your tutee how you would do it (while you think out loud).



Finishing the question

- ▶ **Sum-it-up** – have your tutee summarise the key steps in doing the maths question.
 - Point out any errors or gaps, then summarise the key steps to yourself.
- ▶ **Link-it-up** – the tutor and tutee should talk about how the learning might be used to do another similar question.



Practice the approach

- ▶ Get into pairs
- ▶ One person be the tutor and the other the tutee
- ▶ Practice completing the maths activity sheets



Getting Started

- ▶ Help your tutee pick an activity to carry out from the selection of folders (labelled with difficulty)
- ▶ It may be better to start with something easier to build their confidence
- ▶ Keep all the activity sheets you use in your folder

If you are worried/ are in doubt....ask the supervising member of staff!!

8.6 Guidelines for TA Comparison Group

Guidance for TA Support Group

Dos

- Help students get started... this might involve help with selecting a worksheet to do, getting equipment out etc.
- Have the students work individually on their own worksheets.
- Give support when student's raise their hand e.g:
 - read out the question if struggling with literacy
 - Break down questions into smaller steps
 - Explain a method they could use to work the question out
 - Direct them to their maths mat to help them.
- Keep students on task/ manage behaviour
- Praise where you feel appropriate 😊

Don'ts

- "Teach" the students all together or in smaller groups
- Get them to work in pairs/ smaller groups
- Work out the question for them/ tell them the answers

8.7 Treatment Integrity Checklist for Paired Maths Group

Date:

School (circle) : A B C

General intervention:

General Element of PM Intervention	Present
Facilitator available if needed?	
Intervention Lasted at least 15 minutes?	
Enough maths activities readily available?	
Additional Maths apparatus available?	

Individual Pair of Participants:

Element of PM Intervention	Present?
Did they read the question together?	
Did the tutor clarify the tutees understanding of the question/ explain key terms?	
Did the tutor encourage the use of apparatus?	
Did the tutor praise the tutee?	
Did the tutor encourage 'thinking out loud'?	
Did the tutor check the correctness of the answer?	
Did the tutor model how to get to the correct answer if incorrect?	
Was the tutee asked to 'sum up' how they had arrived at their answer?	
Did the tutor relate the question to real life?	
Did the tutor relate the question to other maths problems?	
Did the tutor help the tutee choose the activity?	

8.8 Burnett Self- Scale

NAME: DATE: PRE/POST

- | | |
|--|--|
| <p>1. ☐ In general I really like myself.
 ☐ In general I like myself.
 ☐ In general I mostly like myself.
 ☐ In general I do not like myself.
 ☐ In general I really do not like myself.</p> | <p>9. ☐ I feel really proud of myself.
 ☐ I feel proud of myself.
 ☐ I sometimes feel proud of myself.
 ☐ I do not feel proud of myself.
 ☐ I really do not feel proud of myself.</p> |
| <p>2. ☐ I really like maths and sums.
 ☐ I like maths and sums.
 ☐ I sometimes like maths and sums.
 ☐ I do not like maths and sums.
 ☐ I really do not like maths and sums.</p> | <p>10. ☐ I really enjoy maths.
 ☐ I enjoy maths.
 ☐ I sometimes enjoy maths.
 ☐ I do not enjoy maths.
 ☐ I really do not enjoy maths.</p> |
| <p>3. ☐ I really like learning new things.
 ☐ I like learning new things.
 ☐ I sometimes like learning new things.
 ☐ I do not like learning new things.
 ☐ I really do not like learning new things.</p> | <p>11. ☐ I really enjoy learning new things.
 ☐ I enjoy learning new things.
 ☐ I sometimes enjoy learning new things.
 ☐ I do not enjoy learning new things.
 ☐ I really do not enjoy learning new things.</p> |
| <p>4. ☐ I feel really good about myself.
 ☐ I feel good about myself.
 ☐ I sometimes feel good about myself.
 ☐ I do not feel good about myself.
 ☐ I really do not feel good about myself.</p> | <p>12. ☐ I really like being the way I am.
 ☐ I like being the way I am.
 ☐ I sometimes like being the way I am.
 ☐ I do not like being the way I am.
 ☐ I really do not like being the way I am.</p> |
| <p>5. ☐ I feel really pleased with myself.
 ☐ I feel pleased with myself.
 ☐ I sometimes feel pleased with myself.
 ☐ I do not feel pleased with myself.
 ☐ I really do not feel pleased with myself.</p> | <p>13. ☐ I feel really confident in myself.
 ☐ I feel confident in myself.
 ☐ I sometimes feel confident in myself.
 ☐ I do not feel confident in myself.
 ☐ I really do not feel confident in myself.</p> |
| <p>6. ☐ I am really good at maths.
 ☐ I am good at maths.
 ☐ I am OK at maths.
 ☐ I am not good at maths.
 ☐ I really am not good at maths.</p> | <p>14. ☐ I get really good marks in maths.
 ☐ I get good marks in maths.
 ☐ I get OK marks in maths.
 ☐ I do not get good marks in maths.
 ☐ I really do not get good marks in maths.</p> |
| <p>7. ☐ I am really good at learning new things.
 ☐ I am good at learning new things.
 ☐ I am OK at learning new things.
 ☐ I am not good at learning new things.
 ☐ I really am not good at learning new things.</p> | <p>15. ☐ I find learning new things really easy.
 ☐ I find learning new things easy.
 ☐ I sometimes find learning new things easy.
 ☐ I find learning new things hard.
 ☐ I find learning new things really hard.</p> |
| <p>8. ☐ I feel really happy with myself.
 ☐ I feel happy with myself.
 ☐ I sometimes feel happy with myself.
 ☐ I do not feel happy with myself.
 ☐ I really do not feel happy with myself.</p> | <p>16. ☐ I feel really satisfied with myself.
 ☐ I feel satisfied with myself.
 ☐ I sometimes feel satisfied with myself.
 ☐ I do not feel satisfied with myself.
 ☐ I really do not feel satisfied with myself.</p> |

8.9 Burnett Self-Scale Scoring sheet

	Self-esteem	Maths	Learning
	1.	2.	3.
	4.		
	5.		
		6.	7.
	8.		
	9.		
		10.	11.
	12.		
	13.		
		14.	15.
	16.		
	Total		
Average score			

8.10 Behavioural Indicator of Self Esteem Scale

Behavioral Indicators of Self-Esteem (BIOS)

Name: _____ Date: _____

Please circle the number that best describes the frequency of this student's behavior over the past two weeks in the school setting.

1 = never 2 = seldom 3 = sometimes 4 = often 5 = always

- | | | | | | |
|--|---|---|---|---|---|
| 1. Was withdrawn from others | 1 | 2 | 3 | 4 | 5 |
| 2. Was silent and non-communicative | 1 | 2 | 3 | 4 | 5 |
| 3. Gave limited responses | 1 | 2 | 3 | 4 | 5 |
| 4. Initiated communication | 1 | 2 | 3 | 4 | 5 |
| 5. Was alone and isolated | 1 | 2 | 3 | 4 | 5 |
| 6. Made excuses for behavior | 1 | 2 | 3 | 4 | 5 |
| 7. Lacked satisfaction with own performance | 1 | 2 | 3 | 4 | 5 |
| 8. Appeared proud of him/herself | 1 | 2 | 3 | 4 | 5 |
| 9. Was confident in what he/she did | 1 | 2 | 3 | 4 | 5 |
| 10. Was open with others | 1 | 2 | 3 | 4 | 5 |
| 11. Appeared satisfied with him/herself | 1 | 2 | 3 | 4 | 5 |
| 12. Was interested in what was happening | 1 | 2 | 3 | 4 | 5 |
| 13. Was able to accept criticism positively | 1 | 2 | 3 | 4 | 5 |
| 14. Was organized and responsible | 1 | 2 | 3 | 4 | 5 |
| 15. Interacted well with other children | 1 | 2 | 3 | 4 | 5 |
| 16. Volunteered information on a regular basis | 1 | 2 | 3 | 4 | 5 |
| 17. Appeared comfortable with routine | 1 | 2 | 3 | 4 | 5 |
| 18. Showed a hesitancy for new situations | 1 | 2 | 3 | 4 | 5 |
| 19. Needed constant reassurance | 1 | 2 | 3 | 4 | 5 |
| 20. Was slow in settling to a task | 1 | 2 | 3 | 4 | 5 |
| 21. Lacked eye-contact when talking | 1 | 2 | 3 | 4 | 5 |
| 22. Volunteered to do something to help | 1 | 2 | 3 | 4 | 5 |
| 23. Was disruptive | 1 | 2 | 3 | 4 | 5 |
| 24. Was attention seeking | 1 | 2 | 3 | 4 | 5 |
| 25. Displayed leadership qualities | 1 | 2 | 3 | 4 | 5 |
| 26. Achieved quietly | 1 | 2 | 3 | 4 | 5 |
| 27. Was committed to the completion of a task | 1 | 2 | 3 | 4 | 5 |
| 28. Displayed good communication skills | 1 | 2 | 3 | 4 | 5 |
| 29. Was interactive with others | 1 | 2 | 3 | 4 | 5 |
| 30. Showed confidence in making a decision | 1 | 2 | 3 | 4 | 5 |
| 31. Appeared happy with him/herself | 1 | 2 | 3 | 4 | 5 |
| 32. Was confident with you and other adults | 1 | 2 | 3 | 4 | 5 |

8.11 Timeline of Research

School A	School B	School C
18th May 2015- Initial Meeting.	18th May 2015 Initial Meeting.	19th May 2015 Initial meeting.
8th June 2015 School Consent forms Collected, discussions around pragmatics and pupil selection	8th June 2015 School Consent forms Collected, discussions around pragmatics and pupil selection	10th June 2015 School Consent forms Collected, discussions around pragmatics and pupil selection
6th July 2015 Collected consent forms, randomly allocated Pupils and meet with TAs	7th July 2015 Collected consent forms, randomly allocated Pupils and meet with TAs	8th July 2015 Collected consent forms, randomly allocated Pupils and meet with TAs
SUMMER HOLIDAYS		
14th September 2015 T1 measures.	15th September 2015 T1 measures.	16th September 2015 T1 measures.
21st September 2015 Training For Year 10 Tutors and Supporting TA	22nd September 2015 Training For Year 10 Tutors and Supporting TA	23rd September 2015 Training For Year 10 Tutors and Supporting TA
Intervention Groups 28th September- 20th November 2015		
23rd November 2015 T2 Measures	24th November 2015 T2 Measures	25th November 2015 T2 Measures
CHRISTMAS HOLIDAYS		
11th January 2016 Training offered to Control Tutees and additional Tutors trained up	12th January 2016 Training offered to Control Tutees and additional Tutors trained up	13th January 2016 Training offered to Control Tutees and additional Tutors trained up
July 2016 Debrief meeting.	July 2016 Debrief meeting	July 2016 Debrief meeting

8.12 Confirmation of Ethical Clearance Letter



The University of
Nottingham

UNITED KINGDOM • CHINA • MALAYSIA

School of Psychology

The University of Nottingham

University Park

Nottingham

NG7 2RD

T: +44 (0)115 8467403 or (0)115 9514344

AS/wb

Ref: 648R

Tuesday, 21 April 2015

Dear Sarah Cooper & Nathan Lambert,

Thank you for submitting an account of your proposed research 'Paired Maths with Secondary Aged Students: a Randomised Control Trial investigating the impact on maths learning and self esteem.'

That proposal has now been reviewed by the Ethics Committee and I am pleased to tell you that your submission has met with the committee's approval.

Final responsibility for ethical conduct of your research rests with you or your supervisor. The Codes of Practice setting out these responsibilities have been published by the British Psychological Society and the University Research Ethics Committee. If you have any concerns whatever during the conduct of your research then you should consult those Codes of Practice. The Committee should be informed immediately should any participant complaints or adverse events arise during the study.

Independently of the Ethics Committee procedures, supervisors also have responsibilities for the risk assessment of projects as detailed in the safety pages of the University web site. Ethics Committee approval does not alter, replace, or remove those responsibilities, nor does it certify that they have been met.

Yours sincerely

Dr Alastair D. Smith Acting Chair, Ethics Committee

8.13 Information Letter and Consent Form for School

Information Sheet for Schools

Research Project on *Paired Maths*

Researcher: Sarah Cooper lpxsc8@nottingham.ac.uk

University Supervisor: Dr Nathan Lambert nathan.lambert@nottingham.ac.uk

Supervising Senior Educational Psychologist: (anonymised)

Dear *****,

This is an invitation to take part in a research study to evaluate 'Paired Maths'.

Paired maths is a structured form of peer tutoring used to improve maths knowledge/skills. Peer-tutoring is an approach identified in the Sutton Trust Pupil Premium report as being of great significance in raising pupil attainment (*Toolkit of Strategies to Improve Learning*, Higgins, Kokotsaki and Coe, 2011) and involves pairs of children using materials to facilitate and scaffold learning. Within these pairs, one peer is more knowledgeable (the tutor) and therefore supports the learning of the less knowledgeable peer (the tutee). This approach focuses on the process of learning, and involves the tutor-tutee pairings working through a range of maths activities following a specific framework that helps to facilitate and scaffold the learning of the tutee.

There have been a number of studies that have indicated Paired Maths programmes can show improvements on assessments of math skills, in addition to improving attitudes to maths, increasing self-esteem and developing social interaction skills, especially with at risk and low achieving children in primary schools (Topping, 2005; Topping et al., 2003; Thurston and Topping, 2011).

Recently, some research was carried out within this local authority to investigate the effect of a paired maths intervention on improving maths knowledge (for tutees) and self-esteem (for tutees AND tutors) within primary school settings, with very encouraging results. The aim of this research would be to investigate whether the effects can too be seen in secondary schools.

More specifically this research aims to investigate if a 7 week Paired Maths intervention programme can help enhance:

- Maths attainment in Year 8 students (currently in Year7) in the role of Tutees.

- Self-esteem in both Year 8 students (currently in Year 7) in the role Tutees and Year 10 students (currently in year 9) in the role of Tutors. .

I feel that this could be a beneficial intervention to both your students and your staff. Before you decide if you wish to take part, it is important for you to understand what the research will involve. Please take time to read the following information carefully.

The intervention will be aimed at mixed year 8 pupils (for the role of the tutee) who are struggling in maths. It will also involve mixed year 10 students (for the role of the tutor). I will provide a criterion to help you make these decisions, should you wish to take part.

Year 8 pupils will be allocated to one of 3 groups (A, B or C). Group A will receive the Paired Maths intervention, taking the role of the Tutee; Group B will receive a comparison intervention, where they will be supported in small groups by a Teaching Assistant, and Group C will receive their usual maths teaching (autumn term, 2015).

Year 10 pupils will be allocated to one of 2 groups (A or B). Group A take part in the Paired Maths intervention, taking the role of the tutor and group B, who will receive teaching as usual (autumn term, 2015)

The allocation to groups will be conducted *randomly*. This is necessary to ensure the research is highly valid. To ensure equal treatment, groups B and C in Year 8 and group B in Year 10 will also be given the opportunity to receive Paired Maths after the spring term of 2016.

One teaching assistant will need to be available to oversee and supervise the Paired Maths interventions for 15 minutes 3 times a week, initially over 7 weeks from 28.09.15 and again for 7 weeks in the spring term. They will also need to be available for a 2 hour training session. In addition, 1 Teaching assistant will need to be available to support the comparison group B for 15 minutes, 3 times a week over the same initial 7 week period in the autumn term. Training will also be provided for the teaching assistant. Resources needed for the maths activities will be provided by myself (apart from writing equipment and paper).

In order to evaluate paired maths, self-esteem in both Year 8 and Year 10 will be assessed in two ways: (1) the student's tutor will be asked to fill out a questionnaire the students current behaviour in school and (2) the students will be asked to complete an individual questionnaire around their self esteem (these will be conducted on a small group basis by myself). Year 7s (to be Year 8s) will also complete the Basic Number Screening Test as a measure of maths attainment.

All types of assessments will occur twice, once before the initial intervention phase on 23.09.15, then after 7 weeks (and half term) at the end of the initial intervention stage on 25.11.15.

All data that I collect will be anonymised and remain confidential. Individual scores will be only be presented as group scores. These will be communicated to yourselves and parents/pupils in writing, with the option to attend a debrief meeting after the research has concluded. The findings and implications in full, as part of my doctoral research, will be published at a later date.

I hope that you feel you are able to help with this study. If you do consent but later feel that you do not wish to take part, at any stage, your school is free to withdraw.

If you would like to discuss the research further, please contact me on ***** or ***** and I will be more than happy to do this. Thank you for taking the time to consider the research.

Yours Sincerely,

A handwritten signature in dark ink, appearing to read 'Sarah Cooper', on a light yellow background.

Sarah Cooper

Trainee Educational Psychologist and student at the University of Nottingham

Research Supervisor: Dr Nathan Lambert

If you have any complaints about the study, please contact:

Stephen Jackson (Chair of Ethics Committee)

stephen.jackson@nottingham.ac.uk

School of Psychology

Consent Form: School



**The University of
Nottingham**

UNITED KINGDOM • CHINA • MALAYSIA

Ethics Approval Number or Taught Project Archive Number:

Researcher(s):...Sarah Cooper lpxsc8@nottingham.ac.uk

Supervisor(s):Dr Nathan Lambert nathan.lamber@nottingham.ac.uk

*Supervising Senior Educational Psychologist ****

The participant should answer these questions independently:

- Have you read and understood the School Information Sheet? YES/NO
- Have you had the opportunity to ask questions about the study? YES/NO
- Have all your questions been answered satisfactorily? YES/NO
- Do you understand that the school is free to withdraw from the study? YES/NO
(at any time and without giving a reason)
- I give permission for the data from this study to be shared with other
researchers provided that their anonymity is completely protected. YES/NO
- Do you agree for your school to take part in the study? YES/NO

“This study has been explained to me to my satisfaction, and I agree for our school to take part. I understand that our school is free to withdraw at any time.”

Signature of Staff Member:

Date:

Name (in block capitals)

I have explained the study to the above staff member and he/she has agreed for their school to take part.

Signature of researcher:

Date:

8.14 Information Letter and Consent Form for Parents

Information Sheet for Parents/Carers

Research Project on Paired Maths

Researcher: Sarah Cooper lpxsc8@nottingham.ac.uk

University Supervisor: Dr Nathan Lambert nathan.lambert@nottingham.ac.uk

*Supervising Senior Educational Psychologist ****

Dear Parents/Carers

Your child has the opportunity to participate in a mathematics intervention called Paired Maths that is being offered to a small number of students currently in year 7 and year 9 during the next academic year when they will be in year 8 and year 10.

Paired Maths is a structured form of peer tutoring used to address particular maths skills. This involves pairs of children, where one peer is more knowledgeable (the tutor) and therefore supports the learning of the less knowledgeable peer (the tutee). The tutor-tutee pairings work together through a range of maths activities following a specific framework that helps to facilitate and scaffold the learning of the tutee, tapping into children's 'learning to learn' skills (meta-cognition).

Recently, some research was carried out within the local authority to investigate the effects of a Paired Maths intervention on improving maths knowledge (for tutees) and self-esteem (for tutees AND tutors) within primary school settings, with very encouraging results. The aim of this research would be to investigate whether the effects can too be seen in secondary schools.

Before you decide if you wish for your child to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully.

This research aims to investigate if a 7 week Paired Maths intervention programme can help enhance:

- Mathematics attainment in Year 8 students in the role of Tutees.
- Self-esteem in both Year 8 students in the role Tutees and Year 10 students in the role of Tutors.

If your child participates and is currently in Year 7, they will be allocated to one of 3 groups (A, B or C). Group A will receive the Paired Maths intervention (taking the role of the Tutee); Group B will take part in a comparison group where they will spend the

same time on the same maths activities, but in small groups with a teaching assistant, and Group C will receive their teaching as normal.

If your child participates and is currently in Year 9, they will be allocated to one of 2 groups (A or B). Group A take part in the Paired Maths intervention, taking the role of the Tutee and Group B will received their teaching as normal.

The allocation to groups will be conducted *randomly*. This is necessary to ensure the research is highly valid. To ensure equal treatment, groups B and C currently in Year 7 and group B currently in Year 9 will also be given the opportunity to receive Paired Maths after Christmas.

The first round of intervention groups will commence for 7 weeks beginning in the Autumn term(28.09.15- 20.10.15). The Paired Maths group A and Comparison group B will be facilitated by staff who currently work within school. These staff members will receive training from myself before the start of the interventions. The groups will meet for 15 minutes 3 times a week.

In order to measure progress, all Year 8s and Year 10s (currently Year 7s and Year 9s) in the project will meet with me twice. Once before the initial intervention phase during the week beginning 21.09.15, then at the end of the initial intervention stage during the week beginning the 23.10.15. Your child will complete a questionnaire exploring elements of self esteem, this should take about half an hour and will be done individually within small groups. In addition, your child's teacher will be asked to complete a questionnaire exploring their observations of your child's behaviours which are seen as indicative of self-esteem on the above dates.

If your child is currently in Year 7, they will also meet with me on these dates to measure their mathematics skills. This will involve answering a selection of number skill questions, starting with some easy questions and increasing in difficulty. These questions will be given orally in a group context, but the children will work independently.

The research will analyse *group* progress on the measures to evaluate the effectiveness of Paired Maths. The progress of individual pupils will therefore not be available, only the progress of the groups. These results will be sent by letter to you after the research has concluded. This letter will also provide a date for a feedback session, should you wish to meet with me to discuss any aspects of the research further. The findings and implications in full, as part of my doctoral research, will be published at a later date.

It is important to be aware that participation in this study is voluntary and your child is under no obligation to take part. Your child is free to withdraw at any point during the

study. All of the data collected will be kept confidential, anonymous and used for research purposes only.

Finally, we ask that you ensure that your child fully understands the procedures specified on his/her information sheet before signing the consent form attached. We also ask that your child signs/writes his/her name on their form to indicate that he/she is aware of the procedures and happy to take part.

The consent forms will then need to be returned to***by Friday 3rd of July for your child to take part.

Throughout research I'll be working with **** from school. If you have any questions or concerns please do not hesitate to ask either of us further. I am contactable on the above email or through *****.

Thank you for your time,

Yours Faithfully,

A handwritten signature in black ink, appearing to read 'Sarah Cooper', on a light yellow background.

Sarah Cooper

Trainee Educational Psychologist and student at the University of Nottingham

Research Supervisor: Dr Nathan Lambert

If you have any complaints about the study, please contact:

Stephen Jackson (Chair of Ethics Committee)

stephen.jackson@nottingham.ac.uk

School of Psychology

Consent Form:

Parent(s)/Carer(s)



**The University of
Nottingham**

UNITED KINGDOM • CHINA • MALAYSIA

Ethics Approval Number or Taught Project Archive Number:

Researcher(s):...Sarah Cooper lpxsc8@nottingham.ac.uk

Supervisor(s):Dr Nathan Lambert nathan.lamber@nottingham.ac.uk

*Supervising Senior Educational Psychologist ****

The parent/carers should answer these questions independently:

- Have you read and understood the Parent/ Carer Information Sheet? YES/NO
- Have you had the opportunity to ask questions about the study? YES/NO
- Have all your questions been answered satisfactorily? YES/NO
- Do you understand that your child is free to withdraw from the study?
YES/NO(at any time and without giving a reason)
- I give permission for my child's data from this study to be shared with other
researchers provided that their anonymity is completely protected. YES/NO
- Do you agree for your child to take part in the study? YES/NO

"This study has been explained to me to my satisfaction, and I agree for my child to take part. I understand that my child is free to withdraw at any time."

Signature of the Parent/ Carer:

Date:

Name (in block capitals):

Parent/Carer of:

I have explained the study to the above parent/carers and he/she has agreed for their child to take part.

Signature of researcher:

Date:

8.15 Information Letter and Consent Form for Participants

Information Sheet for Pupils

Research Project on *Paired Maths*

Researcher: Sarah Cooper lpjsc8@nottingham.ac.uk

University Supervisor: Dr Nathan Lambert nathan.lambert@nottingham.ac.uk

*Supervising Senior Educational Psychologist: *****

Dear Student,

You and some of your fellow students in Year 8 and Year 10 have been invited to take part in a project that looks at a maths programme called 'Paired Maths.' The project aims to find out if Paired Maths is a good way to help children improve their maths skills and their self esteem (how they see themselves).

This letter will tell you a little bit about the project to help you decide if you would like to take part. If you are unsure about anything or have any questions, please ask ***** or someone at home.

Paired Maths involves pairs of students working together, where one student (the tutor) helps and supports the other (the tutee) to solve maths problems. The tutors are trained to support the tutee by following a set of steps which help them to explore questions and think about how they are answering them.

If you take part, Paired Maths will take place for 15 minutes, 3 times a week for 7 weeks. Current Year 7 (you will be in Year 8 when you do the intervention) students will be paired with and helped by Current Year 9 students (You will be in Year 10 when you do the intervention) to work through some activities aimed at helping the Year 8 students to improve their maths skills. Year 10 students will attend a 2 hour session before the Paired Maths starts to train how to be a good tutor.

If you choose to take part and you are a Year 7: You will be placed in 1 of 3 groups. You will find that some groups in your school will do Paired Maths straight away, whereas others won't. Do not worry if you do not, this means that you will be doing Paired Maths later on. Some groups will work in a small group on some maths activities 3 times a week with a Teaching Assistant before they have Paired Maths.. The group you are in will be picked out of a hat and you will not be able to change from this group I'm afraid!

If you choose to take part and you are a Year 9: You will be placed in 1 of 2 groups. Some students will be trained (involving a 2 hour training session) and be a tutor for the Paired Maths intervention straight away, whereas for other students this will

happen later on. Don't worry, all Year 10s on the project will be offered the chance to train and become a tutor for Paired Maths.

To help find out how good Paired Maths is, all Year 8s and Year 10s in the project will meet with me twice. Here you will do a questionnaire which asks you some questions about what you think about yourself, this should take about half an hour and will be done in a small group (you will be doing it on your own though and won't have to share your answers with any other students). Your teacher will also be asked to fill out a questionnaire which will ask them to think about how you are in school.

If you are a Year 8, you will also meet with me twice to look at your number skills. Here, I will ask you to have a go at some number questions. Some of the questions will be easy, and some of the questions will be a bit trickier, but you will only be asked to try your best! I'll read out the questions to you in a group, but you will be working out and answering the questions on your own. This part will also take around half an hour.

When the research has finished I will let you know what we found out about your group. To do this your parents/carers will receive a letter from me. I will be unable to tell you how you did alone.

If you do not wish to take part, that is ok. If you do wish to take part, you will need to complete the consent form with your parent/carer. You are free to leave the study at any time if you decide to. If you decide at a later time that you do not wish to take part you can do this by speaking to ***** or your parents/carers.

Again, if you are unsure about anything or have any questions, please feel free to ask **** or someone at home.

Yours Faithfully,

A handwritten signature in blue ink, appearing to read 'Sarah Cooper', with a stylized flourish at the end.

Sarah Cooper

Trainee Educational Psychologist and student at the University of Nottingham

Research Supervisor: Dr Nathan Lambert

If you have any complaints about the study, please contact: Stephen Jackson (Chair of Ethics Committee) stephen.jackson@nottingham.ac.uk

School of Psychology

Consent Form:

Pupils

Research Project on Paired Maths

Investigators: Sarah Cooper; Dr Nathan Lambert; Sarah D Cooper

School of Psychology, University of Nottingham

The participant should complete the whole of this sheet himself/herself. Please cross out as necessary

- Have you read and understood the participant information sheet? YES/NO
- Have you had the opportunity to ask questions and discuss the study YES/NO
- Have all the questions been answered clearly? YES/NO
- Have you received enough information about the study? YES/NO
- Do you understand that you are free to withdraw from the study:
At any time? YES/NO
Without having to give a reason? YES/NO
- Do you agree to take part in the study? YES/NO

“This study has been explained to me to my satisfaction, and I agree to take part. I understand that I am free to withdraw at any time.”

Signature of the Participant:

Date:

Name (in block capitals):

I have explained the study to the above participant and he/she has agreed to take part.

Signature of researcher:

Date:

8.16 Information Letter and Consent Form for Intervention Facilitators

School of Psychology

Information Sheet for School Staff

Research Project on *Paired Maths*

Researcher: Sarah Cooper lpjsc8@nottingham.ac.uk

University Supervisor: Dr Nathan Lambert nathan.lambert@nottingham.ac.uk

Supervising Senior Educational Psychologist: Sarah D Cooper,
sarahd.cooper@lincolnshire.gov.uk

Dear Staff Member,

Your school has been invited, and accepted to take part in a research study evaluating 'Paired Maths'. Paired Maths is a structured form of peer tutoring, where two students, one in the role of the tutor and one in the role of the tutee, work together to solve maths questions. The tutor provides support and mediates the learning processes for the tutee through discussions following a set framework.

This research aims to investigate if a 7 week Paired Maths intervention programme can help enhance:

- Mathematics attainment in Year 8 students in the role of Tutees
- Self-esteem in Year 8 students in the role Tutees and also in Year 10 students in the role of Tutors

Your school SENCo, felt that you may wish to take part in supervising the Paired Maths intervention or supporting a comparison group as part of this research. Before you decide if you wish to take part, it is important for you to understand what the research will involve. Please take time to read the following information carefully.

The intervention will be aimed at mixed year 7 pupils that the school feels are struggling in maths and would benefit from intervention that could enhance self-esteem. It will also be aimed at mixed year 9 students (for the role of the tutor) who could benefit from intervention that could enhance self esteem.

Year 7 pupils will be allocated to one of 3 groups (A, B or C). Group A will receive the Paired Maths intervention, taking the role of the Tutee; Group B will receive a comparison intervention, where they will be supported in small groups by a Teaching Assistant, and Group C will receive their usual maths teaching. Year 9 pupils will be allocated to one of 2 groups (A or B). Group A take part in the Paired Maths intervention, taking the role of the Tutee and group B, who will receive teaching as usual.

The allocation to groups will be conducted *randomly*. This is necessary to ensure the research is highly valid. To ensure equal treatment, groups B and C in Year 7 and group B in Year 9 will also be given the opportunity to receive Paired Maths after the Christmas holidays.

If you decide to participate, you will be allocated to either:

1. Supervise the Paired maths intervention. This will involve providing supervision and support to six pairs of yr8-yr10 students working through maths problems using the Paired Maths framework , 3 times a week for 15 minutes for 7 weeks from 28.09.15 to 20.11.15. Before the intervention begins, there will be a 1 hour training session for yourself and the yr 10 tutors around the Paired Maths intervention. All resources and activities for the intervention will be provided by myself.

OR 2.Run the small comparison group. This will involve supporting a group of six year 8 students to work individually through the same maths problems This will also occur 3 times a week for 15 minutes for 6 weeks from 28.09.15 to 20.11.15.Details of the type of support to be expected will be given.

In order to evaluate the intervention, I will come into school before, during and after the intervention and work with the students taking part in all of the groups. I will also come into school on 3 occasions over the 6 weeks to observe the sessions. The purpose of these visits is **not** to make a judgement on teaching quality; it is to support you and to evaluate how the training given (mainly to the Year 10 tutors) is reflected in practice. The dates of these visits will be arranged with you in advance.

As part of the research, if you do consent to the observations and inclusion of this data in the research, but later feel that you do not wish for this to happen, at any stage, you are free to withdraw from this.

As supporting either the Paired Maths or comparison group will be part of school practice, if you consent to supporting a group but later feel that you do not wish to take part, you will need to raise this with your school SENCO.

I hope that you feel you are able to help with this study. To participate, you will need to complete the consent form attached. If you would like to discuss the research further, please contact me on *****, or the email address above and I will be more than happy to do this. Thank you for taking the time to consider the research.

Yours faithfully,

A handwritten signature in dark ink on a light yellow background. The signature is stylized and appears to read 'Sarah Cooper'.

Sarah Cooper

Trainee Educational Psychologist and student at the University of Nottingham
Research Supervisor: Dr Nathan Lambert

If you have any complaints about the study, please contact:

Stephen Jackson (Chair of Ethics Committee)

stephen.jackson@nottingham.ac.uk

8.17 Debriefing Letter

University of Nottingham

School of Psychology

Debriefing Sheet for Parents/Carers/ Participants

Research Project on *Paired Maths*

Researcher: Sarah Cooper lpjsc8@nottingham.ac.uk

University Supervisor: Dr Nathan Lambert Nathan.lambert@nottingham.ac.uk

Supervising Senior Educational Psychologist: *****

Dear Parents/Carers and Pupil,

I would like to firstly thank you for agreeing to take part in the research. Below is a brief summary of the outcomes. In addition, if you would like to know more about the research, or have any queries or concerns, the details of an optional debriefing session are given.

At the beginning of the research, Year 8 pupils were allocated into either: group A (Paired Maths), group B (Comparison group support) or group C (Instructions as usual). Results from the research indicated that all groups showed an increase in mathematics skills, but there were no statistically significant differences between the groups on either the mathematics or self-esteem measures taken .This was unexpected and may be due to the small sample size used and some of the difficulties experienced in the implementation of the intervention.

Year 10 pupils were allocated into either: group A (Paired Maths) or group B (instruction as normal). Results from the research indicated that the intervention group did not show statistically significant increases in comparison to group B . This was also unexpected and could be due to the small sample size used.

If you wish to discuss the research any further, a debriefing session will be held at **XXXXXX** in school on **XXXXXXXXXXXXX**. Both the researcher and **(SENCO)** will be available at this time to answer any further questions that you may have.

I would again like to thank you for agreeing to participate in the research.

Thank you for your time,

Yours Faithfully,

A handwritten signature in black ink, appearing to read 'Sarah Cooper', on a light yellow rectangular background.

Sarah Cooper

Trainee Educational Psychologist and student at the University of Nottingham

Research Supervisor: Dr Nathan Lambert

If you have any complaints about the study, please contact:

Stephen Jackson (Chair of Ethics Committee)

stephen.jackson@nottingham.ac.uk