The Sutton Trust-Education Endowment Foundation Teaching and Learning Toolkit: Technical Appendices

Professor Steve Higgins

Maria Katsipataki

(School of Education, Durham University)

Dr Dimitra Kokotsaki

Professor Robert Coe

(CEM Centre, Durham University)

Dr Lee Elliot Major (The Sutton Trust) and Robbie Coleman (Education Endowment Foundation)

June 2013

Contents

Acknowledgements

The first version of the Toolkit was originally commissioned by the Sutton Trust and produced as the 'Pupil Premium Toolkit' by Durham University in May 2011. The Sutton Trust-EEF Teaching and Learning Toolkit has been developed from this initial analysis, since the Education Endowment Foundation's launch in 2011.

The Toolkit is written by Professor Steve Higgins, Maria Katsipataki (School of Education, Durham University), Dr Dimitra Kokotsaki, Professor Rob Coe (CEM Centre, Durham University), Dr Lee Elliot Major (The Sutton Trust). and Robbie Coleman (Education Endowment Foundation). The authors also thank Isabella McDonald, Laura Evans and Sarah Whiteway for their help producing the January 2013 update.

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

Overview

The aim of these appendices is to set out some of the methods and assumptions used in the comparative synthesis of effect sizes in the Sutton Trust-EEF Teaching and Learning Toolkit. The primary aim of the Toolkit is to provide schools with evidence from education research which will help them to make informed decisions about spending to support the learning of disadvantaged pupils. Our emphasis is on identifying comparative messages from existing research. In summarising each particular field a number of judgements have had to be made about the applicability of the research evidence to the challenge of supporting learners from disadvantaged backgrounds in schools. This appendix sets out the rationale and sources of evidence for these decisions.

We believe that educational research can help schools get the maximum “educational bang for their buck", both in terms of making an initial choice between possible strategies, and in implementing a strategy as effectively as possible. However there are, of course, some limitations and caveats to the meta-analytic approach we have taken and these are also set out here. The quality of the evidence within any area is variable and one of the issues in meta-analysis is that some of the subtleties of these issues are lost in aggregation. There is also considerable variation in each of the themes that have been summarised for the Toolkit. There are examples within each area where interventions have been successful in improving attainment and others that have been unsuccessful. The most successful approaches on average have had their failures and the least successful their triumphs. What we are saying is that the existing evidence so far suggests provides information and insight which we believe is useful to schools as they make decisions about their spending and teaching priorities. What we are not saying is that approaches which are unsuccessful on average can *never* work or that approaches like feedback and metacognitive approaches will *always* work in a new context, with different pupils, a different curriculum and undertaken by different teachers.

Overall we think that the messages in the Toolkit are encouraging for teachers. The evidence summarised in the Toolkit shows that they can make a difference and that they are the most important people in the education system who are able make that difference to children and young people’s learning. However, we think that the evidence indicates that that the challenge is to get the *pupils* to work harder, not the teachers. Learners need to engage in activities which make them think harder, more deeply and more frequently. They also need to learn what is expected in different subjects and to develop strategies to help them when they get stuck. Above all they should believe they should succeed through effort and that they should be able to seek and respond to feedback to improve their learning.

We should also make it clear that we do not believe that there are any guarantees from the evidence. Teachers and schools will need to try out these ideas and evaluate their usefulness in improving learning. Sometimes this needs perseverance or effort to create the conditions in which learners can respond to feedback or take more responsibility for their learning. Another way of looking at these approaches is seeing them as means to set up a context in which learning is more or less likely to improve. The actual improvement will depend on the extent to which learners actually think harder, more deeply or more frequently about what is being learned and their teachers can support, challenge, extend and develop this thinking.

Section 1: Resources and pupil learning

It is difficult to establish a clear link between educational expenditure and pupils' learning in schools. Analysis of spending per pupil and scores on the Third International Maths and Science Study (TIMSS) found 'no association between spending levels and average academic achievement' even after controlling for variables such as family background and school characteristics' (Hanushek & Woessman, 2010). However, most of the studies have been undertaken at the system level (e.g. whole countries, states or local authorities) where the relationship between allocation of resources and differences in schools, teachers and pupils is highly complex. It may seem obvious that more money offers the possibilities for a better or higher quality educational experience, but the evidence suggests that it is not simply a question of spending more to get better results. This may be because in the UK and other developed countries we broadly spend reasonably efficiently, and increased effectiveness comes at much greater cost (Steele et al., 2007). Much of the early research in this area failed to find a convincing connection for a range of reasons (Burtless, 1996), though meta-analyses of such studies indicated there was a sufficient connection to warrant increased spending (e.g. Greenwald et al. 1998). More recent research suggests that there is a link between spending and outcomes, but that it is a complex picture (e.g. Vignoles et al., 2000) and that higher quality data sets are required to understand the mechanisms by which spending and learning are associated (Levačić & Vignoles, 2002). Some analyses suggest that the effects of greater spending tend to influence mathematics and science more than English in UK secondary schools (Steele et al., 2007) and that disadvantaged pupils may benefit more (Holmund et al. 2008; Pugh et al. 2011).

Over the period 1997-2011 per capita spending in England increased by 85% in real terms (based on projections in DCSF, 2009). During the same period improvements pupil outcomes were marginal on most international and comparative measures (e.g. Tymms, 2004; Tymms and Merrell, 2007; NFER, 2011; OECD, 2011).

Investing for better learning, or spending so as to improve learning, is therefore not easy, particularly when the specific aim is to support disadvantaged learners whose educational trajectories are harder to influence. Much depends on the context, the school, the teachers (their levels of knowledge and experience), the learners (their level of attainment and their social background) and the educational outcomes that you want to improve (knowledge, skills or dispositions). Improving test scores in arithmetic in the short term, for example, may not raise students' aspirations for what further learning in mathematics may accomplish for them. There is some evidence where interventions have been costed that spending can be used effectively to bring about measurable improvement. However these estimates vary considerably. Wiliam (2002), for example, estimated the cost of a formative assessment project with an effect size of 0.32 on pupil attainment was about £2,000 *per teacher* per year. A recent evaluation of Every Child a Reader (Tanner et al., 2011) estimated costs of £3,100 in the first year and £2,600 per year subsequently *per child* with an average reading gain of 13% (non-significant, p142) (estimated at an effect size of about 0.14: Glass, McGaw & Smith, 1981: 136).

We interpret the lack of a clear causal link between general additional spending and learning to mean that it is difficult to spend additional resource effectively to improve learning and to increase attainment, but that there must be some areas which offer better prospects than others. This is what this *Toolkit* seeks to provide. We also think that the evidence shows that if schools want to use any additional resource, such as the Pupil Premium, to benefit disadvantaged learners they should not assume that any increased allocation alone will improve learning, but they will need to decide specifically and deliberately how it should be spent, and then evaluate the impact of this for themselves. The existing research indicates that this is a challenging but achievable task.

Section 2: Cost effectiveness estimates

Cost estimates are based on the likely costs of adopting or implementing an approach with a class of twenty-five pupils. Where an approach does not require an additional resource, estimates are based on the cost of training or professional development which may be required to support establishing new practices. Approaches marked with £££ or less could be funded from the 2012-13 pupil premium allocation of £623 per eligible pupil. For example, at least 40 pupils receiving the Pupil Premium will be needed to employ an additional teacher in 2012-13 (assuming Main Pay Scale 3 (£25,168) or Outer London MPS1 (£25,117). This drops to 28 pupils eligible for the £900 Pupil Premium in 2013-4. If the Pupil Premium eventually increases to £1,200, this will be reduced to about 20 pupils.

In terms of cost effectiveness it may also be useful for schools to consider the kind of investment they are making. Reducing class sizes only last for as long as the funding maintains smaller classes. Technology equipment typically lasts for up to five years or so (with some maintenance costs). Developing teachers' feedback skills through professional development is potentially more valuable, as it may make a more lasting change in their effectiveness and build capacity within the school.

The scale used in the costing assumptions is as follows:

£	*Very low:* up to about £2,000 per year per class of 25 pupils, or less than £80 per pupil per year.
££	*Low:* £2,001-£5,000 per year per class of 25 pupils, or up to about £170 per pupil per year.
£££	*Moderate:* £5,001 to £18,000 per year per class of 25 pupils, or up to about £700 per pupil per year. This represents the 2012/13 Pupil Premium allocation (£623).
££££	*High:* £18,001 to £30,000 per year per class of 25 pupils, or up to £1,200 per pupil.
£££££	*Very High:* over £30,000 per year per class of 25 pupils, or over £1,200 per pupil. By 2014/5, the Pupil Premium is projected to rise to approximately £1,200 per pupil.

Other estimates, based on costs per class or per teacher are as follows:

Expenditure	Rate	Cost estimate
Teacher	£25-£30k per year (Scale point 3 England & Wales, Inner London Scale Point 3)	£27,500 per year
Teaching Assistant	£16-20k per year	£18,000 per year
Supply cover	£150-£200 per day	£175 per day
Computer	Total cost of ownership estimated at £3,000	£600 per year
CPD day course	£60-£500 per day	£200 per day
CPD programme	Training, support and cover for a 5 day programme with classroom development	£2,000 per year

Section 3: Effect size: what it is, what it means and how it is calculated

What is an effect size?
An effect size[1] is a key measure in intervention research and an important concept in the methodology of the *Toolkit*. It is basically a way of measuring the *extent* of the difference between two groups. It is easy to calculate, readily understood and can be applied to any measured outcome for groups in education or in research more broadly.

The value of using an effect size is that it quantifies the effectiveness of a particular intervention, relative to a comparison group. It allows us to move beyond the simplistic, 'Did it work (or not)?' to the far more important, 'How *well* did it work across a *range* of contexts?' It therefore supports a more scientific and rigorous approach to the accumulation of knowledge, by placing the emphasis on the most important aspect of the intervention – the size of the effect – rather than its statistical significance, which conflates the effect size and sample size. For these reasons, effect size is the most important tool in reporting and interpreting effectiveness, particularly when drawing comparisons about *relative* effectiveness of different approaches.

The basic idea is to compare groups, relative to the distribution of scores. This is the standardised mean difference between two groups. There has been some debate over the years about exactly how to calculate the effect size (see below), however in practice most of the differences in approaches are small in the majority of contexts where effect sizes are calculated using data on pupils' learning. It is important to remember that, like with many other statistics, the effect size is based on the average difference between two groups. It does not mean that all of the pupils will show the same difference.

For those concerned with statistical significance, it is still readily apparent in the confidence intervals surrounding an effect size. If the confidence interval includes zero, then the effect size would be considered not to have reached conventional statistical significance. The advantage of reporting effect size with a confidence interval is that it lets you judge the size of the effect first and then decide the meaning of conventional statistical significance. So a small study with an effect size of 0.8, but with a confidence interval which includes zero, might be more interesting educationally that a larger study with a negligible effect of 0.01, but which is statistically significant.

What does it mean?
As an example, suppose we have two classes of 25 students, one class is taught using a feedback intervention, the other is taught as normal. The classes are equivalent before the intervention. The intervention is effective with an effect size of 0.8. This means that the average person in the class receiving the feedback intervention (i.e. the one who would have been ranked 12th or 13th in their class) would now score about the same as the person ranked 6th in a control class which had not received the intervention. Visualising these two individuals provides a valuable interpretation of the difference between the two effects (see Figure 1).

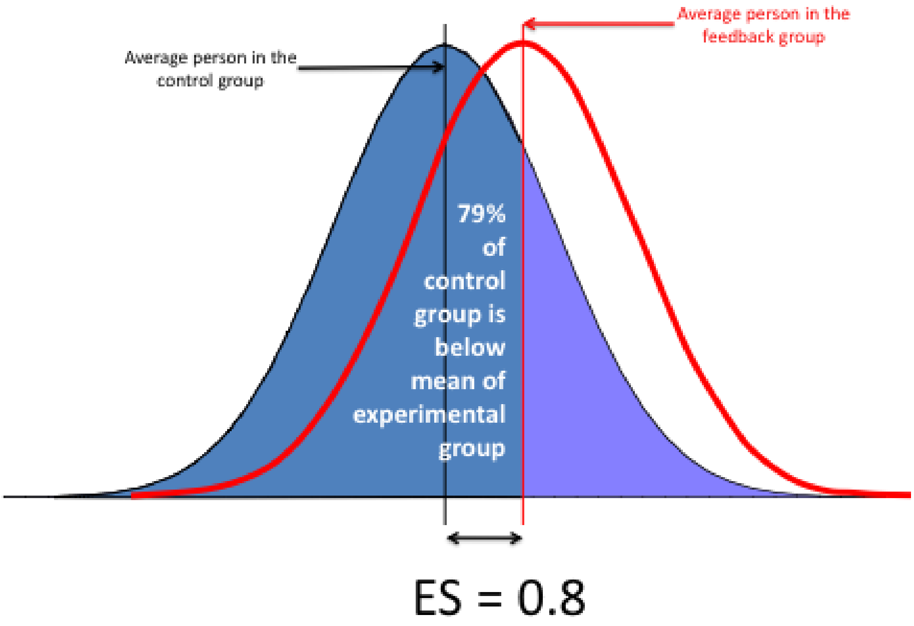
Another way to interpret effect sizes is to compare them with effect sizes of differences that are familiar. For example, Cohen (1969, p23) describes an effect size of 0.2 as 'small', and gives to illustrate the point an example that the difference between the heights of 15 year old and 16 year old girls in the US corresponds to an effect of this size. An effect size of 0.5 is described as 'medium' and is 'large enough to be visible to the naked eye'. A 0.5 effect size corresponds to the difference between the heights of 14 year old and 18 year

[1] Effect sizes can be thought of in two broad categories: first, those which compare the extent of the differences between two groups or standardised mean differences, such as Cohen's d or Hedges g; and second, variance-accounted for effect sizes, such as η^2, ω^2 or R^2) which report the extent to which overlap of key variables is explained. A third group of corrected effect sizes (Vacha-Haase and Thompson, 2002), are variations of these two, but which attempt to adjust for sampling issues. Some effect sizes can be converted mathematically into others (d to r, for example). However, it is important to bear in mind the research design from which data is analysed and the precise calculation method used in understanding the comparability of particular effect size measures (Coe, 2004). The Toolkit focuses on standardised mean difference as a measure of the impact of different interventions.

old girls. Cohen describes an effect size of 0.8 as 'grossly perceptible and therefore large' and equates it to the difference between the heights of 13 year old and 18 year old girls.

As a further example he states that the difference in IQ between holders of the PhD and 'typical college freshmen' is comparable to an effect size of 0.8.

FIGURE 1: AN EFFECT SIZE OF 0.8



Although this labelling also corresponds with the overall distribution of effects found in education research with an average around 0.4 (Sipe and Curlette, 1997; Hattie and Timperley, 2007), a 'small' effect may be educationally important if, for example, it is easy or cheap to attain or is achievable with groups who are otherwise hard to influence. Similarly a large effect size may not be as important if is unrealistic to bring about in normal circumstances. Cohen does acknowledge the danger of using terms like 'small', 'medium' and 'large' out of context. Glass and colleagues (1981, p104) are particularly critical of this approach, arguing that the effectiveness of a particular intervention can only be interpreted in relation to other interventions that seek to produce the same effect. They also point out that the practical importance of an effect depends entirely on its relative costs and benefits. In education, if it could be shown that making a small and inexpensive change would raise academic achievement by an effect size of even as little as 0.1, then this could be a very significant improvement, particularly if the improvement applied uniformly to all students, and even more so if the effect were cumulative over time.

As a standardised metric an effect size can also be converted to other measures for comparison: e.g. "students at Phoenix Park outperformed those at Amber Hill in the national school-leaving examination (the General Certificate of Secondary Education, or GCSE) by, on average, one third of a grade, equivalent to a standardized effect size of 0.21" (Wiliam et al. 2004: 50). So using this conversion, an effect size of 0.8 would be equivalent to an improvement of just over one GCSE grade.

In the Toolkit we have equated school progress in months to effect size as a crude but meaningful equivalent (see Table 1, below). We have assumed that a year of progress is about equivalent to one standard deviation per year and corresponds with Glass' observation that "the standard deviation of most achievement tests in elementary school is 1.0 grade equivalent units; hence the effect size of one year's instruction at the elementary school level is about +1" (Glass, 1981: 103). However, in is important to note that the correspondence of one standard deviation to one year's progress can vary considerably for different ages and types of test.

It is also the case that effect size difference reduces with age. Bloom and colleagues (2008) estimate annual progress on tests drops from 1.52 to 0.06 for reading and from 1.14 to 0.01 for mathematics in the US from Kindergarten to Grade 12. Wiliam (2010) estimates "apart from the earliest and latest grades, the typical annual increase in achievement is between 0.3 and 0.4 standard deviations". In the UK, data[2] from National Curriculum tests (DfES, 2004) indicates annual gains representing an effect size of about 0.8 at age 7 (at the end of Key Stage 1), falling to 0.7 at 11 (at the end of Key Stage 2) and only 0.4 at age 14 (end of Key Stage 3). One implication of this is that our estimates of improvement may underestimate the gains achievable for older pupils. If 11 year old pupils tend to make 0.7 standard deviations progress over a year, then the potential gain in terms of months estimated from meta-analytic effect sizes would increase by nearly a third. However, we think this would overestimate the gains achievable for younger children, particularly when effect sizes are re-estimated as months of possible additional progress. On the other hand, part of the reason that the same effect corresponds to more 'months gain' in older pupils is that their overall rate of gain slows down. By the end of secondary school age, the difference between the attainments of successive age groups is relatively small, especially compared with the spread within each. For these older pupils it may be a bit misleading to convert an effect size into typical months' gain: one month's gain is typically such a small amount that even quite a modest effect appears to equate to what would be gained in a long period of teaching.

TABLE 1: CONVERTING EFFECT SIZE TO MONTHS' PROGRESS

Months' progress	Effect Size from ...	... to	Description
0	-0.01	0.01	Very low or no effect
1	0.02	0.09	Low
2	0.10	0.18	Low
3	0.19	0.26	Moderate
4	0.27	0.35	Moderate
5	0.36	0.44	Moderate
6	0.45	0.52	High
7	0.53	0.61	High
8	0.62	0.69	High
9	0.70	0.78	Very high
10	0.79	0.87	Very high
11	0.88	0.95	Very high
12	0.96	>1.0	Very high

There are other reasons for preferring a more conservative estimate of what it likely to be achievable in practice. One problem is that estimates of the effects of interventions come from research studies that may

[2] http://www.education.gov.uk/rsgateway/DB/SBU/b000481/b02-2004v2.pdf, with thanks in particular to Michelle Weatherburn and Helen Evans at the Department for Education for identifying this data and providing support with the interpretation of National Test data.

optimise rather than typify their effects. For example, research is often conducted by advocates of a particular approach; considerable care is often taken to ensure that the intervention is implemented faithfully in the research setting; outcome measures used in research studies may be better aligned with the aims and focus of the intervention than other more general measures. For these reasons it may be unrealistic to expect schools to achieve the gains reported in research whose impact may be inflated (this is what Cronbach and colleagues (1980) called 'super-realisation bias'). Other evidence suggests that effect sizes will also be smaller as interventions are scaled up or rolled out (Slavin & Smith, 2008). A further problem is that part of the learning gain typically achieved in a year of schooling may be a result of maturational gains that are entirely independent of any learning experiences that are, or could be, provided by the school. For example, Luyten (e.g. 2006; Luyten et al., 2006) has shown that a substantial part (sometimes more than half) of the difference between the attainments of pupils in successive school grades is accounted for by differences in the ages of pupils who have experienced exactly the same schooling. The implication seems to be (though this is somewhat speculative) that any potential accelerating effect of using the kinds of strategies we have discussed in this report may be limited to changing the part of the year's gain that is due to schooling, while the growth that is due to pure maturation may be harder to affect. For these reasons we have selected what we see as a more conservative estimate, based on effect size estimates for younger learners, which can be improved or refined as more data becomes available about effect size transfer from research studies to practice.

Methods of calculation

Over the years there have been a number of methods proposed to calculate the appropriate standard deviation for an effect size. The main approaches are listed below.

Glass's Δ

Gene V. Glass (1977) proposed an estimator of the effect size that uses only the standard deviation of the control group, this is commonly referred to as Glass's Δ (delta). He argued that if several interventions or treatments were compared with the control group it would be better to use just the standard deviation from the control group, so that effect sizes would not differ under equal means and different variances.

Cohen's d

Cohen's *d* is defined as the difference between two means divided by an unspecified standard deviation for the data. This definition of Cohen's d is termed the 'maximum likelihood estimator' by Hedges and Olkin (1985).

Hedges's g

Hedges's *g*, suggested by Larry Hedges (1981) is based on a standardized mean difference, like the other measures, but the pooled standard deviation is computed slightly differently from Cohen's *d*.

d or g (corrected)?

Hedges's *g* is biased for small sample sizes. However, this bias can be adjusted (*g* (corrected)). Hedges and Olkin (1985) refer to this unbiased estimate as *d*, but it is not the same as Cohen's *d*. In most recent meta-analyses when an effect size is referred to as Hedges's *g* it is the bias-corrected formula which has been used, though some studies also refer to this as *d*.

Final issues

There are some notes of caution in comparing effect sizes across different kinds of interventions. Effect size as a measure assumes a normal distribution of scores. If this is not the case then an effect size might provide a misleading comparison. If the standard deviation of a sample is decreased (for example, if the sample does not contain the full range of a population) or inflated (for example, if an unreliable test is used), the effect size is affected. A smaller standard deviation will increase the effect size, a larger standard deviation will reduce it. Another key issue is which standard deviation is chosen (Bloom et al., 2008) as this primarily determines the comparability of the effect size (Coe, 2004). This explains the variation in methods advocated above.

There is also evidence that there is some systematic variation in effect sizes in education. One factor, for example, is the age of the pupils, where studies with younger learners tend to have higher effect sizes. One reason for this is likely to be the narrower distribution of scores producing a smaller standard deviation and therefore a larger effect size, though there is also a relationship with the subject (e.g. mathematics or English) being researched (Hill, Bloom & Lipsey, 2009). In England the standard deviations of National Test scores[1] increase from 3.9 at age 7, to 4.3 at age 11, and 6.8 at 14 as the distribution of scores widens and flattens (DfES, 2004).

There is also some variation associated with the type of outcome measure with larger effect sizes typically reported in mathematics and science compared with English (e.g. Higgins et al., 2005) and for researcher designed tests and teacher assessments compared with standardised tests and examinations (e.g. Hill et al., 2007: 7).

Slavin and Smith (2009) also report that there is a relationship between sample size and effect size in education research, with smaller studies tending to have larger effect sizes. The correlation found was -0.28 (p503), suggesting that is explains about 8% of the variation between large and small studies. The issue is important in terms of comparing effects between different kinds of interventions which tend to be small scale (such as areas of research looking at interventions to address special needs for example) and others which tend to have larger samples (class size interventions for example).

Other systematic factors may also affect such comparisons. Studies reporting effect sizes with groups from either end of the distribution (high attaining or low attaining learners) are likely to be affected by regression to the mean if they don't compare like with like (Shagen & Hogden, 2009). This would inflate effect sizes for low attaining pupils (who are more likely to get higher marks on re-test) and depress effect sizes for high performing students when they are compared with 'average' pupils. If the correlation between pre-test and post-test is 0.8, regression to the mean may account for as much as 20% of the variation in the difference between test and retest scores when comparing low and average students.

The aim of the *Toolkit* is not to provide definitive claims as to what *will* work to bring about improvement in a new context. Rather it is an attempt to provide the best possible estimate of what is likely to be beneficial based on existing evidence. In effect it summarises what *has worked* as a 'best bet' for what might work in the future. The applicability of this information to a new context is always likely to need active enquiry and evaluation to ensure it helps to achieve the desired effects.

Section 4: Meta-analysis and ‘super-synthesis’ of intervention research in education

Meta-analysis is a method of combining the findings of similar studies to provide a combined quantitative synthesis or overall ‘pooled estimate of effect’. The results of, say, interventions seeking to improve low attaining students’ learning in mathematics can be combined so as to identify clearer conclusions about which interventions work and what factors are associated with more effective approaches. The advantages of meta-analysis over other approaches to reviewing are that it combines or ‘pools’ estimates from a range of studies and should therefore produce more widely applicable or more generalisable results.

In addition, it can show whether the findings from similar studies vary more that would be predicted from their samples so that the causes of this variation can be investigated (moderator analysis). In education research this is particularly valuable as the results from small studies can be combined to provide answers to questions without being so dependent on the statistical significance of each of the individual studies which relates closely to sample size. Many small studies with moderate or low effects may not reach statistical significance and if you review the field by simply counting how may were statistically significant, you may be misled into thinking that the evidence is less conclusive than if you combine these studies into one combined study or meta-analysis. The statistical techniques to undertake meta-analysis form a set of transparent and replicable rules which are open to scrutiny (Aguinis et al. 2010).

Another key advantage of meta-analysis is that it helps to deal with the quantity of information in education research which can overwhelm other approaches (Chan and Arvey, 2012). This is particularly important when trying to draw relative inferences across different areas of education research. The number of studies available to review in any area of education is extensive, so techniques to aggregate and build up knowledge to propose further research and test theories and ideas are invaluable. In fields like psychology and medicine meta-analysis is relatively uncontroversial as a synthesis technique with nearly 40 years development of the principles and methods involved.

‘Super-synthesis’

It is also tempting to look at results across different kinds of studies with a common population, so to provide more general or comparative inferences. This approach is, of course, vulnerable to the classic ‘apples and oranges’ criticism which argues that you can’t really make a sensible comparison between different kinds of things. However as Gene Glass (2000) said, “Of course it mixes apples and oranges; in the study of fruit nothing else is sensible; comparing apples and oranges is the only endeavor worthy of true scientists; comparing apples to apples is trivial.”

A number of publications have attempted to take meta-analysis this stage further, by synthesising the results from a number of existing meta-analyses – producing what has been called a ‘meta-meta-analysis’ (Kazrin, Durac & Agteros, 1979), a ‘mega-analysis’ (Smith 1982), ‘super-analysis’ (Dillon, 1982) or ‘super-synthesis’ (e.g. Sipe & Curlette, 1997). However, one can make a clear separation of types within these studies. Some use the meta-analyses as the unit of analysis in order to say something about the process of conducting a meta-analysis and identifying statistical commonalities which may be of importance (e.g. Ioannidis & Trikalinos, 2007; Lipsey and Wilson, 1993). Others, however, attempt to combine different meta-analyses into a single message about a more general topic than each individual meta-analysis can achieve (e.g. Bloom, 1984; Walberg, 1984; Hattie, 1992; Sipe & Curlette, 1997). Even here, there appears to be a qualitative difference – some retain a clear focus, either by using meta-analyses as the source for identifying original studies with an overarching theoretical focus (e.g. Marzano, 1998) in effect producing something might best be considered as a series of larger meta-analyses rather than a meta-meta-analysis. Others, though, make claims about broad and quite distinct educational areas by directly combining results from identified meta-analyses (e.g. Hattie, 1992; Sipe & Curlette, 1997). In terms of the apples and oranges analogy, this is a little like asking which fruit is best for you, as a lot depends on what you mean by ‘best’ and how this is measured.

Hattie (2009) synthesized more than 800 meta-analyses and came up with some interesting findings. First of all, he concluded that most things in education 'work' as the average effect size is about 0.4. He then uses this to provide a benchmark for what works above this 'hinge' point. There are, of course, some reservations about this 'hinge' as small effects may be valuable if they are either cheap or easy to obtain, or tackle an otherwise intractable problem. Similarly, large effect sizes may be less important if they are unrealistic and if they cannot be replicated easily in classrooms by teachers. Further reservations about combining effect sizes of different kinds suggest that intervention effects should be distinguished from maturational differences or correlational effects sizes. The underlying distributions may be of different kinds, so that unlike comparing fruit, it is more like comparing an apple with a chair (Higgins & Simpson, 2011).

Although there are clearly limitations to the application of quantitative synthesis in this way, the data from meta-analysis offers the best source of information to try to answer comparative questions between different areas of educational research. It is hard to compare areas without some kind of benchmark. If you have two narrative reviews, one arguing that, say, parental involvement works and another arguing that digital technology is effective, and both cite studies with statistically significant findings showing they each improve reading comprehension, it is hard to choose between them in terms of which is likely to offer the most benefit. Meta-analysis certainly helps to identify which researched approaches have made, on average, the most difference, in terms of effect size, on tested attainment of pupils in reading comprehension or other areas of attainment. We suggest that this comparative information should be treated cautiously, but taken seriously. If effect sizes from a series of meta-analysis in one area, such as meta-cognitive interventions for example, all tend to be between 0.6 and 0.8, and all of those in another area, such as individualised instruction, are all between -0.1 and 0.2, then this is persuasive evidence that schools should investigate meta-cognitive approaches to improve learning, rather than focus on individualised instruction. Some underlying assumptions are that the research approaches are sufficiently similar (in terms of design for example), that they compared sufficiently similar samples or populations (of school pupils) with sufficiently similar kinds of interventions (undertaken in schools) and similar outcome measures (standardised tests and curriculum assessments). So, if you think that a meta-analysis of intervention research into improving reading comprehension has a set of broadly similar set of studies, on average, to a meta-analysis investigating the development of understanding in science, then you might be tempted to see if any approaches work well in both fields (such as reciprocal teaching) or, indeed, don't work well in both fields (such as individualised instruction). Our argument is that so long as you are aware of the limits of the inferences drawn, then the approach has value. We suggest that this provides the best evidence we have so far, particularly where we have no studies providing direct comparisons.

Toolkit themes

The initial themes for the Toolkit were based on expectations of how schools seemed likely to spend the Pupil Premium when it was first announced. A number of areas were specifically included at the request of teachers who have been consulted at different stages in the development of the Toolkit. Thanks in particular go to ARK and teachers from the TeachFirst Future Leaders programme, a group of Hammersmith and Ealing deputy headteachers and a number of teachers in the North-East of England who were generous with their time in attending conference or workshop presentations about earlier drafts of the Toolkit. Some of these areas (e.g. School Uniforms, Performance Pay) did not have any quantitative systematic reviews or meta-analyses to support a pooled estimate of effect. Inferences drawn from single studies or projects are limited, so these topics have a lower overall quality assessment in terms of the overall warrant from the research evidence. Feedback from schools and teachers forms an important part of the development of the Toolkit.

Search and inclusion criteria

The initial source of studies for the Toolkit was a database of meta-analyses of educational interventions developed for an ESRC Researcher Development Initiative.[3] Additionally repeated systematic searches have been undertaken for systematic reviews with quantitative data (where effect sizes are reported but not

[3] ESRC Grant RES-035-25-0037: 'Training in the Quantitative synthesis of Intervention Research Findings in Education and Social Sciences'.

pooled) and meta-analyses (where effect sizes are combined to provide a pooled estimated of effect) of intervention research in education in each of the areas of the Toolkit. These searches have been applied to a number of information gateways including Web of Knowledge, FirstSearch, JSTOR, ERIC, Google Scholar and ProQuest Dissertations. In addition a number of journals were hand searched (e.g. Review of Educational Research and Education Research Review). Journal publishers' websites offering full-text searching (Elsevier, Sage, Wiley-Blackwell) were also searched for meta-analyses. Relevant references and sources in existing super-syntheses (e.g. Sipe & Curlette, 1997; Marzano, 1998; Hattie, 2009) were identified and obtained where possible. A record of the search strategy used and studies found are kept for each of the Toolkit themes. Other studies found during the search process are also consulted in each area to provide additional contextual information.

Estimating overall impact
In each area of the Toolkit an overall estimate of the effects is identified. Where the data is available a weighted mean is used. This is based on calculating a weight for each meta-analysis according to its variance, based on the reciprocal of the square of the standard error (Borenstein et al. 2010). Where the data is not available for this an estimate is given based on the available evidence and a judgement made about the most applicable estimate to use (such as the impact on disadvantaged pupils, or the most rigorous of the available meta-analyses). Where no meta-analyses of educational interventions in a given area could be found an effect size is estimated from correlational studies or large scale studies investigating the relationship under review. If there is no information in this area, then individual studies are identified which can provide a broad estimate of effect..

Weight of evidence and quality assessment
The weight of evidence in each field was assessed according to the criteria in Table 2 below and a judgement made about how well the descriptors matched each area included in the Toolkit. These criteria are weighted to identify consistency in terms of the findings (both the overall pooled effect the pattern of effects relating to moderator variables) and to give weight to ecological validity (where studies took place in schools with interventions managed by teachers rather than researchers). The focus of the Toolkit is on providing advice to schools about how to spend additional resource to benefit disadvantaged learners, so these were judged to to be important criteria.

TABLE 2: QUALITY ASSESSMENT CRITERIA

Rating	Description
★	*Very limited:* Quantitative evidence of impact from single studies, but with effect size data reported or calculable. No systematic reviews with quantitative data or meta-analyses located.
★★	*Limited:* At least one meta-analysis or systematic review with quantitative evidence of impact on attainment or cognitive or curriculum outcome measures.
★★★	*Moderate:* Two or more rigorous meta-analyses of experimental studies of school age students with cognitive or curriculum outcome measures.
★★★★	*Extensive:* Three or more meta-analyses from well-controlled experiments mainly undertaken in schools using pupil attainment data with some exploration of causes of any identified heterogeneity.
★★★★★	*Very Extensive:* Consistent high quality evidence from at least five robust and recent meta-analyses where the majority of the included studies have good ecological validity and where the outcome measures include curriculum measures or standardised tests in school subject areas.

Section 5: Data table of meta-analyses and other studies used to estimate effect sizes

Meta-analysis	Pooled effect	ES	SE	SD	CI lower	CI upper	Min ES	Max ES	No. studies	No. Effects	Number of pupils	Mod. analysis	Pub bias
Ability grouping													
Kulik & Kulik 1982 (secondary - all)	0.10	Δ	0.05	0.32	0.01	0.19	-1.25	1.50	52	36		Yes	
Kulik & Kulik 1984 (elementary - all)	0.07	Δ	0.04						28			Yes	
Lou et al 1996 (on low attainers)	-0.12	g	-0.06		-0.01	-0.24	-1.96	1.52	103	51	16073	Yes	
Slavin 1990 b (secondary low attainers)	-0.06	Δ	-0.03	-0.12					29	15		Yes	
Indicative effect size (weighted mean)	**-0.007**												
Ability Grouping: Gifted and Talented													
Kulik & Kulik, 1987 (within class grouping)	0.62											Yes	
Kulik & Kulik, 1987 (between class grouping)	0.33											Yes	
Kulik & Kulik, 1992 (accelerated classes)	0.87	d							23			No	
Kulik & Kulik, 1992 (enriched classes)	0.41	d							25			No	
Rogers, 2007 (promotion)	1.00								32			No	
Rogers, 2007 (starting school early)	0.49								68			No	
Rogers, 2007 (ability grouping)	0.49								32			No	
Rogers, 2007 (pull-out groups)	0.65								7			No	
Rogers, 2007 (subject acceleration)	0.59								21			No	
Rogers, 2007 (G&T collaborative groups)	0.26								3			No	
Steenbergen-Hu & Moon, 2010 (all studies)	0.18	g	0.128		-0.072	0.431			28	274		Yes	No
Steenbergen-Hu & Moon, 2010 (school pupils with peers)	0.40	g	0.187		0.029	0.762			13			Yes	No
Vaughn, Feldhunsen & Asher, 1991 (pull-out)	0.65		0.19						3			No	
Indicative effect size (median)	**0.49**												
Adventure Education													
Cason & Gillis, 1994 (all)	0.31	d		0.62			-1.48	4.26	43	147	11238	Yes	
Cason & Gillis, 1994 (school grades)	0.61	d		1.527			-1.48	4.26	43	147		Yes	
Gillis & Speelman, 2008 (overall)	0.43	d					-0.24	2.83	44	390	2796	Yes	
Gillis & Speelman, 2008 (academic achievement)	0.26	d					-0.24	2.83	44	390		Yes	
Hattie et al. 1997 (all)	0.34	d	0.09						96	1728	12057	Yes	

Hattie et al. 1997 (academic outcomes)	0.45	d	0.23						96	1728		Yes	
Laidlaw, 2000	0.17			0.39			-0.43	1.38	48	389	3,550	Yes	
Indicative effect size (weighted mean)	**0.23**												
After school programmes													
Crawford, 2011	0.40	d	0.05		0.30	0.50	0.02	1.70		23		Yes	Yes
Durlak & Weissberg 2007	0.16	g	0.08	0.57	0.01	0.14	-0.16	0.67	55	66			No
Fashola 1998	*NPE*	*d*					*0.11*	*0.90*					
Lauer, Akiba & Wilkerson 2006	0.07	g	0.03		0.01	0.11			15	21		Yes	Yes
Scott-Little et al 2002	*NPE*	*d*					*0.38*	*0.50*	*23*				*No*
Tanner et al. 2011	0.14		0.11										
Zief et al. 2006 (GPA)	0.08	d			-0.03	0.20			5			No	No
Zief et al. 2006 (reading)	0.03	d			-0.10	0.16			2			No	No
Indicative effect size (median)	**0.10**												
Arts participation													
Lewis, 2004	0.20	d	0.15	0.09					5				
Newman et al. 2010 (pri/EY cognitive)	0.45	g(c)	0.09	0.19	0.28	0.62	-0.06	1.13	5	10			
Newman et al. 2010 (sec Eng)	0.05	g(c)	0.02	0.04	0.01	0.09	-0.01	0.08	1	3			
Newman et al. 2010 (sec maths)	0.03	g(c)	0.02	0.03	0.00	0.06	-0.01	0.05	1	3			
Newman et al. 2010 (sec sci)	0.06	g(c)	0.01	0.01	0.05	0.07	0.05	0.06	1	3			
Standley 2008	0.32	d	0.05	0.25	0.23	0.41	-0.23	1.70	30			Yes	Yes
Winner & Cooper 2000 (Literacy)	0.02	d	0.01	0.05			-0.25	0.66	24	24	19277	Yes	Yes
Winner & Cooper 2000 (Maths)	0.04	d	0.02	0.14			-0.14	0.34	15	15	18736	Yes	Yes
Indicative effect size (weighted mean)	**0.15**												
Aspiration interventions													
Cummings et al. 2012(parental interventions-aspirations)	0.24-0.66	d							60				
Cummings et al. 2012(mentoring-aspirations)	0.11-0.24	d							60				
Cummings et al. 2012 (extra curricular-aspirations)	0.043-0.155	d							60				
Cummings et al. 2012 (parental aspirations-attainment)	0.17-0.45	d							60				
Cummings et al. 2012 (extra curricular-attainment)	0.032-0.092	d							60				
Cummings et al. 2012 (mentoring-attainment)	0.09-0.22	d							60				
Petscher, 2010 (reading)	0.32	Zr			0.28	0.36			32	118	214615	Yes	No
Gollwitzer & Sheeran, 2006	0.65	d			0.6	0.7			63	94	8461		

Ma & Kishor, 2007 (maths)	0.12	r			0.12	0.13			113	108	82941	Yes	No
Indicative effect size	**0.00**												
Behaviour interventions													
Chitiyo et al. 2012 (students with disabilities)	0.87	r							5		25	No	
Gansle, 2005	-0.11	Q		0.28					9			Yes	
Gonzalez et al. 2004	0.49	Zr			0.43	0.55			19	7		Yes	
Quinn et al. 1999(emotional disorder)	0.05	M	0.14	0.6					17			Yes	No
Reddy et al. 2009 intervention-emotional disturbance	1.78	d							29	18	1405	No	Yes
Reddy et al. 2009 prevention-emotional disturbance	0.28	d							29	118		No	Yes
Sander et al. 2012	0.02	d			-0.18	0.22			15	134		Yes	No
Wilson & Lipsey, 2007	0.22				0.15	0.25			249			Yes	
Indicative effect size	**0.32**												
Block scheduling and timetabling													
Dickson et al. 2010 (achievement)	0.11	g (c)	0.06		-0.01	0.22	-0.14	0.48	12	7		Yes	No
Dickson et al. 2010 (maths)	-0.02	g (c)	0.07		-0.16	0.11	-0.14	0.10	12	6		Yes	No
Dickson et al. 2010 (science)	0.20	g (c)	0.07		0.06	0.33	0.13	0.42	12	4		Yes	No
Lewis et al. 2005 (maths)	-0.10	g	0.01		-0.11	-0.08	-0.15	-0.09	7	5	82463	Yes	82463
Lewis et al. 2005 (English)	-0.17	g	0.01		-0.18	-0.15	-0.25	-0.05	7	3		Yes	
Lewis et al. 2005 (science)	-0.12	g	0.01		-0.13	-0.10	-0.16	0.11	7	2		Yes	
Indicative effect size	**0.00**												
Collaborative Learning													
Igel, 2010	0.44	g			0.22	0.66	-0.08	2.45	20		2412	Yes	No
Johnson et al. 1981	0.78	d		0.99					16		70	Yes	
Johnson et al. 2000 (academic controversy)	0.86	d	0.1									Yes	
Johnson et al. 2000 (cooperative integrated read & composition)	0.18	d	0									Yes	
Johnson et al. 2000 (group investigation)	0.62	d	0.44									Yes	
Johnson et al. 2000 (jigsaw)	0.09	d	0.11									Yes	
Johnson et al. 2000 (learning together)	0.91	d	0.04						164	194		Yes	No
Johnson et al. 2000 (team assisted individualization)	0.19	d	0.04									Yes	
Johnson et al.2000 (student-team achievement)	0.28	d	0.07									Yes	
Romero, 2009	0.4	g			0.255	0.574			30			Yes	No

Indicative effect size (weighted mean)	**0.42**												
Digital technology													
Bayraktar 2001 (science)	0.27	g	0.02	0.11	0.24	0.31	-0.69	1.295	42	108		Yes	
Camnalbur & Erdoğan 2008 (in Turkey)	1.05	d	0.07	0.07	0.91	1.19			78			No	
Cheung & Slavin, 2011 (maths)	0.15	d			0.12	0.21			85		60000	Yes	No
Christmann & Badgett, 2003	0.34	d											
Li & Ma 2010 (maths)	0.28	d	0.08	0.28	0.13	0.43	-0.66	3.76	46	85			No
Liao 2005	0.55	d	0.06	0.73	0.43	0.67	-0.768	1.914	52	134			Yes
Lou, Abrami, d'Apollonia 2001	0.16	gc	0.02	0.20	0.12	0.20	-1.14	3.37	100	178			Yes
Means et al 2009	0.16	g	0.10	0.69	-0.04	0.59	-0.04	0.356	46	51			No
Moran, et al. 2008	0.49	gc	0.11	0.74	0.27	0.71	-0.204	2.679	7	7			Yes
Morphy & Graham, 2012	0.52	d*	0.10		0.33	0.71							
Onuoha 2007	0.26	d	0.05	0.28	0.17	0.35	-0.38	1.12	35	67			No
Oostdam, Otter, Overmaat 2002	0.19	d	0.06	0.40	0.08	0.30	0.13	0.25	42	50			Yes
Pearson, et al. 2005	0.49	g	0.11	0.74	0.27	0.71			30	89		Yes	
Rosen & Salomon, 2007	0.46	d	0.01	0.62	0.44	0.48	-1.152	2.003	32				Yes
Sandy-Hanson 2006 (gen. academic)	0.25	d	0.02	0.47	0.22	0.28	-1.04	1.33	31	31		Yes	
Seo & Bryant 2009	0.37	d	0.01	0.03	0.43	0.67			11				No
Sitzman et al. 2006	0.15	gc	0.02	0.17	0.11	0.19			71				
Strong, Torgerson, Torgerson, Hulme 2011	0.08	d	0.09	0.21	-0.09	0.25	-1	0.17	6	8			No
Tamim et al. 2009	0.35		0.04		0.27	0.43						Yes	
Tamim et al. 2011	0.33	d	0.04	0.20	0.25	0.41	-0.09	0.55	25	574			No
Tokpah, 2008	0.38	d	0.03	0.14	0.34	0.43	-0.47	1.23	31	102			Yes
Torgerson & Elbourne 2002	0.37	g	0.20	0.53	-0.02	0.77	-0.11	1.15	7	6			
Torgerson & Zhu 2003	0.02	g	0.19	0.38	-0.17	0.58				4		Yes	Yes
Torgerson & Zhu 2003	-0.05	g	0.14	0.29	-0.33	0.24				4		Yes	Yes
Torgerson & Zhu 2003	0.89	gc	0.33	0.47	0.25	1.54				2		Yes	Yes
Vogel et al 2006	0.07		0.01	0.06	0.05	0.09			32				
Waxman, Connell, and Gray, 2002	0.30	Δ	0.15	0.63	0.00	0.60	-0.154	0.39	20	138			Yes
Waxman, Lin, Michko 2003	0.45	Δ	0.14	0.72	0.17	0.72			42	29			
Zhao 2003	0.81	d	0.13	0.72	0.55	1.07	0.28	2.82	9	29			Yes
Indicative effect size (weighted mean)	**0.28**												

Early years intervention													
Anderson et al. 2003	0.35	*Δ*	0.18	0.62			-0.61	0.89	12	29		No	
Camilli et al. 2010	0.23								123				
Gilliam & Zigler 2001	NPE	*Δ*					0.07	0.50	13			No	No
Gorey, 2010 (esti.on long-term impact)	0.55	*U3*							35	80	18000	Yes	Yes
Karoly et al. 2005	0.28												
LaParo & Pianta 2000	0.51		0.26	2.18					70			No	
Lewis & Vosburgh 1988	0.41		0.04	0.39	0.33	0.73	0.21	0.96	65	46	3194	No	
Manning et al. 2010 (adolescent education)	0.53	d			0.40	0.68			23			Yes	Yes
Nelson et al. 2003	0.52	g	0.27	1.55			0.01	1.25	34			Yes	No
Indicative effect size (weighted mean)	**0.45**												
Extended School Time													
Cooper et al. 2003 (district level comparison) with comparison group	0.06	d			-0.02	0.02			13	39	44000	Yes	No
Cooper et al. 2003 (with matched controls)	0.11	d							13			Yes	No
Baker et al. 2004 (international comparison -maths in UK)	*0.12*	*r*											
Indicative effect size	**0.11**												
Feedback													
Bangert-Drowns et.al. 1991	0.26		0.06	0.38			-0.83	1.42	40	58			
Fuchs & Fuchs 1985	0.72	U3	0.09	0.88					21	95		Yes	
Kingston & Nash, 2011 (AfL)	0.20	Q	0.08		0.19	0.21			13	42		Yes	
Kluger & De Nisi, 1996	0.41	d	0.09	1.03	0.23	0.59			131	607	23663	Yes	Yes
Lysakowski & Walberg 1982	0.97	d	0.49	1.53			-1.09	4.99	54	94	14689	Yes	Yes
Tenenbaum & Goldring 1989	0.72		0.37	1.42					15	16			
Walberg 1982	0.81	d	0.41	1.80					19	19			
Indicative effect size (weighted mean)	**0.62**												
Homework													
Cooper, Robinson & Patal 2006	0.60	d	0.26	0.64	0.38	0.82	0.39	0.97	6	9		Yes	
Paschal, Weinsten & Walberg 1984	0.36	*Δ*	0.18	0.24			-0.60	1.96	15	81			
Indicative effect size (Primary)	**0.07**												
Indicative effect size (weighted mean - Secondary)	**0.44**												

Individualised instruction													
Aiello & Wolfe, 1980 (science)	0.35	*d*							115	182		Yes	
Bangert, Kulik & Kulik, 1983	0.10	*Δ*	0.05	0.37	0.00	0.20	-0.84	1.24	49	49		Yes	
Horak, 1981	-0.07	*Δ*	-0.04	-0.28			-1.49	0.53	60	129			
Willett, Yamashita & Anderson, 1983	0.17	*Δ*	0.09	0.41			-0.87	1.74	130	341			
Indicative effect size	**0.10**												
Learning styles													
Garlinger & Frank, 1986	-0.03	d							7	7			
Kavale & Forness, 1987	0.14	d	0.06	0.28	0.02	0.27			39		3087		
Lovelace, 2005 (Dunn & Dunn Model)	0.67	d					0.67	0.80	76	168	7196	Yes	Yes
Slemmer, 2002 (ICT context)	0.13	d	0.03		0.08	0.19			48	51		Yes	Yes
Tamir, 1985	0.02	d							54	13			
Indicative effect size	**0.10**												
Mastery Learning													
Bangert, Kulik & Kulik, 1983	0.10	*Δ*	0.053						49			Yes	No
Guskey & Piggott, 1998 (language)	0.60	gc			0.18	1.02	0.02	1.70	46	78		Yes	
Guskey & Piggott, 1998 (maths)	0.70	gc			0.50	0.90	0.02	1.70	46	78		Yes	
Kulik, Kulik & Bangert-Drowns, 1990	0.52	g	0.033						108			Yes	
Waxman et.al. 1985	0.39	*Δ*					-1.18	1.73	38	309	7200	No	No
Indicative effect size (weighted mean)	**0.40**												
Mentoring													
Bernstein et al. 2009 (maths)	-0.05	d							32		2573	Yes	
Bernstein et al. 2009 (reading)	-0.04	d							32			Yes	
Bernstein et al. 2009 (science)	-0.03	d							32			Yes	
DuBois et al. 2002 (academic)	0.11	d			-0.08	0.08			55	575		Yes	No
Wheeler, Keller & DuBois, 2010 (maths)	-0.02	d			-0.02	0.03			3			No	
Wheeler, Keller & DuBois, 2010 (reading)	-0.01	d			-0.05	0.04			3			No	
Wood & Mayo-Wilson, 2012 (academic performance)	-0.01	g			-0.11	0.08			6		4769	Yes	
Indicative effect size	**0.05**												

Meta-cognition and self-regulation strategies													
Abrami et al. 2008	0.34	gc	0.01	0.61	0.31	0.37	-1.00	2.75	117	161	20698	Yes	Yes
Chiu 1998	0.67	gc	0.34	0.68			-1.25	2.75	43	123	3475	Yes	
Dignath et al. 2008	0.62	d*	0.05		0.52	0.72	0.44	1.00	48	263		Yes	
Haller et al. 1988	0.71	d	0.36	0.81			0.25	3.80	20	115	1553	No	
Higgins et al. 2005	0.62	gc	0.09	0.38	0.45	0.80	-0.17	1.61	19	19		No	Yes
Klauer & Phye 2008	0.69	gc	0.05		0.59	0.79	0.59	0.94	73		3600	Yes	
Indicative effect size (weighted mean)	**0.62**												
One-to-one tutoring													
Cohen, Kulik & Kulik 1982 (on tutees)	0.40	*Δ*	0.07	0.50	0.26	0.54			52			Yes	Yes
Cohen, Kulik & Kulik 1982 (on tutors)	0.33	*Δ*	0.09		0.15	0.51			38			Yes	Yes
Elbaum et al. 2000	0.41	*Δ*	0.05	0.25	0.32	0.49	-1.32	3.34	29		1539	Yes	Yes
Jun, Ramirez & Cumming, 2010 (by adults)	0.70	*d*			0.48	0.93			12			Yes	No
Ritter et al. 2009	0.30	*g*	0.06	0.32	0.18	0.42	0.26	0.45	21		1676	Yes	No
Slavin et al. 2011 (1-1 phonics tutoring)	0.62	*d*							10			Yes	
Tanner et al. 2011	0.14	*d*											
Wasik & Slavin 1993	NPE	*Δ*					0.20	1.16	16				
Indicative effect size (weighted mean)	**0.44**												
Parental involvement													
Bus et al. 1995 (joint book reading)	0.59	d							16	33	3410	Yes	Yes
Jeynes 2005	0.27	gc	0.14	0.57			0.00	1.78	41	17	20000	Yes	
Jeynes 2007	0.25	gc	0.07		0.11	0.39	0.01	0.83	52	20	30000	Yes	
Layzer et al. 2001 (across school age)	0.27	d							562	11.112		Yes	No
Layzer et al. 2001 (preschool)	0.37	d							562	11.112		Yes	No
Nye, Turner & Schwartz, 2006	0.43	g			0.30	0.56			19			Yes	No
Senechal & Young, 2008 (family literacy)	0.65	d			0.53	0.76			16	1.340		Yes	
Van-Steensel et al. 2011 (family literacy)	0.18	d	0.06						30	152		Yes	No
Indicative effect size (weighted mean)	**0.26**												
Peer tutoring/ peer-assisted learning													
Cohen, Kulik & Kulik 1982 (on tutees)	0.40	*Δ*	0.07	0.50	0.26	0.54	-1.00	2.30	52			Yes	Yes
Cohen, Kulik & Kulik 1982 (on tutors)	0.33	*Δ*	0.09						11			Yes	Yes
Ginsburg-Block et al. 2006	0.48	g	0.24	0.39			0.38	0.78	36	36		Yes	

Jun, Ramirez & Cumming, 2010 (cross-age)	1.05	*d*			0.45	1.44			12			Yes	
Ritter et al. 2009	0.30	gc	0.06	0.32	0.18	0.42	0.26	0.45	28				Yes
Rohrbeck et al. 2003	0.59	gc	0.10	0.90	0.40	0.78	0.21	0.62	90			Yes	Yes
Indicative effect size	**0.48**												
Performance pay *No meta-analyses or systematic reviews. ES estimated from:*													
Woessman 2010 (correl)	0.25		0.13										
Martins 2009	-0.09		-0.05										
Indicative effect size	**0.00**												
Phonics													
Camilli, Vargas & Yurecko, 2003	0.24	d							40			Yes	No
Ehri et al. 2001	0.41	d	0.03		0.36	0.47	-0.47	3.71	66	65		Yes	Yes
Jeynes, 2008	0.30	gc	0.10		0.10	0.50	-1.21	2.02	22		5000	Yes	Yes
Slavin et al. 2011 (1-1 phonics tutoring)	0.62	d							10			No	
Slavin et al. 2011 (small groups)	0.35	d							22			No	
Torgerson, Brooks & Hall, 2006	0.27	d	0.09		0.10	0.45	-0.19	2.69	14			Yes	Yes
Indicative effect size (weighted mean)	**0.35**												
Physical Environment *No meta-analyses or systematic reviews to estimate ES.*													
Indicative effect size	**0.00**												
Reducing class sizes													
Goldstein, Yang, Omar, Turner & Thompson, 2000	0.20	d	0.10	0.31			-0.07	0.60	9	36			
Glass & Smith 1978	0.01	*Δ*	0.00	0.04					77	725	9000		No
McGiverin, Gilman & Tillitski 1989	0.34	d	0.13		0.09	0.59	-0.74	2.24	10	24			
Slavin 1990 (a)	0.17	*Δ*	0.09										
Indicative effect size	**0.20**												
Repeating Years													
Allen et al. 2009 (low quality studies)	-0.30	d							22	207		No	No
Allen et al. 2009 (medium & high quality studies)	0.04	d							22	207		No	No
Bright, 2011	-0.50	d					0.11	1.17	26	245		No	
Jimerson, 2001	-0.31	d							20	246			

Holmes & Matthews, 1984	-0.34	d	0.036						44	575	11132	Yes	
Yoshiba, 1989	-0.60	*Δ*		0.61			-1.98	0.75	34	242		Yes	
Indicative effect size (weighted mean)	**-0.32**												
School uniforms													
No meta-analyses or systematic reviews. ES estimated from:													
Samuels 2003 - language arts	0.03	g(c)	0.11	0.16	-0.18	0.23	-0.06	0.03	1	2	9585	No	No
Samuels 2003 - mathematics	-0.06	g(c)	0.11	0.16	-0.26	0.15	-0.06	0.03	1	2		No	No
Sowell, 2012 mathematics	0.02										1152		
Indicative effect size	**0.00**												
Small Group Tuition													
Elbaum et al. 2000 (pairs)	0.4	*Δ*			0.24	0.56			19	116		Yes	
Elbaum et al. 2000 (small group-*NB only 1 study*)	1.61	*Δ*			0.75	2.48						Yes	
Lou et al. 2001 (with ICT), individual	0.16	d			0.12	0.2			122	486		Yes	
Lou et al. 2001 (with ICT), small group	0.31	d			0.2	0.43			122	486		Yes	
Lou et al. 2001 (with ICT), pairs compared with groups of 3-5	0.08	d							122	486	11317	Yes	
Slavin et al. 2011	0.31								20				
Indicative effect size (weighted mean)	**0.34**												
Social & Emotional Learning													
Durlak et al. 2011	0.27	g			0.15	0.39			213		270034	Yes	No
Multon et al. 1991	0.26	r							36	38	4998	Yes	No
Payton et al. 2008	0.28	d			0.14	0.41			29			Yes	
Indicative effect size (weighted mean)	**0.27**												
Sports participation													
Newman et al. 2010 (academic outcomes)	0.19	gc	0.08	0.12	0.03	0.35	0.15	0.34	2	2		No	Yes
Newman et al. 2010 (mathematics)	0.80	gc	0.11	0.16	0.58	1.02	0.66	0.98	1	2		No	Yes
Lewis, 2004	0.10	d		0.13					5			Yes	No
Shulruf, 2010 GPA	0.15	d			0.07	0.23			29	15	59804	Yes	
Indicative effect size (weighted mean)	**0.18**												
Summer schools													
Lauer, Akiba & Wilkerson 2006	0.05	g	0.01	0.04	0.01	0.11			14			Yes	Yes
Cooper et al 2000	0.26	d	0.01	0.06	0.24	0.28	-0.20	2.70	30				

Lewis, 2004	0.10	d	0.13			5			Yes	Yes
Indicative effect size (weighted mean)	**0.19**									
Teaching assistants										
No meta-analyses or systematic reviews. ES estimated from:										
Gerber et al. 2001 (with regular classes)	NPE (0.0 est)	d		ns	ns			1985		
Gerber et al. 2001 (with small classes)	NPE (-.15 est)	d		-0.13	-0.26			1985		
Blatchford et al. 2009	0.00							7578		
					Total studies	**8147**	**Total pupils**	**1132463**		

KEY		
Single studies with ES reported in italics		
Types of effect size		
Control group SD	Glass	*Δ*
SD unspecified	Cohen's d	d
Pooled SD	Hedges g	g
Pooled SD corrected for small sample bias	Hedges g corrected	gc
gc is also sometimes confusingly referred to as an 'unbiased estimator' or d		d*
Values in red calculated		
Distribution overlap, percentage of scores in the lower-mean group exceeded by the average score in the high-mean group.		U3
Fisher z transformed correlation coefficient.		Zr
No pooled effect	NPE	

Section 6: Bibliography of meta-analyses and other studies used to estimate effect sizes

Abrami, P.C., Bernard, R.M., Borokhovski, E., Wade, A., Surkes, M.A., Tamim, R., & Zhang, D. (2008). Instructional Interventions Affecting Critical Thinking Skills and Dispositions: A Stage 1 Meta-Analysis. *Review of Educational Research 78*.4 pp 1102-1134.

Aiello, N.C. (1981). *A meta-analysis comparing alternative methods of individualized and traditional instruction in science.* Virginia Polytechnic Institute and State University). ProQuest Dissertations and Theses, p 142. Retrieved from http://search.proquest.com/docview/303140546 (30/6/12).

Aiello, N.C. & Wolfle, L.M. (1980). *A Meta-Analysis of Individualized Instruction in Science* Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 7-11, 1980). ERIC ED190404.

Anderson, L.M., Shinn, C., Fullilove, M.T., Scrimshaw, S.C., Fielding, J.E., Normand, J., Carande-Kulis, V.G. (2003). The Effectiveness of Early Childhood Development Programs: A Systematic Review. *American Journal of Preventative Medicine* 24: 32-46.

Baker, D. P., Fabrega, R., Galindo, C., & Mishook, J. (2004). Instructional time and national achievement: Cross-national evidence. *Prospects*, 34:3, 311-334.

Bangert, R.L., Kulik, J.A., Kulik, C.C. (1983). Individualized Systems of Instruction in Secondary Schools. *Review of Educational Research*, 53:2. pp. 143-158.

Bangert-Drowns, R.L.Chen-Lin, C., Kulik, J.A., & Kulik, C.C. (1991). The Instructional Effect of Feedback in Test-Like Events. *Review of Educational Research, 61*:2, 213-238.

Bayraktar S. (2000). A Meta-Analysis of the Effectiveness of Computer Assisted Instruction in Science Education. *Journal of Research on Technology in Education, 42*:2, 173-188.

Bernstein, L., Rappaport, C. D., Olsho, L., Hunt, D., & Levin, M. (2009). *Impact Evaluation of the US Department of Education's Student Mentoring Program. Final Report*. NCEE 2009-4047. Washington US Department of Education National Center for Education Evaluation and Regional Assistance.

Blatchford, P., Bassett, P., Brown, P., Koutsoubou, M., Martin, C., Russell, A. and Webster, R., with Rubie-Davies, C. (2009). *The impact of support staff in schools. Results from the Deployment and Impact of Support Staff project*. (Strand 2 Wave 2) (DCSF-RR148). London: Department for Children, Schools and Families.

Bus, A. G., Van Ijzendoorn, M. H., & Pellegrini, A. D. (1995). Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy. Review of Educational Research, 65:1, 1-21.

Camilli, G., Vargas, S., & Yurecko, M. (2003). Teaching Children to Read. The Fragile Link between Science and Federal Education Policy. *Educational Policy Analysis Archives, 11*:15, ISSN 1068-2341.

Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-Analysis of the effects of early education interventions on cognitive and social development. Teachers College Record, 112:3, pp. 579–620.

Camnalbur, M., & Erdogan, Y. (2010). A Meta-Analysis on the Effectiveness of Computer-Assisted Instruction: Turkey Sample. *Educational Sciences: Theory & Practice, 8*:2, 497-505.

Cheung, A.C.K. & Slavin, R.E. (2011). The Effectiveness of Education Technology for Enhancing Reading Achievement: A Meta-Analysis. *From the Center for Research and Reform in Education, John Hopkins University: http://www. bestevidence. org/reading/tech/tech. html*.

Chitiyo, M., Makweche-Chitiyo, P., Park, M., Ametepee, L.K. & Chitiyo, J. (2011). Examining the Effect of Positive Behaviour Support on Academic Achievement of Students with Disabilities. *Journal of Research in Special Educational Needs, 11*:3, 171-177.

Chiu, C.W.T. (1998). *Synthesizing Metacognitive Interventions: What Training Characteristics Can Improve Reading Performance?* Paper presented at the Annual Meeting of the American Educational Research Association San Diego, CA, April 13-17, 1998.

Christmann, E. P., & Badgett, J. L. (2003). A meta-analytic comparison of the effects of computer-assisted instruction on elementary students' academic achievement. *Information Technology in Childhood Education Annual*, *2003*(1), 91-104.

Cohen, P.A., Kulik, J.A., Kulik, C.C. (1982). Educational Outcomes of Tutoring: A Meta-Analysis of Findings. *American Educational Research Journal*, 19.2: pp. 237-248.

Cooper, H, Charlton, V., Muhlenbruck, M., Borman, G.D. (2000). Making the Most of Summer School: A Meta-Analytic and Narrative Review. *Monographs of the Society for Research in Child Development*, 65.1, pp. 1-127

Cooper, H., Valentine, J. C., Charlton, K., & Melson, A. (2003). The effects of modified school calendars on student achievement and on school and community attitudes. Review of Educational Research, 73(1), 1-52.

Cooper, H., Robinson, J.C., Patall, E.A. (2006). Does Homework Improve Academic Achievement? A Synthesis of Research, 1987-2003. *Review of Educational Research*, 76. 1 pp. 1-62.

Crawford, S.T. (2011). Meta-analysis of the impact of after-school programs on students reading and mathematics performance. University of North Texas). ProQuest Dissertations and Theses, Retrieved from http://search.proquest.com/docview/919077210 (20/6/12).

Cummings, C., Laing, K., Law, J., McLaughlin, J., Papps, I., Todd, L., et.al. (2012). *Can Changing Aspirations and Attitudes Impact on Educational Attainment? A Review of Interventions.* Joseph Rowntree Foundation Report.

Dickson K., Bird K., Newman M. & Kalra N. (2010). *Effect of Block Scheduling on Academic Achievement in Secondary Schools: A Systematic Review of Evidence.* The Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre), Institute of Education, University of London.

Dignath, C, Buettner,G and Langfeldt. H (2008). How can primary school students learn self-regulated learning strategies most effectively? A meta-analysis on self-regulation training programmes. *Educational Research Review* 3.2 pp. 101-129.

DuBois, D. L., Holloway, B. E., Valentine, J. C., & Cooper, H. (2002). Effectiveness of mentoring programs for youth: A meta-analytic review. *American Journal of Community Psychology*, 30(2), 157-197.

Durlak, J.A. & Weissberg, R.P. (2007). *The impact of after-school programs that promote personal and social skills*. Chicago, IL: Collaborative for Academic, Social, and Emotional Learning.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82(1), 405-432.

Ehri, C.L., Nunes, S.R., Stahl, S.A., & Willows, D.M. (2001). Systematic Phonics Instruction Helps Students Learn to Read: Evidence from the National Reading Panel's Meta-Analysis. *Review of Educational Research, 71*, (3) 393-447.

Elbaum, B., Vaughn, S.M., Hughes, M.T. & Moody, S.M. (2000). How Effective Are One-to-One Tutoring Programs in Reading for Elementary Students at Risk for Reading Failure? A Meta-Analysis of the Intervention Research. *Journal of Educational Psychology* 92.4 (2000): 605-619.

Elbaum, B., Vaughn, S., Hughes, M. T., Moody, S. W., & Schumm, J. S. (2000) (b). How reading outcomes of students with disabilities are related to instructional grouping formats: A meta-analytic review. In R. Gersten, E.P. Schiller & S. Vaughn (Eds) *Contemporary special education research: Syntheses of the knowledge base on critical instructional issues*, 105-135. Mahwah, NJ: Lawrence Erlbaum Associates.

Fashola. O.S. (1998). *Review Of Extended-Day And After-School Programs And Their Effectiveness Report No. 24* Baltimore, Center for Research on the Education of Students Placed At Risk (CRESPAR), Johns Hopkins University.

Fuchs, L.S. & Fuchs, D. (1985). Effects of systematic formative evaluation A meta-analysis. *Exceptional Children*, 53.3 pp 199-208.

Gansle, K.A. (2005). The Effectiveness of School-Based Anger Interventions and Programs: A Meta-Analysis*. Journal of School Psychology, 43,* 321-341.

Garlinger, D.K. & Frank, B.M. (1986). Teacher-student cognitive style and academic achievement: a review and a mini-meta analysis. *Journal of Classroom Interaction* 21.2, 2-8.

Gerber, S.B., Finn, J.D., Achilles, C.M. and Boyd-Zacharias, J. (2001). Teacher aides and students' academic achievement. *Educational Evaluation and Policy Analysis* 23.2 pp 123-143.

Gilliam, W. & Zigler, E.F. (2000). A critical meta-analysis of all evaluations of state-funded preschool from 1977 to 1998: implications for policy, service delivery and program evaluation. *Early Childhood Research Quarterly*, 15(4), 441-473.

Ginsburg-Block, M.D., Rohrbeck, C.A., & Fantuzoo, J.W. (2006). A Meta-Analytic Review of Social, Self-Concept and Behavioral Outcomes of Peer-Assisted Learning. *Journal of Educational Psychology, 98* (4), 732-749.

Glass, G.V. & Smith, M.L. (1978). *Meta-analysis of research on the relationship of class size and achievement.* San Francisco: Far West Laboratory for Educational Research and Development.

Goldstein, H., Yang, M., Omar, R., Turner, R., & Thompson, S. (2000). Meta-analysis using multilevel models with an application to the study of class size effects. *Journal of the Royal Statistical Society: Series C (Applied Statistics),* 49(3), 399-412.

Gollwitzer, P.M., & Sheeran, P. (2006). Implementation Intentions and Goal Achievement: A Meta-Analysis of Effects and Processes. *Advances in Experimental Social Psychology, 38*, 70-119.

Gonzalez, J.E., Nelson, R.J., Gutkin, T.B., Saunders, A., Galloway, A. & Shwery, C.S. (2004). Rational Emotive Therapy with Children and Adolescents: A Meta-Analysis. *Journal of Emotional and Behavioral Disorders, 12*:4, 222-235.

Gorey, K.M. Early childhood education: A meta-analytic affirmation of the short- and long-term benefits of educational opportunity. School Psychology Quarterly, 16(1), 9-30. doi: 10.1521/scpq.16.1.9.19163

Haller, E.P., Child, D.A. & Walberg, H.J. (1988). Can Comprehension be taught? A Quantitative Synthesis of "Metacognitive Studies". *Educational Researcher*, 17.9 pp 5-8.

Higgins, S., Hall, E., Baumfield, V., & Moseley, D. (2005). *A meta-analysis of the impact of the implementation of thinking skills approaches on pupils*. In: Research Evidence in Education Library. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Horak, V.M. (1981). A Meta-Analysis of Research Findings on Individualized Instruction in Mathematics. *Journal of Educational Research*, 74:4 p 249.

Igel, C.C., 2010. *The Effect of Cooperative Learning Instruction on K-12 Student Learning: A Meta-Analysis of Quantitative Studies from 1988 to 2009.* University of Virginia, Dissertation retrieved from Pro-Quest Dissertations.

Jeynes, W.H. (2005). A Meta-Analysis of the Relation of Parental Involvement to Urban Elementary School Student Academic Achievement. *Urban Education*, 40.3 pp 237-269.

Jeynes, W.H. (2007). The Relationship Between Parental Involvement and Urban Secondary School Student Academic Achievement: A Meta-Analysis. *Urban Education*, 42.1 pp 82-110.

Jeynes, W.H. (2008). A Meta-Analysis of the Relationship between Phonics Instruction and Minority Elementary School Student Academic Achievement. *Education and Urban Society. 40* (2), 151-166.

Johnson, D.W., Maruyama, G., Johnson, R., & Nelson, D. (1981). Effects of Cooperative, competitive and individualistic goal structures on Achievement: A meta-analysis. *Psychological Bulletin, 89*:1, 47-62.

Johnson, D.W., Johnson, R.T. & Beth-Stanne, M. (2000*). Cooperative Learning Methods: A Meta-Analysis*. Minneapolis, MN: University of Minnesota.

Jun, S.W., Ramirez, G., & Cumming, A. (2010). Tutoruing Adolescents in Literacy: A Meta-Analysis. *Journal of Education, 45 (2),* 219-238.

Kavale, K.A. & Forness, S.R. (1987). Substance Over Style: Assessing the Efficacy of Modality Testing and Teaching. *Exceptional Children*, 54.3 pp. 228-39.

Karoly, L., Kilburn, R., Cannon, J.S. (2005). *Early Childhood Interventions: Proven Results, Future Promise*. Rand Corporation. Available at: http://www.rand.org/pubs/monographs/2005/RAND_MG341.pdf (30/6/12).

Kingston, N. & Nash, B. (2011). Formative Assessment: A Meta-Analysis and a Call for Research. *Educational Measurement: Issues and Practice, 30*:4, 28-37.

Klauer, K.J. & Phye, G.D. (2008). Inductive Reasoning: A Training Approach. *Review of Educational Research*, 78.1 pp 85-123.

Kluger, A.N. & De Nisi, A. (1996). Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119:2 pp 254-284.

Kulik, C.C. & Kulik, J.A. (1982). Effects of Ability Grouping on Secondary School Students: A Meta-Analysis of Evaluation Findings. *American Educational Research Journal*, 19:3 pp. 415-428.

Kulik C-L.C & Kulik J.A. (1984). Effects of Ability Grouping on Elementary School Pupils: A Meta-Analysis. *Annual Meeting of the American Psychological Association*

Kulik, J.A., & Kulik, C.L.C. (1987). Effects of ability grouping on student achievement. *Equity and Excellence in Education*, 23:(1-2), 22-30.

Kulik, J. A., & Kulik, C. L. C. (1992). Meta-analytic findings on grouping programs. *Gifted Child Quarterly*, 36:2, 73-77.

La Paro, K.M. & Pianta, R.C. (2000). Predicting Children's Competence in the Early School Years: A Meta-Analytic Review. *Review of Educational Research*, 70:4 pp 443-484.

Layzer, J.I., Goodson, B.D., Bernstein, L. & Price, C. (2001). *National Evaluation of Family Support Programs. Final Report Volume A: A Meta-Analysis*. Administration for Children, Youth and Families (DHHS), Washington DC.

Lauer P., Akiba, M., Wilkerson, S.B., Apthorp, H.S., Snow, D., & Martin-Glenn, M.L. (2006). Out-of-School-Time Programs: A Meta-Analysis of Effects for At-Risk Students. *Review of Educational Research*, 76:2, 275-313.

Lewis, C.P. (2004). *The Relation Between Extracurricular Activities With Academic And Social Competencies In School Age Children: A Meta-Analysis* PhD thesis submitted to the Office of Graduate Studies of Texas A&M University, College Station, Tx (ProQuest Dissertations)

Lewis, R.J. & Vosburgh, W.T. (1988). Effectiveness of Kindergarten Intervention Programs: A Meta-Analysis. *School Psychology International*, 9:4, 265-275.

Lewis, C.W., Winokur, M.A., Cobb, R.B., Gliner, G.S., Schmidt, J. (2005). *Block Scheduling in the High School Setting A Synthesis of Evidence-Based Research* MPR Associates, Inc., Berkeley, CA/ Office of Vocational and Adult Education, U.S. Department of Education.

Liao, Y.C. (2007). Effects of Computer-Assisted Instruction on Students' Achievement in Taiwan: A Meta-Analysis. *Computers and Education*, 48:2 pp 216-233.

Li, Q., & Ma, X. (2010). A meta-analysis of the effects of computer technology on school students' mathematics learning. *Educational Psychology Review*, *22*:3, 215-243.

Lou Y, Abrami P.C., Spence J.C., Poulsen, C., Chambers B., &D'Apollonia S. (1996). Within-Class Grouping: A Meta-Analysis. *American Educational Research Association, 66* (4), 423-458.

Lou, Y., Abrami, P.C., & D'Apollonia, S. (2001). Small Group and Individual Learning with Technology: A Meta-Analysis. *Review of Educational Research*, 71:3, 449-521.

Lovelace, M.K. (2002). A meta-analysis of experimental research studies based on the Dunn and Dunn learning-style model. St. John's University (New York), School of Education and Human Services). ProQuest Dissertations and Theses, p 177. Retrieved from http://search.proquest.com/docview/275698679 (20/6/12).

Lovelace, M.K. (2005) Meta-Analysis of Experimental Research Based on the Dunn and Dunn Model. *The Journal of Educational Research*, 98:3, pp 176-183.

Lysakowski, R.S., & Walberg, H.J. (1982). Instructional Effects of Cues, Participation, and Corrective Feedback: A Quantitative Synthesis. *American Educational Research Journal,* 19:4, 559-578.

Ma, X., & Kishor, N. (1997). Assessing the Relationship between Attitude towards Mathematics and Achievement in Mathematics: A Meta-Analysis. *Journal for Research in Mathematics Education. 28:1,* 26-47.

Manning, M., Hommel, & Smith (2010) "A meta-analysis of the effects of early years developmental prevention programs in at-risk populations on non-health outcomes in adolescence Children and Youth Services Review 32, 506–519

Martins, P.S. (2009). *"Individual Teacher Incentives, Student Achievement and Grade Inflation,"* IZA Discussion Papers 4051, Institute for the Study of Labor (IZA).

McGiverin, J., Gilman, D., & Tillitski, C. (1989). A Meta-Analysis of the Relation between Class Size and Achievement. *The Elementary School Journal*, 90:1, 47.

Morphy P. & Graham S. (2012). Word Processing Programs and Weaker Writers/Readers: A Meta-Analysis of Research Findings, *Reading and Writing, 25*, 641-678.

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology*, 38:1, 30.

Nelson, G., Westhues, A. & MacLeod, J. (2003). A Meta-analysis of Longitudinal Research on Preschool Prevention Programs for Children. *Prevention and Treatment*, 6(2), 46-50.

Newman, M., Bird, K., Tripney, J., Kalra, N., Kwan, I., Bangpan, M., Vigurs, C. (2010). *Understanding the impact of engagement in culture and sport: A systematic review of the learning impacts for young people* In: Research Evidence in Education Library. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Niemiec, R.P. & Walberg, H.J. (1985). Computers and Achievement in the Elementary Schools. *Journal of Educational Computing Research*, 1:4, 435-40.

Nye, C, Turner, J., & Schartz, C (2006). *Approaches to Parent Involvement for Improving the Academic Performance of Elementary School Children*. Campbell Collaboration, Campbell Library of Systematic Reviews.

Paschal, R.A., Weinstein, T. & Walberg, H.J. (1984). The effects of homework on learning: A quantitative synthesis. *The Journal of Educational Research*, 78:2, 97-104.

Patall, E.A., Cooper, H., & Allen, A.B. (2010). Extending the School Day or School Year: A Systematic Review of Research (1985-2009). *Review of Educational Research, 80*:3, 401-436.

Payton, J., Weissberg, R. P., Durlak, J. A., Dymnicki, A. B., Taylor, R. D., Schellinger, K. B., & Pachan, M. (2008). *The positive impact of social and emotional learning for kindergarten to eighth-grade students.* Chicago, IL: Collaborative for Academic, Social, and Emotional Learning (CASEL).

Pearson, D.P., Ferdig, R.E., Blomeyer, R.L. & Moran, J. (2005). *The Effects of Technology on Reading Performance in the Middle-School Grades: A Meta-Analysis With Recommendations for Policy* Naperville, Il: University of Illinois/North Central Regional Educational Laboratory.

Petscher, Y. (2009). A Meta-Analysis of the Relationship between Student's Attitudes towards Reading and Achievement in Reading. *Journal of Research in Reading, 33*:4, 335-355.

Reddy, L.A., Newman, E., De Thomas, C.A., Chun, V. (2009). Effectiveness of School-Based Prevention and Intervention Programmes for Children and Adolescents with Emotional Disturbance: A Meta-Analysis. *Journal of School Psychology, 47,* 77-99.

Ritter, G.W., Barnett, J.H., Genny, C.S., & Albin, G.R. (2009). The Effectiveness of Volunteer Tutoring Programs for Elementary and Middle School Students: A Meta-Analysis. *Review of Educational Research, 79:3*, 3-38.

Rogers, K. B. (2007). Lessons Learned About Educating the Gifted and Talented A Synthesis of the Research on Educational Practice. *Gifted Child Quarterly*, 51:4, 382-396.

Rohrbeck, C., 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*, 95:2, 240-257.

Romero C.C. 2009. *Cooperative learning instruction & science achievement for secondary and early post-secondary students: A systematic Review*. Dissertation, Colorado State University.

Samuels, K.S. (2003). *The relationship of school uniforms to students' achievement, attendance, discipline referrals and perceptions: an analysis of one school district.* PhD dissertation University of Alabama at Birmingham (ProQuest Dissertations and Theses)

Sander, J.P., Patall, E.A., Amoscato, L.A., Fisher, A.L., & Funk, C. (2012). A Meta-Analysis of the Effects of Juvenile Delinquency Interventions on Academic Outcomes. *Children and Youth Services Review, 34,* 1695-1708.

Sandy-Hanson A.E. (2006). A Meta-Analysis of the Impact of Computer Technology versus Traditional Instruction on Students in Kindergarten through Twelfth Grade in the United States: A Comparison of Academic Achievement, Higher Order Thinking Skills, Motivation, Physical Outcomes and Social Skills. *Doctoral Dissertation, Howard University*

Scott-Little, C. (2002). Evaluations of After-School Programs: A Meta-Evaluation of Methodologies and Narrative Synthesis of Findings. *American Journal of Evaluation*, 23:4, 387-419.

Senechal, M. & Young, L. (2008). The Effect of Family Literacy Interventions on Children's Acquisition of Reading from Kindergarten to Grade 3: A Meta-Analytic Review. *Review of Educational Research, 78*:4, 880-907.

Shulruf, B. (2010). Do extra-curricular activities in schools improve educational outcomes? A critical review and meta-analysis of the literature. International Review of Education, 56(5), 591-612.Sitzmann, T., Kraiger, K., Stewart, D. & Wisher, R. (2006). The Comparative Effectiveness Of Web-Based And Classroom Instruction: A Meta-Analysis. *Personnel Psychology* 59, 623–664.

Slavin, R.E. (1989). Class Size and Student Achievement: Small Effects of Small Classes. *Educational Psychologist*, 24. Pp 25-77

Slavin, R.E. (1990). Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis. *Review of Educational Research*, 60:3, 471.

Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2011). Effective programs for struggling readers: A best-evidence synthesis. *Educational Research Review*, 6:1, 1-26.

Slemmer, D.L. (2002). The effect of learning styles on student achievement in various hypertext, hypermedia and technology enhanced learning environments: a meta-analysis Unpublished PhD dissertation, Boise State University, Boise, Idaho (ProQuest Dissertations and Theses).

Standley, J.M. (2008). Does Music Instruction Help Children Learn to Read? Evidence of a Meta-Analysis. Update. *Applications of Research in Music Education*, 27:1, pp. 17-32.

Steenbergen-Hu, S., & Moon, S. M. (2011). The effects of acceleration on high-ability learners: A meta-analysis. *Gifted Child Quarterly*, 55:1, 39-53.

Sowell, R. E. (2012). The relationship of school uniforms to student attendance, achievement, and discipline (Doctoral dissertation, Liberty University: ProQuest Dissertations and Theses).

Tamim, R.M. (2009). *Effects of Technology on Students' Achievement: A Second-Order Meta-Analysis.* Unpublished PhD dissertation Concordia University, Montreal Canada (ProQuest Dissertations and Theses).

Tamim R.M., Bernard R.M., Borokhovski E., Abrami P.C., & Schmid R.F. (2011). What Forty Years of Research Says about the Impact of Technology on Learning: A Second-Order Meta-Analysis and Validation Study. *Review of Educational Research, 81*, 4-28.

Tamir, P. (1985). Meta-analysis of cognitive preferences and learning. *Journal of Research in Science Teaching* 22:1 pp 1-17.

Tanner, E., Brown, A., Day, N. Kotecha, M., Low, N., Morrell, G., Brown, V., Collingwood, A., Chowdry, H., Greaves, E., Harrison, C., Johnson, G., & Purdon, S. (2011). *Evaluation of Every Child a Reader (ECar):* Research Report DFE-RR114 London: Department for Education.

Tenenbaum, G., & Goldring, E. (1989). A Meta-Analysis of the Effect of Enhanced Instruction: Cues, Participation, Reinforcement and Feedback, and Correctives on Motor Skill Learning. *Journal of Research and Development in Education*, 22(3), 53-64.

Torgerson, C.J. & Elbourne, D. (2002). A systematic review and meta-analysis of the effectiveness of information and communication technology (ICT) on the teaching of spelling. *Journal of Research in Reading*, 25, 129-143.

Torgerson, C, & Zhu, D. (2003). *A systematic review and meta-analysis of the effectiveness of ICT on literacy learning in English, 5-16*. In: Research Evidence in Education Library. London: EPPI-Centre, Social Science Research Unit, Institute of Education.

Torgerson C.J., Brooks, G. & Hall, J. (2006). A Systematic Review of the Research Literature on the use of Phonics in the Teaching of Reading and Spelling. *Research Report, No.711*

Van Steensel, R., McElvany, N. Kurvers J. & Herppich S. (2011). How Effective Are Family Literacy Programs? : Results of a Meta-Analysis. *Review Of Educational Research* 81:1 pp 69-96.

Vaughn, V. L., Feldhusen, J. F., & Asher, J. W. (1991). Meta-analyses and review of research on pull-out programs in gifted education. *Gifted Child Quarterly,* 35:2, 92-98.

Walberg, H.J. (1982). What makes schooling effective? *Contemporary Education Review*, 1, 1–34.

Wasik, B. A., & Slavin, R. E. (1993). Preventing Early Reading Failure with One-to-One Tutoring: A Review of Five Programs. *Reading Research Quarterly*, 28:2, 179–200.

Waxman, H.C., Connell, M.L., & Gray, J. (2002). *A Quantitative Synthesis of Recent Research on the Effects of Teaching and Learning With Technology on Student Outcomes*. Naperville, Illinios.

Waxman, H.C., Lin, M. & Michko, G.M. (2003). *A Meta-Analysis of the Effectiveness of Teaching and Learning With Technology on Student Outcomes A Meta-Analysis of the Effectiveness of Teaching and Learning With Technology on Student Outcomes* December 2003. Technology. Naperville, Illinois.

Wheeler, M. E., Keller, T. E., & DuBois, D. L. (2010). Review of Three Recent Randomized Trials of School-Based Mentoring. *Social Policy Report*, 24:3,1-27

Wiliam, D. (2002). '*Linking Research and Practice: knowledge transfer or knowledge creation?*' Plenary presentation at the annual conference of the North American Chapter of the International Group for the Psychology of Mathematics Education, Athens, Georgia, USA, October 2002.

Willett, J.B., Yamashita, J.J. & R.D. Anderson (1983). "A Meta-Analysis of Instructional Systems Applied in Science Teaching." *Journal of Research in Science Teaching* 20:5, 405-17.

Wilson, S.J., & Lipsey, M.W. (2007). School-Based Interventions for Aggressive and Disruptive Behaviour. Update of a Meta-Analysis. *American Journal of Preventive Medicine, 33*, 130-143.

Winner, E., & Cooper, M. (2000). Mute Those Claims: No Evidence (Yet) for a Causal Link between Arts Study and Academic Achievement. *Journal of Aesthetic Education*, 34:3/4, 11.

Woessmann, L. (2010). *Cross-Country Evidence on Teacher Performance Pay.* CESifo Working Paper No. 3151 Category 5: Economics Of Education Munich: CESifo.

Wood, S., & Mayo-Wilson, E. (2012). School-Based Mentoring for Adolescents A Systematic Review and Meta-Analysis. *Research on Social Work Practice*, 22:3, 257-269.

Zief, S.G., Lauver, S., Maynard, R.A. (2006). *Impacts of After-School Programs on Student Outcomes* Oslo: The Campbell Collaboration

Zvoch K., & Stevenes J.J, (2013). Summer School Effects in a Randomized Field Trial. *Early Childhood Research Quarterly, 28*, 24-32.

Technical Appendices References

Aguinis, H., Pierce, C. A., Bosco, F. A., Dalton, D. R., & Dalton, C. M. (2011). Debunking myths and urban legends about meta-analysis. *Organizational Research Methods*, 14(2), 306-331.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher,* 13(6), 4-16.

Bloom, H.S., Hill, C.J., Black, A.R. and Lipsey, M.W. (2008). Performance Trajectories and Performance Gaps as Achievement Effect-Size Benchmarks for Educational Interventions. *Journal of Research on Educational Effectiveness*, 1.4 pp 289-328.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1(2), 97-111.

Burtless, G. (ed.) (1996). *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success.* Washington, D.C.: Brookings Institute Press.

Chan, M. E., & Arvey, R. D. (2012). Meta-Analysis and the Development of Knowledge. *Perspectives on Psychological Science*, 7(1), 79-92.

Coe R, (2002). *It's the Effect Size, Stupid: What effect size is and why it is important* Paper presented at the Annual Conference of the British Educational Research Association, University of Exeter, England, 12-14 September 2002. Available at: http://www.leeds.ac.uk/educol/documents/00002182.htm (accessed 2/3/11).

Coe, R. (2004) 'Issues arising from the use of effect sizes in analysing and reporting research' in I. Schagen and K. Elliot (Eds) *But what does it mean? The use of effect sizes in educational research.* Slough, UK: National Foundation for Educational Research. Available at: http://www.nfer.ac.uk/nfer/publications/SEF01/SEF01.pdf

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Cronbach, L.J., Ambron, S.R., Dornbusch, S.M., Hess, R.O., Hornik, R.C., Phillips, D.C., Walker, D.F. & Weiner, S.S. (1980). *Toward reform of program evaluation: Aims, methods, and institutional arrangements*. San Francisco, Ca.: Jossey-Bass.

Department for Education and Skills (DfES) (2004). *Statistics of Education: Variation in Pupil Progress* 2003 London: Department for Education and Skills. Available at: http://www.education.gov.uk/rsgateway/DB/SBU/b000481/b02-2004v2.pdf (accessed 15/06/11).

Department for Children, Schools and Families (DCSF). *Departmental Report* 2009. London: Department for Children Schools and Families. Available at: http://media.education.gov.uk/assets/files/pdf/d/departmental%20report%202009.PDF (accessed 18/05/13).

Dillon, J.T., (1982). Superanalysis. *American Journal of Evaluation* 3(4) pp 35-43.

Glass G.V., McGaw, B. & Smith, M.L. (1981). *Meta-analysis in social research* Beverly Hills, Ca: Sage.

Glass, G.V. (1977). Integrating Findings: The Meta-Analysis of Research. *Review of Research in Education*, 5: 351-379.

Greenwald, R, Hedges, LV, and Laine, RD (1996). 'The Effect of School Resources on Student Achievement'. *Review of Educational Research*, 66(3), 361-396.

Hanushek, E., and Woessman, L. (2010). The economics of international differences in educational achievement. *NBER Working Paper 15949*, April 2010, p18

Hattie, J.A. (1992). Measuring the effects of schooling. *Australian Journal of Education*, 36, 5-13.

Hattie, J.A. (2008). *Visible Learning.* London: Routledge.

Hedges, L.V. & Olkin, I. (1985). *Statistical Methods for Meta-Analysis.* Orland, Florida: Academic Press.

Higgins, S. & Simpson, A. (2011). Book Review: Visible Learning. *British Journal of Educational Studies* 59.2 pp 90-92.

Higgins, S., Hall, E., Baumfield, V., Moseley, D. (2005). *A meta-analysis of the impact of the implementation of thinking skills approaches on pupils*. In: Research Evidence in Education Library. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Hill, C.J., Bloom, H.S., Black, A.R. & Lipsey, M.W. (2007). *Empirical Benchmarks for Interpreting Effect Sizes in Research* MDRC Working Papers on Research Methodology New York: MDRC Avaliable at: www.mdrc.org/publications/459/full.pdf (accessed 17/5/11).

Holmlund, H. McNally, S. and Viarengo, M. (2008) : *Does money matter for schools?,* IZA discussion papers, No. 3769, http://nbn-resolving.de/ urn:nbn:de:101:1-20081126249

Ioannidis, J.P.A. & Trikalinos, T.A. (2007). The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. *Canadian Medical Association Journal* 176 p 8.

Levačić, R., & Vignoles, A. (2002). Researching the links between school resources and student outcomes in the UK: a review of issues and evidence. *Education Economics* 10 (3), 312-331.
Luyten, H. (2006). 'An empirical assessment of the absolute effect of schooling: regression/ discontinuity applied to TIMSS-95'. *Oxford Review of Education*, 32: 3, 397-429.
Luyten, H., Peschar, H., Coe, R. (2008). Effects of Schooling on Reading Performance, Reading Engagement, and Reading Activities of 15-Year-Olds in England. *American Educational Research Journal* 45 (2) pp 319-342.
Pugh, G., Mangan, J., & Gray, J. (2011). Do increased resources increase educational attainment during a period of rising expenditure? Evidence from English secondary schools using a dynamic panel analysis. *British Educational Research Journal*, 37(1), 163-189.
Marzano, R.J. (1998). *A Theory-Based Meta-Analysis of Research on Instruction*. Aurora, Colorado, Mid-continent Regional Educational Laboratory. Available at: http://www.mcrel.org:80/topics/products/83/ (viewed 31/05/11).
National Foundation of Educational Research (2011). *Appendix A: Progress in International Reading Literacy Study (PIRLS) 2011 Overview.* Slough: NFER. Available at http://www.nfer.ac.uk/publications/PRTZ01/PRTZ01App.pdf (accessed 18/05/13).
OECD (2009). *PISA 2009 Results: Learning Trends.* Paris: OECD. Available at http://www.oecd.org/pisa/pisaproducts/48852742.pdf (accessed 18/05/13).
Schagen, I. & Hodgen, E. (2009). *How Much Difference Does It Make? Notes on Understanding, Using, and Calculating Effect Sizes for Schools* Wellington NZ: Research Division, Ministry of Education Available at: http://www.educationcounts.govt.nz/publications/schooling/36097/2 (accessed 5/6/11).
Sipe, T. & Curlette, W.L. (1997). A Meta-Synthesis Of Factors Related To Educational Achievement: A Methodological Approach To Summarizing And Synthesizing Meta-Analyses*. International Journal of Educational Research* 25.7. pp 583-698.
Slavin, R. & Smith, D. (2009). The Relationship Between Sample Sizes and Effect Sizes in Systematic Reviews in Education. *Educational Evaluation and Policy Analysis* 31.4: 500-506.
Smith, N.L. (1982). Evaluative Applications of Meta- and Mega-Analysis. *American Journal of Evaluation* 3(4) pp 43.
Steele, F., Vignoles, A., & Jenkins, A. (2007). The effect of school resources on pupil attainment: a multilevel simultaneous equation modeling approach. *Journal of Royal Statistical Society* Series A,.170(3), 801-824.
Tanner, E., Brown, A., Day, N., Kotecha, M., Low, N., Morrell, G., Brown, V., Collingwood, A., Chowdry, H., Greaves, E., Harrison, C., Johnson, G., & Purdon, S. (2011). *Evaluation of Every Child a Reader (ECar): Research Report DFE-RR114* London: Department for Education.
Tymms, P. (2004). Are standards rising in English primary schools? *British Educational Research Jounrnal*, 30(4), 477-494.
Tymms, P. & Merrell, C. (2007). *Standards and quality in English primary schools over time: the national evidence.* Primary Review Research Survey 4/1. Cambridge: The Primary Review.
Vignoles, A., Levačić, R., Walker, J., Machin, S. & Reynolds, D. (2000). *The Relationship Between Resource Allocation and Pupil Attainment: A Review* London: Department for Education and Employment.
Walberg, H. J. (1984). Improving the productivity of America's schools. *Educational Leadership*, 41(8), 19-27.
Wiliam, D. (2002). '*Linking Research and Practice: knowledge transfer or knowledge creation?*' Plenary presentation at the annual conference of the North American Chapter of the International Group for the Psychology of Mathematics Education, Athens, Georgia, USA, October 2002.
Wiliam, D. (2010). 'Standardized Testing and School Accountability'. *Educational Psychologist*, 45: 2, 107-122.
Wiliam, D., Lee, C., Harrison, C. & Black P. (2004). Teachers developing assessment for learning: impact on student achievement. *Assessment in Education*, 11, 1 49-65.