Case Study 1: An Evidence-Based Practice Review Report

Theme: Interventions Implemented by Parents

How effective are numeracy interventions implemented by parents at helping their children increase mathematical knowledge?

Summary

Numeracy interventions (NIs) implemented by parents are defined as any type of systematic intervention that focuses on strengthening mathematical knowledge and mathematical understanding and consequently, academic attainment through the active participation of parents with their children. This systematic literature review aims to evaluate the effectiveness of four different NIs conducted between 1999 and 2014 with pre-secondary typically developing school children. Although the effect sizes ranged from low to medium, thus showing that the effectiveness of these interventions are only moderate; the findings also suggest that their flexible and affordable models mean they could work well as potential programs to enhance children's readiness for school.

Introduction

Description of numeracy interventions

In this review NIs are understood as any structured and systematic programme implemented at home by a parent (or significant adult) to help children improve their mathematical knowledge and consequently their academic performance.

In this review ‘mathematical knowledge’ will be used interchangeably with ‘numeracy’. Its definition has been taken from the National Numeracy Strategy for primary school (DfEE,1999):

‘*Numeracy is a proficiency which involves confidence and competence with numbers and measures. It requires an understanding of the number system, a repertoire of computational skills and an inclination and ability to solve number problems in a variety of contexts*’ (p.4).

The four interventions examined similar areas of numeracy which made the assessment of their methodology easier. For example, three of the four studies focus on developing children’s mathematical knowledge through strengthening their number and mathematical skills, in order to increase their readiness for school. As most of these studies focus on pre-school mathematics, basic skills such as: reciting and recognising numbers from 1 to 10; understanding number concepts from 1 to 5; understanding simple mathematical concepts of size and position; and using appropriate terminology to describe portions, were targeted for improvement (Ford, McDougall, & Evans, 2009; Lavelle-Lore, 2013; Starkey & Klein, 2000). The only study to address more advanced problem solving skills aimed to bridge the gap between ‘school maths’ and ‘real-life maths’. Children were taught to make sense of number problems and identify the operations needed to solve them (Topping, Kearney, McGee, & Pugh, 2004).

All the studies required parents to have a short period of training to ensure the fidelity of the implementation. In addition, the interventions consisted of parents sharing time with their children on a one to one basis. This included playing mathematical games or solving real life problems (e.g. what time do we need to leave if the train departs at 4:45 and the station is 5 miles away?).

Even though all the studies aimed to improve children's mathematical knowledge and academic performance, not all of them had underachieving pupils in their samples. Instead, most studies consisted of children who could be at risk of underachievement in the future, e.g. families from low SES and ethnic minority groups. Therefore, most studies intended to tackle early mathematical performance to avoid future numeracy issues.

Psychological Theory

Children are said to be born with a natural inclination to understand life in 'mathematical terms'. Butterworth (1999) suggests that mathematics is 'as natural as breathing' and by way of example, says that without numbers people couldn't even enjoy their favourite sports, as they all use numbers to keep the score. Furthermore, basic laws of perception could not be understood without the use of mathematical concepts either. For example, to see a light get brighter and realise its absolute increase in energy requires the notion of quantity.

Piaget (1973) argued that cognitive development also embraces the comprehension of mathematical concepts. In his Pre-operational stage, he suggests that two year-olds have the capacity to start learning to count from 1 to 10. However, only when children reach the ages of 7-11 and move into the Concrete operations stage, they can understand that the numeral, e.g. '1' actually stands for one object.

Conversely to Piaget's thinking, modern research has shown that numeracy development begins as early as nine months (Blevins-Knabe & Musun-Miller, 1996) and by the age of two, children show knowledge of how different numbers/words relate to different amount of objects (Aunio et al., 2008).

For many years research has shown that learning is most effective when it is done through a thoughtfully and thoroughly 'scaffolded' tutoring procedure (Topping, 2000). For this reason parental involvement has been considered to be a key ingredient in children's academic and personal development.

In England and the United States, both policy oriented and empirical research have demonstrated how parents' support at home has an important effect in their children's academic success, as well as their general wellbeing. These studies suggest that children with supportive parents enjoy school more, attend lessons more regularly, get better academic results and have fewer behavioural problems (DFES, 2007; Harris, 2008). Furthermore, parents' involvement goes beyond academic achievement as it also has a positive impact on their children's social skills, they are less vulnerable to stress and have better general health (DFES, 2007). In other words, as Desforges and Abouchaar (2003) state, regardless of class or income, the influence of the parent [is] the single most significant factor in a child's life.

Moreover, there is an increasing recognition that affective factors play an important role in the teaching and learning of mathematics, in particular the presence of anxiety as a constant variable throughout the school years (Ma, 1999). For this reason, parental NIs can act as a bridge between children learning maths and feeling confident with numbers (Bempechat, 2004; DCSF, 2008)

Rationale

Mathematical knowledge has been put on an academic pedestal for many centuries. In fact, one of the key elements of any cognitive assessment is the level of performance in a series of numerical activities, as mathematical proficiency seems to be considered an index of general reasoning ability.

Research has shown that basic numerical skills taught in early years and primary school will set the foundation for more complex problem solving activities not only in secondary school but in daily life too (Aubrey, Godfrey, & Dahl, 2006).

In the UK the importance of children acquiring solid mathematical skills from an early age has become a priority along with the need to strengthen literacy skills. Only four years ago the Secretary of State claimed that mathematical understanding is critical in the development of a nation; it underpins scientific and engineering development, it is the foundation of technological breakthroughs and opens the doors of economic progress (Ofsted, 2011).

Additionally, the failure to acquire strong mathematical foundations can have serious consequences on people's lives. Simple activities such as checking the price at the checkout or helping children with their maths homework can produce a significant amount of fear and anxiety in people. People can lose confidence in their academic abilities, therefore avoiding any unnecessary contact with numbers, and as a result trusting blindly in any bills they are asked to pay or avoiding applying to a course that has maths content (Frederickson & Cline, 2009).

Thus Educational Psychologists (EPs) have the opportunity to change this by providing the necessary insight to parents and teachers to help children learn maths by:

- Introducing evidence based interventions into the life of a child to minimise the possible negative experiences noted above.
- Involving parents and giving them the opportunity to establish a culture of learning and talking about maths with their children, as well as strengthening the parent-child relationship.
- Influencing the way mathematics is taught in schools and encouraging adults to treat numeracy issues as an ‘outside the child’ problem instead of an ‘inside the child’ problem.
- Helping educators to consider the intellectual, emotional, social and cultural context in which the child encounters mathematics. For example, how maths is presented to children and the language used, as mathematical ideas are usually more complex than other subjects. This can be a particular challenge for EAL students, children with learning difficulties or children who are not exposed to a wide range of vocabulary at home (Frederickson, Miller, & Cline, 2009).

Review Question

How effective are numeracy interventions implemented by parents at helping their children increase mathematical knowledge?

Critical Review of the Evidence Base

A comprehensive literature search was carried out between 15th and 27th December, 2014, which included the following databases:

- PsycINFO
- PsycEXTRA
- PsycARTICLES
- ERIC (EBSCO)
- ERIC (PROQUEST)

The search terms are outlined in Table 1. Included sets in combination indicate that the studies chosen:

1. Included a numeracy intervention as the main independent variable, or was one of the main elements of the independent variable.
2. Described parents (or significant adults) as the ones who implemented the intervention.
3. Described the intervention being implemented at home.
4. Involved pre-secondary typically developing school children, aged between 3–11 years.

Table 1. Search Terms used in Databases

1	2	3	4
Numeracy	Parent*	Home* intervention	Children
Math*	Parent* involvement	Home based	Primary school

Math* intervention	Parent* tutoring	Family based	Pre-school
Numeracy intervention	Parent* participation	Family	Pre-kindergarten
Arithmetic		Family program	Young children
Arithmetic intervention			Early years
Math* homework			NOT secondary school
Numeracy homework			NOT high school

As a topic, NIs implemented by parents proved to be under researched yielding a limited amount of peer-reviewed results. Consequently, the search was opened to every type of study. Furthermore, due to the review question, only randomised controlled trials (RCT) with pre and post measures were included, as these types of designs are more effective in ruling out alternative explanations for any observed effect of the intervention. However, only studies conducted in English from an 'Organisation for Economic Co-operation and Development' (OECD) country were included, as their samples could be considered more similar in terms of economic development, policies and educational systems.

A relevant word search was carried out and duplicates removed. Initially every study was screened by title and then by abstract. A flow diagram of the study selection process is illustrated below. The inclusion and exclusion criteria are presented in Table 2 and the three articles which needed to be fully read in order to be excluded are presented in Appendix 3.

The 5 studies that met the inclusion criteria for critical analysis are presented in Table 3.

Figure 1. Flow diagram of Literature Search

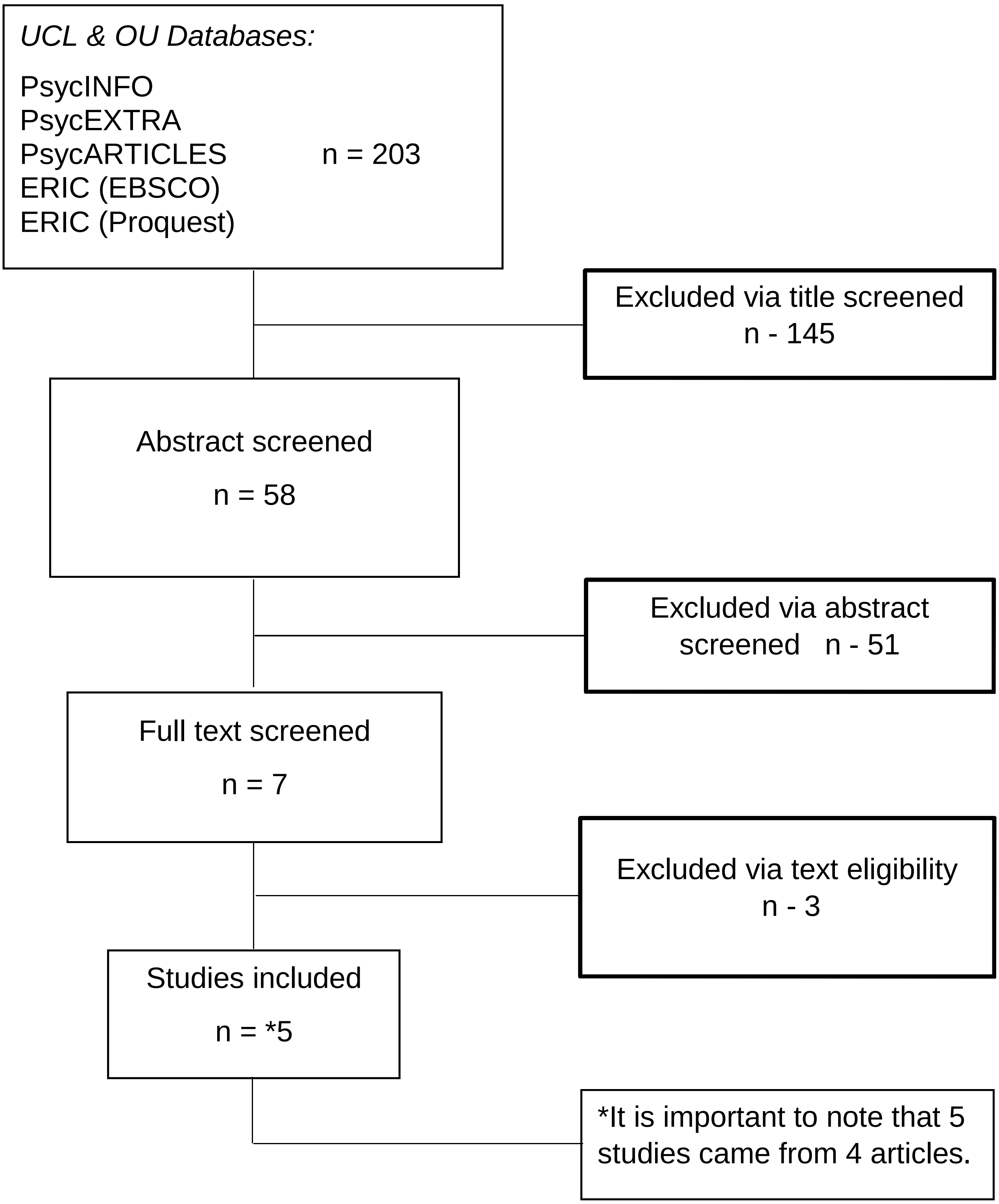


Table 2. Inclusion and Exclusion Criteria

	Inclusion Criteria	Exclusion Criteria	Rationale
Type of article	Any article, e.g. peer reviewed, non-peer reviewed, grey literature, pilot studies, etc.	Books, books' chapters, etc.	The availability of studies and possibility of finding relevant studies in unpublished work.
Type of study	Only cross-sectional studies	Longitudinal studies	Collecting data in a short period of time does not compare to collecting data across several years, as other variables need to be taken into account for the latest.
Language	Only studies in English from an OECD country e.g. UK, USA, Canada, Australia, New Zealand, etc.	Studies not written in English and/or not belonging to an OECD country.	Due to social, educational and financial similarities.
Year of publication	From 1999 to 2014	Studies carried out before 1999 or after 2014.	Due to the launch of the UK National Numeracy Strategy in 1999.
Type of data	The study must include primary empirical data.	The study does not include primary empirical data e.g. Meta-analysis or Literature Review.	This type of data ensures first-hand information.
Intervention	Studies which included a numeracy intervention, or had a clear and strong numeracy element to it.	Studies that do not include numeracy in their interventions.	Numeracy is the topic of this review.
Outcomes	The study measures any type of numeracy outcomes which matched the concept of mathematical knowledge.	Studies that look at other numeracy outcomes that don't fully meet the 'mathematical knowledge' concept previously outlined.	Mathematical knowledge is the basis for more complex arithmetic and problem solving skills in higher education (Aubrey et al., 2006)
Design	The study is a RCT with pre and post measures.	The study is not a RCT with pre and post measures.	This design is more effective in isolating the effect of the intervention (Gough, 2007).
Participants	Pre-secondary, typically developing school	Older children	Pre-secondary school mathematics sets the

	children (aged 3 to 11 years)		bases for more complex maths in the future (Aubrey et al., 2006)
	Parents or a significant adult that shares time with the child at home (e.g. carer, grandparent, etc.).	Adults who do not share time with the child at home, e.g. teachers.	Parents are considered 'children's first educators' (DCSF, 2008).
Setting	Studies that looked at home-based interventions.	Any other setting	This is the focus of this review.
Analysis	Quantitative	Qualitative	The operational definition of 'mathematical knowledge' in this review and the need for objective measures.
Outcomes	The study needs to show changes in the child or parent-child population as a unit of analysis.	The studies show changes in the parent population only.	This is the population of interest with respect to current Educational Psychology practice.

Table 3. Studies Included in this Review

1.	Ford, R. M., McDougall, S. J. P., & Evans, D. (2009). Parent-delivered compensatory education for children at risk of educational failure: Improving the academic and self-regulatory skills of a Sure Start preschool sample. *British Journal of Psychology (London, England : 1953)*, *100*, 773–797.
2.	Lavelle-Lore, M. D. (2013). *Parent-Child Home Numeracy Intervention and the Mathematics Scores of First Grade Students in Urban Catholic Schools by A Dissertation Presented to Saint Joseph ' s University Graduate Board In Partial Fulfillment of the Requirements for the Degree of Doc*.
3.	Starkey, P., & Klein, A. (2000). Early Education and Development Fostering Parental Support for Children ' s Mathematical Development : An Intervention with Head Start Families. *Early Education and Development*, *11*(5), 659–680.
4.	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.

Comparison of selected studies

As shown above, the five studies analysed came from four articles as one of them carried two studies with different samples (Starkey & Klein, 2000). All of the studies were compared on methodological features checking for the quality of the execution and its standards (WoE A). In addition, a critical appraisal looking at the appropriateness of the research design (WoE B) and how well matched the study was to the review question (WoE C) was carried out. The ‘weight of evidence’ (WoE) framework developed by Harden and Gough (2012, in Gough, Oliver, & Thomas, 2012) was used to ensure that the conclusions of this review are based on an objective, trustworthy, and appropriate analysis of the outcomes.

Table 5 provides a summary of the WoE framework and Table 6 shows the qualitative scores of each study’s WoE. The following table helps explain the scores equivalences in numerical terms; this criteria has been applied to all WoEs.

Table 4. Scores Equivalences

Evidence	Scores equivalences	Average scores
Strong	High	2.5 – 3
Promising	Medium	1.5 – 2.4
Weak	Low	1.4 or less
No/limited evidence	Zero	0

The Task Force on Evidence-Based Interventions in School of Psychology by Kratochwill (2003) Coding protocol (Group-based design) was used to assess the quality of the execution of the study, according to the generic and accepted standards associated with RCT studies.

Not all of the protocol was relevant to this review, as only primary or secondary outcomes related to mathematical knowledge were looked at. Additionally, the intervention samples were not clinical and only the quality of measures, comparison group and statistical analyses were examined. For this reason, some of the tables have been removed while others have been included across this paper. An example of a completed coding protocol can be found in Appendix 4. Further information about the WoE frameworks, their rationale and the score of each study can be found in Appendix 2.

Table 5. Framework for Weight of Evidence

Weight of Evidence A	Weight of Evidence B	Weight of Evidence C	Weight of Evidence D
Quality of execution of the study in relation to quality standards for studies of that type.	Appropriateness of research design for addressing the Review Question.	Appropriateness of the focus of the study to the Review Question.	Considering A, B & C to rate the overall degree to which the study contributes to answering the Review Question.
(Methodological Quality)	(Methodological Relevance)	(Topic Relevance)	(Overall weight of evidence)

Table 6. Weight of Evidence for each study

Studies	WoE A Methodological Quality	WoE B Methodological Relevance	WoE C Topic Relevance	WoE D Overall weight of evidence
Ford, McDougall & Evans (2009)	Low	High	Medium	Medium
Lavelle-lore (2013)	Low	Medium	Medium	Medium
a) Starkey & Klein (2010)	a) Medium	Medium	Medium	Medium
b) Starkey & Klein (2010)	b) Medium	Medium	Medium	Medium
Topping et al. (2004)	Low	Medium	High	Medium

Integrated critical review

Research Design

Appendix 1 summarises the following relevant factors: design, sample, control group, group equivalence, measures, follow up, and statistical analysis.

Design: although only RCT studies were included in this review, one study's sample was not randomly drawn from a normal population as their participants were children underachieving in maths. The pupils were then randomly assigned to either the control or intervention group. Nevertheless, the authors acknowledged this issue and carried out both parametric and non-parametric tests to avoid any biases. They found little difference between both analyses hence reporting the parametric results (Topping, Kearney, McGee, & Pugh, 2004).

Sample: the age range across all the studies was 3 to 6 years, with the exception of the Topping et al (2004) study, which focused on primary school children from 9 to 10 years.

A common feature across the studies was their *underpowered sample sizes*. Overall the number of participants ranged from 28 – 60 'units of analysis' which in some cases included only children and in others dyads of parent-child. This was judged accordingly to Cohen's (1992) suggestion that a necessary sample size for a Power of .80 to detect a medium effect size is at least 64 participants per group. A medium effect size in mathematics interventions with small samples is in line with what Kroesbergen and Van Luit (2003) found in their meta-analysis.

Across the studies, the *attrition rate* was low (20% or less) (Kratochwill, 2003) which can show parents' desire and willingness to take part in the programmes. However, given the small number of participants this can also be an indicator about the potential bias of the samples. Special care needs to be taken with the Lavelle-lore (2013) study as there was a high attrition rate, which means the outcomes were obtained from parents who were particularly interested in their children's readiness for school. This suggests the remainder of participants are not an accurate representation of the population.

Control group: only one study had an 'active control' which carried on receiving a typical intervention in the form of normal maths homework (Topping, et al.).

Group equivalence: all the studies showed some degree of equivalence between the control and intervention group. The equivalence tried to control demographics, SES and level of mathematical ability. Nevertheless, the Topping et al. (2004) study reported that the control group at pre-test 'had the same mean level but greater variability in attainment' than the intervention group (p.363) and for this reason, obtained a lower WoE B in this section.

Measures: only the Lavelle-lore (2013) study used a standardised test (GMADE^{TM}) with a high alternate form of reliability (.94). The other studies used a range of nursery tests or 'child friendly' assessments that the authors believed best fitted their purpose. Perhaps this is due to the early age of most participants as the measures needed to engage with them and assess their change in mathematical knowledge. Despite the lack of standardised tests, as many of these assessments were devised by the authors, Starkey and Klein (2000) provided a high inter coder reliability >.98 in all measured tasks. For this reason, their studies obtained slightly higher scores in the WoE C.

Follow up: none of the studies carried out a follow up assessment. This significantly reduces the possibility of generalising the durability of the findings outside the intervention period. However, one of the studies undertook a second post-test assessment after four months of the intervention, so this particular study obtained a higher score in the WoE B (Ford, McDougall, & Evans, 2009).

Application of Intervention:

Given the topic of the review and the implied homogeneity of settings (i.e. at home), it was not considered necessary to produce a summary table for these elements.

Treatment duration: research shows that NIs for pre-secondary school children seems to be more effective when it is short term. In a meta-analysis of mathematical interventions for SEN children (including low achievers and those at risk) conducted by Kroesbergen and Van Luit (2003), they found that the duration of the intervention correlated negatively with the effect size. In other words, the shorter the intervention the higher the effect size, as short interventions tend to focus on a specific domain of knowledge enabling the child to acquire a specific ability. This is supported by Williams (2008) when he suggests that a '12 week, one to one intervention [is] probably about right' for early years and primary school pupils' (p.50). However, it should be pointed out that both of these reviews did not include parental involvement and were conducted in the classroom.

On average the interventions analysed in this review lasted between 1-12 months and had an average intensity of approximately 20 minutes per day, twice a week. This is in line with the short term duration suggested by both reviews previously referred to (Kroesbergen & Van Luit, 2003; Williams, 2008).

Components of the treatment:

All studies used similar components such as:

- One-to-one instruction from parent to child.
- Activities based on mathematical games with a range of interactive and entertaining materials such as toys, cards, books, counters, etc.
- A manual to teach and remind parents of the main procedures of the intervention.

The Topping et al. (2004) study also provided a web page that included ‘more tips for tutors’ if required by the participant. This study was also the only one to report qualitative information about how much parents and children enjoyed the intervention.

Findings: Outcomes and Effect sizes (ES)

All the studies’ outcomes matched the definition of mathematical knowledge of this review. Table 8 shows a summary of the outcomes, ES and WoE D. Given the homogeneity of the outcome measures, design and the statistical analysis; a Cohen’s d (1992) score was calculated by dividing the difference in the means of each group (pre and post-test) by the pooled standard deviation (Hedge’s g).

Table 7. Cohen’s d Interpretation (1992)

Small	.2
Medium	.5
Large	.8

When assessing the WoEs as a whole, a very homogeneous picture emerges. There is a trend of low to medium scores across the studies in most aspects of the analysis.

The ES also reveal a great deal of similarity between the studies. Most of them obtained an overall medium ES which is considered appropriate for social sciences (Cohen, 1992). However, it is important to note that it was particularly difficult to conduct this analysis for two of the studies. Even though the study carried out by Ford et al. (2009) obtained high ES in half of its outcomes, all three points of data collection (pre, 1st post and 2nd post) were made with different measures. This means it is difficult to compare the efficacy of the intervention across time.

Moreover, it was not possible to continue with the analysis for the Lavelle-Lore (2013) study as the means and standard deviations were identical in too many cases (p.98). Statistically, this is almost impossible so the author was approached to confirm the information but due to the lack of response, ES could not be performed for this intervention. Furthermore, there was a disparity between the modest ES of .25 on mathematical performance reported by Topping et al. (2004) and this review's findings of an even lower ES of .09.

Table 8. Summary of Effect Sizes

Study	Outcome	Effect sizes		Overall Quality	WoE D
		Pre- Post	**Descriptor (Cohen's d)**		
*Ford, McDougall & Evans (2009)	Quantity judgements	(1st. Post test) .91	Large	Medium	Medium
	Number Recognition	.22	Small		
	Object Counting	.95	Large		
	Free Counting	.31	Small		
	Number	(2nd. Post-test) .33	Small	Medium	
	Mathematics	.83	Large		
Lavelle-lore (2013)	Concepts and Communication	Means and SD seemed incorrect			Medium
	Operations and Computation	Means and SD seemed incorrect			
	Process and Application	Means and SD seemed incorrect			
a) Starkey & Klein (2010)	Enumeration	.30	Small	Small	Medium
	Numerical Reasoning	.31	Small		
b) Starkey & Klein (2010)	Enumeration	.26	Small	Small	Medium
	Numerical Reasoning	.67	Medium		
Topping et al. (2004)	Mathematical performance	.09	Small	Small	Medium

*The outcomes were obtained with three different measures at the point of data collection hence only post test results were used in the mean difference formula.

Nevertheless there are some key themes that can be identified from the studies:

- None of the studies were peer-reviewed but instead considered 'early stage programs' according to the coding protocol (e.g. pilot studies and dissertation). This could be a reason for the lack of statistical strictness and limited methodological quality, which usually characterises published articles.
- None of the studies reported the effect size they were aiming for and as already mentioned, they were all underpowered (Cohen, 1992). However, it is important to note that the size of the effect cannot be attributed to the lack of power but to other factors such as the study design and the intervention per se.
- The lack of qualitative or quantitative information on the enjoyment/dislike of the interventions from parents and children. This information is key in 'early stage programmes' as it provides a valuable insight on how to improve the intervention, the study design and the likelihood of the parents continuing the programme at home.

Conclusions

The aim of this review was to find out how effective numeracy interventions implemented by parents were at increasing children's mathematical knowledge and consequently readiness for school. ES and outcomes suggest that these types of NIs are only moderately efficient in helping pre-secondary typically developing school children raise their mathematical knowledge. Nevertheless, this is line with other reviews carried out in the area of academic performance and parental involvement (Nye, Turner, & Schwartz, 2006).

Additionally in contrast to other type of interventions such as literacy; NIs heavily depend on the parents' knowledge and level of confidence with the subject. Therefore, it is necessary to consider the training that parents had as well as their previous experience with maths could have affected the outcomes of the intervention (Ginsburg, Rashid, & English-Clarke, 2008; O'Toole & de Abreu, 2005).

Even though a small to medium ES may be a weak statistical predictor of the efficiency of the intervention, in reality it may be just the support a child needs to not fall behind in the future. According to the Sutton Trust - Education Endowment Foundation (Teaching and Learning Toolkit: Technical Appendices, 2013) a small to medium ES can mean three to six additional months progress in attainment in a child's life. In other words, a moderate ES could be everything a pre-secondary typically developing school child needs to be mathematically ready for school, catch up with their peers and avoid a premature encounter with failure in maths and consequently the emotional distress that can bring (EEF, 2013).

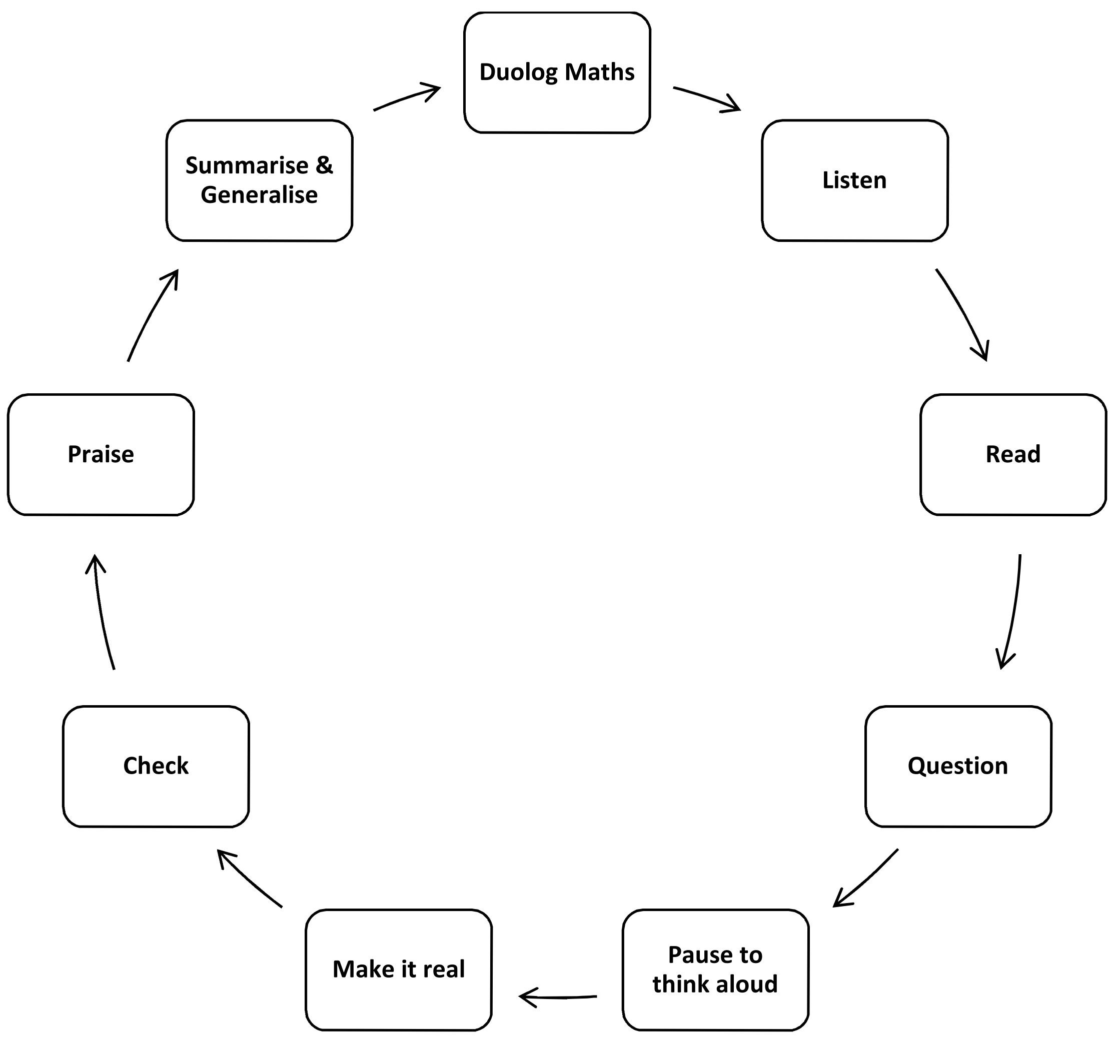
Furthermore, albeit the modest ES found, two interventions can be considered as relatively easy and inexpensive to implement across different settings. For instance, 'Let's play in Tandem' by Ford et al. (2009) made use of a range of educational toys, children's story books (which are comercially available) and paper-based activities that schools or children centres could buy/make and lend to families.

Topping et al. (2004) have also created a dynamic and well structured intervention to encourage pairs to work together, in this case parents and children. 'Duolog Maths' aims to bring 'school maths' to 'real life' scenarios. The activities came from a commercially published collection of mathematical problems (the Heinemann Mathematics 5 Extension booklet) to support the tutoring process. Despite their low

ES, this intervention is probably the most feasile to implement in different settings as it has a simple and flexible model.

Aditionally the intervention focuses on the thinking process rather than answering as being the ultimate goal. Skills which can be easily generalised and transferred to any other school subject. The following diagram illustrates the intervention's sequence (for more information go to Topping et al. (2004) p. 357):

Figure 2: Duolog Maths



Recommendations

Previous research suggest that a great amount of variance in NIs implemented by parents can be explained by two factors: parental involvement (Nye, Turner, & Schwartz, 2006) and parents' previous experience with maths (Ginsburg, et al., 2008; O'Toole & de Abreu, 2005). Consequently, it is important to improve the quality of RCTs on Numeracy Interventions conducted at home to clearly identify the effect of the intervention over other variables, thus ensuring it is the intervention causing change and not other factors. This could be done by strengthening some basic methodological principles.

Firstly, by using a Power analysis stipulating the effect size will indicate how confident the authors are in the intervention producing the expected change, in this case to improve mathematical knowledge. This will also provide an idea of how many participants the study needs to have.

Secondly, by using a bigger sample size will increase the likelihood of detecting a significant finding and consequently decreasing the likelihood of making type II error. However, finding suitable participants is a common issue in this type of research, especially with parents who need to give up their own time in committing to the intervention in addition to the other responsibilities they already have.

Thirdly, although the ES found in this review were modest, the authors reported some positive and significant outcomes immediately after the intervention program had ended. However, there is no evidence on the sustainability of these effects, so a follow up assessment could be used to reinforce these findings (Fishel & Ramirez, 2005).

Lastly, using either qualitative (e.g. a small interview) or quantitative (e.g. a 1-10 scale) methods to check how entertaining and enjoyable the intervention was. Testing ‘enjoyment’ as well as ‘improvement in mathematical knowledge’ will help make the intervention a worthwhile experience for parents and children, in particular by encouraging parents to spend quality time with their children, as research clearly states that parental involvement is the key element for a child’s success not only in school but also in life (Desforges & Abouchaar, 2003).

In addition to methodological improvements, practical elements could enhance the delivery and impact of the intervention. For instance, many parents may have had negative experiences with maths at school and this could be an obstacle in their desire to participate in any NIs with their children. For this reason, the EP could help the school set up a ‘safe space’ for parents to build up these skills, feel comfortable with their mathematical knowledge and consequently support their children better.

This could include a basic maths skills worshop, a Q&A maths session, and/or a parent-child after-school maths club where ludic activities are encouraged. This will not only increase the chance of parental involvement with the intervention but also stregthen the school-parent relationship, which is another predictor of a child’s succes in school (DFES, 2007).

Furthemore, involvement is often easier to achieve with parents of young children as they naturally gravitate to places where support is offered, such as children centres or nurseries. This could be used as an opportunity to involve parents and demonstrate how the intervention could take place at home i.e. by using maths vocabulary in everyday activities, or helping children develop problem solving skills to understand ‘real life maths’ and not just ‘school maths’ (e.g. buying sweets in the shop and knowing the value of coins). Parents could also be encouraged to explore the NI materials so they become familiar with them and are more likely to use them at home.

Finally, developing a flexible approach to NIs will allow parents to fit interventions around their schedule. For example, by breaking the sessions down into manageable activities for parents, or making use of technology to keep a record of the child’s progress, as well as providing ‘tips’ of how to implement the intervention on a website with video examples.

References:

Aubrey, C., Godfrey, R., & Dahl, S. (2006). Early mathematics development and later achievement: Further evidence. *Mathematics Education Research Journal*, *18*(1), 27–46. doi:10.1007/BF03217428

Bempechat, J. (2004). The Motivational Benefits of Homework: A Social-Cognitive Perspective. *Theory Into Practice*, *43*(3), 189–196. doi:10.1207/s15430421tip4303_4

Butterworth, B. (1999). *The Mathematical Brain* (First edition). London: Macmillan.

Cohen, J. (1992). A Power Primer. *Quantitative Methods in Psychology*, *112*(1), 155–159.

Department for Children Schools and Families (DCSF). (2008). *Impact of Parental Involvement on Children's Education*.

Department for Education and Employment (DfEE). (1999). The National Numeracy Strategy: Framework for teaching mathematics from Reception to Year 6. Retrieved from http://www.edu.dudley.gov.uk/numeracy/nns/Introduction.pdf

Department for Education and Skills (DFES). (2007). *A life's work. Every parent matters: helping you help your child*. Retrieved from http://webarchive.nationalarchives.gov.uk/20130401151715/http://www.education.gov.uk/publications/eOrderingDownload/DFES-LKDA-2007.pdf

Desforges, C., & Abouchaar, A. (2003). The Impact of Parental Involvement , Parental Support and Family Education on Pupil Achievements and Adjustment : A Literature Review. *Education*, *30*, 1–110. doi:10.1016/j.ctrv.2004.06.001

Dyson, N. I., Jordan, N. C., & Glutting, J. (2011). A Number Sense Intervention for Low-Income Kindergartners at Risk for Mathematics Difficulties. *Journal of Learning Disabilities*, *46*(2), 166–181. doi:10.1177/0022219411410233

Fishel, M., & Ramirez, L. (2005). Evidence-based parent involvement interventions with school-aged children. *School Psychology Quarterly*, *20*(4), 371–402. doi:10.1521/scpq.2005.20.4.371

Ford, R. M., McDougall, S. J. P., & Evans, D. (2009). Parent-delivered compensatory education for children at risk of educational failure: Improving the academic and self-regulatory skills of a Sure Start preschool sample. *British Journal of Psychology (London, England : 1953)*, *100*, 773–797. doi:10.1348/000712609X406762

The Sutton Trust - Education Endowment Foundation. (2013). Teaching and Learning Toolkit: Technical Appendices. Retrieved from http://educationendowmentfoundation.org.uk/uploads/pdf/Technical_Appendices_(June_2013).pdf

Frederickson, N., & Cline, T. (2009). *Special Educational Needs, Inclusion and Diversity* (2nd edition). Buckingham: OU Press.

Ginsburg, L., Rashid, H., & English-Clarke, T. (2008). Parents learning mathematics: for their children, from their children and with their children. *Adult Learning*, *19*(3/4), 21–26.

Gough, D. (2007). Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, *22*(2), 213–228. doi:10.1080/02671520701296189

Gough, D., Oliver, S., & Thomas, J. (2012). *An Introduction to Systematic Reviews* (First edition). London: Sage.

Harris, A. (2008). Do parents know they matter? Engaging all parents in learning. *Educational Research*, *5*(3), 37–41.

Hojnoski, R. L., Columba, H. L., & Polignano, J. (2014). Embedding Mathematical Dialogue in Parent–Child Shared Book Reading: A Preliminary Investigation. *Early Education and Development*, *25*(4), 469–492. doi:10.1080/10409289.2013.810481

Kratochwill, T. R., & Shernoff, E. S. (2003). Evidence-Based Practice: Promoting Evidence-Based Interventions in School Psychology. *School Psychology Quarterly*, *18*(1), 389–408. doi:10.1521/scpq.18.4.389.27000

Kroesbergen, E. H., & Van Luit, J. E. H. (2003). Mathematics Interventions for Children with Special Educational Needs: A Meta-Analysis. *Remedial and Special Education*, *24*(2), 97–114. doi:10.1177/07419325030240020501

Lefevre, J. A., Kwarchuk, S. L., Smith-Chant, B. L., Fast, L., Kamawar, D., & Bisanz, J. (2009). Home numeracy experiences and children's math performance in the early school years. *Canadian Journal of Behavioural Science*, *41*(2), 55–66. doi:10.1037/a0014532

Lavelle-Lore, M. D. (2013). *Parent-Child Home Numeracy Intervention and the Mathematics Scores of First Grade Students in Urban Catholic Schools.* (Unpublished Doctorate Dissertation). Saint Joseph ' s University, Philadephia.

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education*, *30*(5), 520–540. doi:10.2307/749772

Nye, C., Turner, H., & Schwartz, J. (2006). Approaches to parent involvement for improving the academic performance of elementary school age children. *Campbell Systematic Reviews*, *4*, 1–49. doi:10.4073/csr.2006.4

O'Toole, S., & de Abreu, G. (2005). Parents' past experiences as a mediational tool for understanding their child's current mathematical learning. *European Journal of Psychology of Education*, *20*(1), 75–90. doi:10.1007/BF03173212

Office for Standards in Education, Children's Services and Skills. (2011). Good practice in primary mathematics : evidence from 20 successful schools. *Manchester*, Ofsted.

Piaget, J. (2009). *Main Trends In Psychology.* London: George Allen and Unwin. In *Psychological Medicine*, *3*(04), 538. doi:10.1017/S0033291700054702

Sophian, C. (2004). Mathematics for the future: Developing a Head Start curriculum to support mathematics learning. *Early Childhood Research Quarterly*, *19*(1), 59–81. doi:10.1016/j.ecresq.2004.01.015

Starkey, P., & Klein, A. (2000). Early Education and Development Fostering Parental Support for Children ' s Mathematical Development : An Intervention with Head Start Families. *Early Education and Development*, *11*(5), 659–680. doi:10.1207/s15566935eed1105

Topping, K. (2000). *Tutoring*. *International Academy of Education, UNESCO* (Vol. Educationa).

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. doi:10.1080/030910042000275954

Williams, P. (2008). Independent Review of Mathematics Teaching in Early Years Settings and Primary Schools. *Department for Children, Schools and Families*, 1–83.

Appendix 1: Summary of included studies (Mapping the field)

Study	Locale	Sample	Design	Intervention	Relevant Measures	Primary Outcomes & Findings in relation to relevant measures
Ford, McDougall & Evans (2009)	UK	60 pre-schoolers from low income families. Intervention Group (IG): 30 Control Group (CG): 30	Early stage program. A pre-post test randomised block design (between-subjects) grouped according to child's age, gender, intended school, highest level of education of the primary caregiver, and whether the family had any paid employment. No significant difference was found between the CG and the IG.	A 12 months **compensatory education programme for socio-economically disadvantaged pre-schoolers** (Let's play in Tandem): activities designed to develop numerical skills. * Numeracy intervention was one of several subjects targeted. **Setting:** home **Delivery:** parent **Target outcome:** increase mathematical readiness to school	* 1st Post test: **Preschool test of academic ability (maths section).** This test was devised by the researchers. Blind data collectors.	**Accuracy on the preschool test of academic ability (maths section**). *Independent-samples t tests* indicated that the intervention group outperformed the comparison group in all domains (numerical skills: $t(58) = 3.23$; all $p < .01$ two tail).
					* 2nd Post test: **Four Counties Foundation Phase Profile test** (teachers' ratings of school readiness)	**Scores on 'Four Counties Foundation Phase Profile test'** *Independent-sample t tests* found that the intervention group received significantly higher ratings than the control group for mathematics: $t\ (58) = 2.04$, $p < .05$ two tail).
Lavelle-lore (2013)	USA	Multi-ethnic parents (and 1st grade students attending urban, Catholic schools).	Dissertation A pre-post test randomised-controlled trial: no intervention control group.	**Home numeracy intervention (HNI):** **Setting:** home **Delivery:** parent	**Group Mathematics Assessment and Diagnostic Evaluation (GMADE™): an** untimed, diagnostic	**GMADE Scores** By using 4 *ANCOVAs* with pre-test scores as covariate significant effects of the HNI

		Intervention Group (IG): 30 Control Group (CG): 30	No significant difference was found between the CG and the IG.	**Target outcome:** increase mathematics scores	mathematics test that measures strengths and weaknesses in the areas of Concepts & Communication, Operations & Computation, Process & Application (Alternate forms reliability .94) Parents were asked to fill in a small weekly survey about their children's progress.	were found between the IG & CG on the total Test score: *F*(1, 58) = 8.59, *p* = .005, partial eta squared = .13. And on the sub-tests Concepts and Communication score F (l, 58) = 3.87, *p* = .05, partial eta squared = .06 and on Operations and Computation score F (l, 58) = 8.21, p = .006, partial eta squared = .12.
a) Starkey & Klein (2010)	USA	28 mother-child (pre-kindergarten) dyads from low income families (27 African-American families and 1 Latino family) Intervention Group (IG): 14 Control Group (CG): 14	(a) Early stage program A pre-post test randomized block design (between-subjects) by gender. No significant difference was found between the CG and the IG.	**Bi-generational mathematics intervention:** the intervention group received a series of mathematics curriculum classes and a lending library of maths kits. The families attended the lessons with their pre-kindergarten children. **Setting:** home **Delivery:** parent **Target outcome:** increase mathematical readiness to school	**Individually administered mathematical tasks to the children**. The assessment included two numerical tasks (enumeration and numerical reasoning), a spatial/geometric task (spatial reference), These assessment sessions were videotaped and coded. Inter-coder reliability on all tasks (>.98)	**Development of children's informal mathematical knowledge (readiness to school**) A 2 (groups) x 2 (gender) *ANOVA* was performed. Composite scores were calculated for each child at the pre-test and post-test. Composite scores were considered to be a more sensitive measure of overall growth in mathematical knowledge than his/her score on individual tasks. There was a significant main effect of phase, *F* (1/26) =

						10.36, ***p*** < .005, with posttest scores being higher than pretest scores. The group x phase interaction was also significant, *F* (1/26) = 4.35, ***p*** < .05. The analysis revealed a significant effect of phase for the intervention group, *F* (1/26) = 13.13, p < .002, but not for the comparison group, *F* (1/26) = 0.69, N.S.
b) Starkey & Klein (2010)	USA	31 mother-child (pre-kindergarten) dyads from low income families (all Latino families) Intervention Group (IG): 14 Control Group (CG): 14	(a) Early stage program A pre-post test randomized block design (between-subjects) by gender. No significant difference was found between the CG and the IG.	**Bi-generational mathematics intervention:** the intervention group received a series of mathematics curriculum classes and a lending library of maths kits. The families attended the lessons with their pre-kindergarten children. **Setting:** home **Delivery:** parent **Target outcome:** increase mathematical readiness to school	**Individually administered mathematical tasks to the children.** The assessment included two numerical tasks (enumeration and numerical reasoning), a geometric reasoning task (early mathematical development), These assessment sessions were videotaped and coded. Inter-coder reliability on all tasks (>.98)	**Mathematical tasks scores:** A 2 (group) x 2 (phase: pretest vs. posttest) *ANOVA* was conducted on children's composite mathematics scores before and after the intervention. The analysis revealed a significant main effect of phase, ***F*** (1/29) = 55.09, ***p*** < .001, with posttest scores being higher than pretest scores. The group x phase interaction was also significant, ***F*** (1/29) = 6.43, ***p*** < .02.

						However, intervention children gave more correct answers at the post-test than at the pre-test on the enumeration task, ***t*** (15) = 4.75, ***p*** < .001, one-tailed, on the numerical reasoning task, t (15) = 3.37, ***p*** < *.005,* and on the geometric reasoning task, ***t*** (15) = 2.40, ***p*** < .02. Comparison children improved significantly on the enumeration task, ***t*** (14) = 3.37, ***p*** <.01, but not on any other tasks, ***ps*** > .l0.
Topping et al. (2004)	UK	30 children (9 – 10 years) of below mathematical ability. Intervention Group (IG): 17 Control Group (CG): 13	Pilot study A pre-post test randomised block design (between-subjects). Initially chosen by their teachers on the basis of below average mathematical ability and likelihood of benefiting from the intervention. IG were tutored in mathematical problem solving at home using the method, while the CG received traditional maths problem homework.	**Duolog Maths:** a generic tutoring (pairs) procedure applicable to any mathematical task, but with particular relevance to real-life problem solving. The activities are drawn from the mainstream school maths curriculum. **Setting:** home **Delivery:** parent **Target outcome:** mathematical abilities and attitudes towards mathematics.	1. **Criterion-referenced mathematics test in parallel forms.** (reliability no provided) 2. **Scale of attitudes to mathematics.**	1. **Attainment**: For the difference between experimental pre-test and post-test means, a significant t-value was obtained (t=4.207, df.=16, p < 0.05, two-tailed). 2. **Attitudes towards Mathematics:** No significant differences were found for any of the groups.

		No significant difference was found between the CG and the IG.	

Appendix 2: Weight of Evidence

Weight of Evidence A – Methodological Quality

The Task Force on Evidence-Based Interventions in School of Psychology coding protocol was used to weight all the studies on their quality of methodology (Kratochwill, 2003). The scores from Measures and Comparison group were transferred from the coding protocol but the Statistical analysis criteria is described below.

Statistical Analysis

Weighting of evidence	**Description**
High	Appropriate statistical analysis includes: 1. Appropriate unit of analysis 2. Familywise error rate controlled 3. Sufficiently Large N*
Medium	Appropriate statistical analysis includes two of the following: 1. Appropriate unit of analysis 2. Familywise error rate controlled 3. Sufficiently Large N*
Low	Appropriate statistical analysis includes one of the following: 1. Appropriate unit of analysis 2. Familywise error rate controlled 3. Sufficiently Large N*
Zero	Appropriate statistical analysis did NOT include ANY of the following: 1. Appropriate unit of analysis 2. Familywise error rate controlled 3. Sufficiently Large N*

*Sufficiently large N was judged according to criteria for a 2 group Analysis of Variance (ANOVA) as given by Cohen (1992). Based on a medium effect size according to Cohen's d parameters indicating an alpha level of .05 and a sample size of 64 participants in the control and experimental group. This would result in a power level of 80% (Cohen, 1988).

Understanding the scores equivalences (they apply to all weight of evidences):

Evidence	**Scores equivalences**	**Average scores**
Strong	High	2.5 - 3
Promising	Medium	1.5 – 2.4
Weak	Low	1.4 or less
No/limited evidence	Zero	0

The table below indicates the overall weight of evidence for methodological quality (WoE A) in the 5 studies:

	Weighting Score			
Studies	**Measures**	**Comparison Group**	**Statistical Analysis**	**Overall Methodological Quality**
Ford, McDougall & Evans (2009)	1	2	2	1.6
Lavelle-lore (2013)	2	2	1	1.6
a) Starkey & Klein (2010)	2	2	1	1.6
b) Starkey & Klein (2010)	2	2	1	1.6
Topping et al. (2004)	1	3	1	1.6

Weight of Evidence B – Methodological Relevance

This refers to the appropriateness of the study design and analysis for answering the review question under analysis.

Weighting of evidence	**Description**
High	• There is an 'active' control group (receiving another type of intervention). • A follow up assessment is carried out. • Demonstrate group equivalence statistically. • There is low attrition across the whole study (20% or less) (Kratochwill, 2003). • There is a systematic and clearly structured manual of the intervention. • Adequate sample size[1].
Medium	• There is a 'no experimental group'. • More than one post-test has been done. • Demonstrate group equivalence statistically. • There is a 30% or less attrition across the whole study (Kratochwill, 2003). • The intervention has a written procedure to follow. • Sample size smaller than the suggested number[1].
Low	• There is not a follow up assessment • Group equivalence is not statistically demonstrated. • There is a 40% or more attrition across the whole study (Kratochwill, 2003).

	• There is not any written guidelines of the intervention. • * Sample size smaller than the suggested number[1].

[1]Power $p < .05$, $n = 64$ per intervention & control group (Kratochwill, 2003).

Rationale for WoE B: the criteria described on the 'high' category are used to ensure objective and trustworthy statistical outcomes, the fidelity of the program and rule out any potential biases (Gough, D., Oliver, S., & Thomas, 2012; Kratochwill & Shernoff, 2003).

The table below indicates the overall weight of evidence for methodological relevance (WoE B) in the 5 studies:

Studies	**Weighting Score**					
	Control Group	**Follow up**	**Group Equivalence**	**Attrition Rate**	**Manual**	**Sample Size**	**Overall Weight**
Ford, McDougall & Evans (2009)	2	2	3	3	3	2	2.5
Lavelle-lore (2013)	2	1	3	1	3	2	2
a) Starkey & Klein (2010)	2	1	3	3	3	2	2.3
b) Starkey & Klein (2010)	2	1	3	3	3	2	2.3
Topping et al. (2004)	3	1	2	3	3	2	2.3

Weight of Evidence C – Topic Relevance

This refers to the appropriateness of the focus of each study when answering the review question.

Weighting of evidence	Description
High	a) Numeracy is the only intervention implemented. b) Sample of children underachieving in maths. c) All the measured outcomes match the description of 'mathematical knowledge' d) The focus of the study is only on children's outcomes. e) Use of measures with strong reliability and validity (r = .70 or higher) (Kratochwill, 2003) f) Use the same measures for pre and post phases. g) Data collection is done with multi-methods. h) The intervention is considered 'fun' by the children and parents.
Medium	• Numeracy is considered an important part of the intervention implemented. • Clear rationale to show that sample is in obvious risk of underachieving in maths in the future. • Some of the measured outcomes match the description of 'mathematical knowledge' • The focus of the study is on children's and parents' outcomes. • Use of measures with medium reliability and validity (r at least .70) (Kratochwill, 2003) • Use the same measures for pre and post phases. • Data collection is done with multi-methods. • The intervention is considered 'ok' by the children and parents.
Low	• Numeracy is considered a small part or just an addition of the intervention implemented. • No evidence that the sample used could be at risk of underachieving in maths in the future. • Few of the measured outcomes match the description of 'mathematical knowledge' • The focus of the study is on children's and parents' outcomes. • Use of measures with weak reliability and validity (r at least .50) (Kratochwill, 2003) or no information is provided. • It does not use the same measures for pre and post phases • Data collection did not use multi-methods. • The children and parents did not like the intervention or no information is provided.

Rational for WoE C:

a. The intervention aims to produce change in children's mathematical knowledge instead of being a by-product of another intervention.
b. An intervention is usually done when issues appear, to consequently measure any improvement. If the sample of children already show the need for improvement then it would be easier to measure any change produce by the intervention.
c. If the study measures something similar to the description of 'mathematical knowledge' of this review then it is more accurate to identify the potential use of the intervention.
d. This is the population of interest in Educational Psychology practice.
e. If the measures used in the study are reliable and valid it reassures that the findings reflect the construct that the authors were trying to target.
f. The use of the same measure of similar forms of it assures the compatibility of the outcomes.
g. Triangulation is good practice and confirms findings.
h. If parents and children enjoy the intervention they are more likely to implemented in their daily lives.

The table below indicates the overall weight of evidence for methodological relevance (WoE B) in the 5 studies:

Studies	**Weighting Score**							
	Numeracy Only	**Underachievers**	**'Mathematical Knowledge**	**Children's Outcomes**	**Reliability & Validity**	**Same Pre-post Measures**	**Data Collection Multi-method**	**'Fun'**	**Overall Weight**
Ford, McDougall & Evans (2009)	2	2	3	3	1	1	1	1	1.8
Lavelle-lore (2013)	3	1	3	3	2	3	2	1	2.3
a) Starkey & Klein (2010)	3	2	2	2	2	3	1	1	2
b) Starkey & Klein (2010)	3	2	2	2	2	3	1	1	2
Topping et al. (2004)	3	3	3	3	1	3	3	2	2.7

Weight of Evidence D – Overall Weight of Evidence

This considers weight of evidence A, B and C to rate the overall degree to which the study contributes in answering the review question. The table below indicates the overall weight of evidence (WoE D) of the 5 studies:

Studies	WoE A Methodological Quality	WoE B Methodological Relevance	WoE C Topic Relevance	WoE D Overall weight of evidence
Ford, McDougall & Evans (2009)	1.6	2.5	1.8	1.9
Lavelle-lore (2013)	1.6	2.2	2.3	2
Starkey & Klein (2010)	1.6	2.3	2	1.9
Starkey & Klein (2010)	1.6	2.3	2	1.9
Topping et al. (2004)	1.6	2.3	2.7	2.2

Appendix 3: Table of excluded studies

	Excluded Studies	Reason for Exclusion
1.	Dyson, N. I., Jordan, N. C., & Glutting, J. (2011). A Number Sense Intervention for Low-Income Kindergartners at Risk for Mathematics Difficulties. *Journal of Learning Disabilities*, *46*(2), 166–181. doi:10.1177/0022219411410233	Criteria of participants: the intervention is not delivered by parents but by instructors.
2.	Lefevre, J. A., Kwarchuk, S. L., Smith-Chant, B. L., Fast, L., Kamawar, D., & Bisanz, J. (2009). Home numeracy experiences and children's math performance in the early school years. *Canadian Journal of Behavioural Science*, *41*(2), 55–66. doi:10.1037/a0014532	Criteria of design: the study is not a RCT but a correlational study.
3.	Sophian, C. (2004). Mathematics for the future: Developing a Head Start curriculum to support mathematics learning. *Early Childhood Research Quarterly*, *19*(1), 59–81. doi:10.1016/j.ecresq.2004.01.015	Criteria of participants: the intervention is not delivered by parents but by teachers.

Appendix 4: Example of a Coding Protocol

Coding Protocol: Group Based Design

Domain:

School-and community-based intervention programs for social and behavioural problems	
Academic intervention programmes	
Family and parent intervention programmes	✓
School-wide and classroom-based programmes	
Comprehensive and coordinated school health services	

Name of Coder: V.T **Date:** 29.01.15

Full name of Study in APA format: Ford, R. M., McDougall, S. J. P., & Evans, D. (2009). Parent-delivered compensatory education for children at risk of educational failure: Improving the academic and self-regulatory skills of a Sure Start preschool sample. British Journal of Psychology (London, England : 1953), 100(4), 773–97. doi:10.1348/000712609X406762.

Intervention Name (description from study): Compensatory education programme for socio-economically disadvantaged preschoolers (Let's play in Tandem).

Study ID Number (Unique Identifier): 3

Type of Publication: (Check one)

Book/Monograph	
Journal article	✓
Book Chapter	
Other (specify):	

A. General Characteristics

A1. Random assignment designs

A1.1 Completely randomised design	
A1.2 Randomised block design (between participant variation)	✓
A1.3 Randomised block design (within-subjects variation)	
A1.4 Randomised hierarchical design	

A2. Nonrandomised designs

A2.1 Nonrandomised design	
A2.2 Nonrandomised block design (between participant variation)	
A2.3 Nonrandomised block design (within-subjects variation)	
A2.4 Nonrandomised hierarchical design	

A3. Overall confidence of judgement on how participants were assigned

A3.1 Very low (little basis)	
A3.2 Low (guess)	
A3.3 Moderate (weak inference)	
A3.4 High (strong inference)	
A3.5 Very high (explicitly stated)	✓
A3.6 N/A	
A3.7 Unknown/unable to code	

B. Statistical Treatment/ Data Analysis

	Yes	No	N/A
B1. Appropriate unit of analysis	✓		
B2. Familywise error rate controlled		✓	
B3. Sufficiently large *N*		✓	

B.3.1 Statistical Test: T-test

B3. P Level: 0.5

ES: Medium (.05)

N required: 64 per group

B4. Total size of sample (start of the study): 60

B5. Intervention group sample size: 30

B6. Control group sample size: 30

C. Type of Programme

C.1 Universal prevention programme	
C.2 Selective prevention programme	✓
C.3 Targeted prevention programme	
C.4 Intervention/Treatment	
C.5 Unknown	

D. Stage of the Programme

D.1 Model/demonstration programmes	
D.2 Early stage programmes	✓
D.3 Established/institutionalised programmes	
D.4 Unknown	

E. Concurrent or Historical Intervention Exposure

E.1 Current exposure	
E.2 Prior Exposure	
E.3 Unknown	✓

II. Key Features for Coding Studies and Rating Levels of Evidence/Support

(3 = Strong Evidence; 2 = Promising Evidence; 1 = Weak Evidence; 0 = No Evidence)

A. Measurement

A1. Use of outcome measure produces reliable sources for the majority of primary outcomes.

A1.1Yes	
A1.2 No	
A1.3 Unknown/Unable to code	✓

A2. Multi-method

A2.1 Yes	
A2.2 No	✓
A2.3 N/A	
A2.4 Unknown/Unable to code	

A3. Multi-source

A3.1 Yes	
A3.2 No	✓
A3.3 N/A	
A3.4 Unknown/Unable to code	

A4. Validity of measures reported

A4.1 Yes validated with specific target group	
A4.2 In part, validated for general population only	
A4.3 No	✓
A4.4 Unknown/Unable to code	

Rating for Measurement

3	
2	
1	✓
0	

B. Comparison Group

B1. Type of Comparison Group

B1.1 Typical contact	
B1.2 Typical contact (other) specify:	
B1.3 Attention placebo	
B1.4 Intervention elements placebo	
B1.5 Alternative intervention	
B1.6 Pharmacotherapy	
B1.7 No intervention	✓
B1.8 Wait list/ delayed intervention	
B1.9 Minimal contact	
B1.10 Unable to identify comparison group	

Rating for Comparison Group

3	
2	✓
1	
0	

B2. Overall confidence rating in judgement of type of comparison group

B2.1 Very low (little basis)	
B2.2 Low (guess)	
B2.3 Moderate (weak inference)	
B2.4 High (strong inference)	
B2.5 Very high (explicitly stated)	✓
B2.6 N/A	
B2.7 Unknown/unable to code	

B3. Counterbalancing of Change Agents

B3.1 By change agent	
B3.2 Statistical	
B3.3 Other	✓

There was no mention of counterbalancing change-agents.

B4. Group Equivalence Established

B4.1 Random assignment	
B4.2 Post hoc matched set	✓
B4.3 Statistical matching	
B4.4 Post hoc test for group equivalence	

B5. Equivalent Mortality

B5.1 Low attrition (less than 20% for Post)	✓
B5.2 Low attrition (less than 30% for follow-up)	
B5.3 Intent to intervene analysis carried out Findings: No significant group differences	

C. Primary/Secondary Outcomes are Statistically Significant

C.1 Evidence of appropriate statistical analysis for primary outcomes

C1.1 Appropriate unit of analysis	✓
C1.2 Familywise/experimenterwise error rate controlled when applicable	
C1.3 Sufficiently large *N*	

There wasn't evidence for any of this.

C2. Percentage of primary outcomes that are statistically significant

C2.1 Significant primary outcomes for at least 75% of the total primary measures for each key construct	✓
C2.2 Significant primary outcomes for between 50% and 74% of the total primary measures for each key construct	
C2.3 Significant primary outcomes for between 25% and 49% of the total primary measures for each key construct	

Rating for Primary Outcomes Statistically Significant

3	
2	✓
1	
0	

C3. Evidence of appropriate statistical analysis for primary outcomes

	Yes	No
C3.1 Main effect analyses conducted	✓	
C3.2 Moderator effect analyses conducted specify results:		✓
C3.3 Mediator analyses conducted, specify results:		✓

C. Primary Outcomes Statistically Significant (only list $p \leq .05$)

Outcomes	Primary	Who changed?	What changed?	Source	Treatment Information	Outcome Measure Used	Reliability	ES
Outcome # 1 Accuracy on the preschool test of academic ability (maths)	Primary	Children	Knowledge	Test	Let's play in Tandem	Test devised by the researchers	Unknown	Quantity judgements: .91 Number Recognition: .22 Object Counting: .95 Free Counting: .31
Scores on 'Four Counties Foundation Phase Profile test'	Primary	Children	Knowledge	Standardised Test	Let's play in Tandem	Teachers' ratings of school readiness	No provided	Number: .33 Mathematics: .83

Type of Data Effect Size is Based On	Type of Data Effect Size is Based On
Means and SDs ☑ *t* - value or *F* – value ☑	Highly estimated (e.g., only have *N* p value) ☑ Moderate estimation (e.g., have complex but complete statistics)

Chi-square (*df* = 1) Frequencies or proportions (dichotomous) Frequencies or proportions (polytomous) Other (specify): Unknown	Some estimation (e.g., unconventional statistics that require conversion) Slight estimation (e.g., use significance testing statistics rather than descriptives) No estimation (e.g., all descriptive data is present)

D. Identifiable Components

D1. Evidence for primary outcomes:

3	
2	
1	✓
0	

	Yes	No
D2. Design allows for analysis of identifiable components		✓

E3. Total number of components: 3 components: parental intervention, child and curriculum at the school.

E4. Number of components linked to primary outcomes: 3

<u>Additional criteria to code descriptively:</u>

	Yes	No
E5. Clear documentation of essential components		✓
E6. Procedures for adapting the intervention are described in detail		✓
E7. Contextual features of the intervention are documented	✓	

Rating of Identifiable Components

3	
2	
1	✓
0	

F. Implementation Fidelity

F1. Evidence of Acceptable Adherence

F1.1 Ongoing supervision/consultation	✓
F1.2 Coding intervention sessions/lessons or procedures	
F1.3 Audio/video tape implementation	

F1.3.1 Entire intervention	
F1.3.2 Part of intervention	

F2. Manualization

F2.1 Written material involving a detailed account of the exact procedures and the sequence in which they are to be used	✓

F2.2 Formal training session that includes a detailed account of the exact procedures and the sequence in which they are to be used	
F2.3 Written material involving an overview of broad principles and a description of the intervention phases	✓
F2.4 Formal or informal training session involving an overview of broad principles and a description of the intervention phases	✓

	Yes	No	Unknown
F3. Adaptation procedures are specified		✓	

Rating for Implementation Fidelity

3	
2	✓
1	
0	

G. **Replication** (answer G1, G2, G3, and G4)

G1. Same Intervention	
G2. Same Target Problem	
G3. Independent evaluation	

Rating for Replication

3	
2	
1	
0	✓

H. Site of Implementation

H1. School (if school is the site, select one of the following options)

H1.1 Public	
H1.2 Private	
H1.3 Charter	
H1.4 University Affiliated	
H1.5 Alternative	
H1.6 Not specified/unknown	

H2. Non School Site (if it is a non-school site, select one of the following options)

H2.1 Home	✓
H2.2 University Clinic	
H2.3 Summer Program	
H2.4 Outpatient Hospital	

H2.5 Partial inpatient/day Intervention Program	
H2.6 Inpatient Hospital	
H2.7 Private Practice	
H2.8 Mental Health Center	
H2.9 Residential Treatment Facility	
H2.10 Other (specify):______________________________	
H2.11 Unknown/insufficient information provided	

Rating for Site of Implementation

3	
2	✓
1	
0	

III. Other Descriptive or Supplemental Criteria to Consider

A. External Validity Indicators

	Yes	No
A1. Sampling procedures described in detail	✓	

Specify rationale for selection: Unknown

Specify rationale for sample size: Unknown

	Yes	No
A1.1 Inclusion/exclusion criteria specified		✓
A1.2 Inclusion/exclusion criteria similar to school practice		✓
A1.3 Specified criteria related to concern		✓

A2. Participant Characteristics Specified for Treatment and Control Group (see Appendix 1).

A3. Details are provided regarding variables that:

	Yes	No
A3.1 Have differential relevance for intended outcomes *Specify*: Low SES & no developmental delay according to SGS II	✓	
A3.2 Have relevance to the inclusion criteria *Specify:* Low SES & no developmental delay according to SGS II	✓	

A4. Receptivity/acceptance by target participant population

No data reported for receptivity/acceptance

Participants from Treatment Group	Results (What person reported to have gained from participation in programme)	General Rating
☑ Child/Student ☐ Parent/caregiver ☐ Teacher ☐ School ☐ Other	No evidence	☐ Participants reported benefiting overall from the intervention ☐ Participants reported not benefiting overall from the intervention

A5. Generalisation of Effects:

A5.1 Generalisation over time

	Yes	No
A5.1.1 Evidence is provided regarding the sustainability of outcomes after intervention is terminated. *Specify:*		✓
A5.1.2 Procedures for maintaining outcomes are specified		✓

A5.2 Generalisation across settings

	Yes	No
A5.2.1 Evidence is provided regarding the extent to which outcomes are maintained in contexts that are different from the intervention context. *Specify:* as different measures used across the whole study		✓
A5.2.2 Documentation of the efforts to ensure application of intervention to other settings. *Specify*:		✓
A5.2.3 Impact on implementers or context is sustained. *Specify*:		✓

Unable to code A5.2.3 because there is insufficient information provided.

A5.3 Generalisation across persons

	Yes	No
A5.3.1 Evidence is provided regarding the degree to which outcomes are manifested with participants who are different than the original group of participants for with the intervention was evaluated. *Specify*: At home & school	✓	

B. Length of Intervention

B1. Unknown/insufficient information provided	
B2. Information provided (if information provided, specify one of the following):	✓

B2.1. Weeks:

B2.2 Months: 12

B2.3 Years: __

B2.4 Other: __

C. Intensity/dosage of Intervention

C1. Unknown/insufficient information provided	
C2. Information provided (if information provided, specify one of the following):	✓

C2.1. Length of intervention session: <u>20 minutes</u>

C2.2 Frequency of intervention sessions: <u>1 time a week</u>

D. Dosage Response

D1. Unknown/insufficient information provided	✓
D2. Information provided (if information provided, answer D2.1):	

D2.1 Describe positive outcomes associated with higher dosage:

Programme Implementer

E1. Research Staff	
E2. School Speciality Staff	
E3. Teachers	
E4. Educational Assistants	
E5. Parents	✓
E6. College Students	
E7. Peers	
E8. Others	
E9. Unknown/insufficient information provided	

E. Characteristics of the Intervener

F1. Highly similar to target population on key variables (e.g., race, gender, SES)	✓
F2. Somewhat similar to target participants on key variables	
F3. Different from target participants on key variables	

F. Intervention Style or Orientation

G1. Behavioural	
G2. Cognitive-behavioural	
G3. Experimental	
G4. Humanistic/Interpersonal	
G5. Psychodynamic/insight oriented	
G6. Other, *specify:*	
G7. Unknown/insufficient information provided	✓

G. Cost Analysis Data

H1. Unknown/insufficient information provided	✓
H2. Information provided (if information provided, answer H2.1):	

H2.1 Estimated Cost of Implementation:______________________

H. Training and Support Resources

I1. Simple orientation given to change agents	✓
I2. Training workshops conducted	

# of Workshops provided:

Average length of training:

Who conducted training:

I2.1 Project Director	
I2.2 Graduate/project assistants	✓
12.3 Other, *specify:*	
12.4 Unknown	

I3. Ongoing technical support	✓
I4. Programme materials obtained	✓
I5. Special facilities	
I6. Other, *specify:*	

One way mirror and bug in the ear receiver required for this intervention.

I. Feasibility

J1. Level of difficulty in training intervention agents

J1.1 High	
J1.2 Moderate	
J1.3 Low	✓
J1.4 Unknown	

J2. Cost to train intervention agents (specify if known): unknown

J3. Rating cost to train intervention agents:

J3.1 High	
J3.2 Moderate	
J3.3 Low	
J3.4 Unknown	✓

Summary of Evidence for Group-Based Design Studies

Indicator	Overall Evidence Rating NNR = No numerical rating or 0-3	Description of Evidence (Strong, Promising, Weak, No/limited evidence, or Descriptive Ratings
General Characteristics		
General Design Characteristics	NNR	NNR
Statistical Treatment	NNR	NNR
Type of Programme	NNR	NNR
Stage of Programme	NNR	NNR
Concurrent/Historical Intervention Exposure	NNR	NNR
Key Features		
Measurement	1	Weak
Comparison Group	2	Promising
Primary Outcomes are Statistically Significant	2	Promising
Educational/Clinical Significance	NNR	NNR
Identifiable Components	1	Weak
Implementation Fidelity	2	Promising
Replication	0	No evidence
Site of Implementation	2	Promising
Follow Up Assessment Conducted	1	Weak
Descriptive or Supplemental Criteria		
External Validity Indicators	NNR	NNR
Length of Intervention	NNR	NNR
Intensity/Dosage	NNR	NNR
Dosage Response	NNR	NNR
Programme Implementer	NNR	NNR
Characteristics of the Intervener	NNR	NNR
Intervention Style/Orientation	NNR	NNR
Cost Analysis Data Provided	NNR	NNR
Training and Support Resources	NNR	NNR
Feasibility	NNR	NNR