

Future Academy®'s Multidisciplinary Conference

What Works or What's Worked? Evidence from Education in the United Kingdom

Maria Katsipataki^{a*}, & Steve Higgins^a

^a *Durham University, School of Education, Leazes Rd. Durham, DH1 1TA, UK*

Abstract

This paper deliberates the quality of existing evidence on educational school-based interventions aiming to improve pupils' attainment. An overall picture as well as more specific examples derived from the "Sutton Trust-Education Endowment Foundation (EEF) Teaching and Learning Toolkit" will be discussed presenting different cases and versions of the Toolkit (Early Years Toolkit). The main sources of studies are meta-analyses and systematic reviews with quantitative data, where they have enough information for the calculation of effect sizes (ES). This is an ongoing project; so some representative finding will be discussed providing a best estimate of what is likely to be beneficial in a number of areas of educational practice.

© 2016 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of Future Academy® Cognitive Trading

Keywords: educational interventions; attainment; effect size; meta-analysis

1. Introduction

The present paper will discuss the work and findings from our ongoing research focusing on the following areas; the "Sutton Trust-EEF Teaching and Learning Toolkit"; school-based interventions, effective spending of the pupil premium in English schools; improving academic achievement and the "Early Years Toolkit". Schools in England receive the pupil premium which is an additional funding for public schools to help raise the attainment of disadvantaged pupils (e.g. pupils who receive free school meals) Department for Education (Dfe) 2015. After receiving this extra funding schools have to decide how they will best use the pupil premium to raise attainment.

* Corresponding author. Tel.: +44 191 334 8137

E-mail address: maria.katsipataki@durham.ac.uk

But how can the schools take effective decisions on what programme or intervention will be more likely to work so as to increase children's attainment? Most of us are aware that there is an abundance of programmes promising increased results and significant academic gains. Unfortunately, many of these programmes are not evidence-based and have not been properly evaluated. Therefore, there is a high risk that schools will spend this funding on resources or approaches which may not be effective. The Sutton Trust-EEF Teaching and Learning Toolkit has been developed based on the importance of schools making informed decisions on their spending. We do recognize that it is difficult to establish a straightforward link between school spending and pupils' learning. Some research suggests that there is no association between how much a school spends and the average academic performance of a student even after controlling for variables such as school and family (Hanushek & Woessman, 2010). On the other hand, there is research suggesting that there is a strong link between additional spending and academic outcomes, but since the relationship is complex we need higher quality data to understand the specific patterns where spending and attainment are linked (Vignoles et al., 2000). Therefore, making decisions on spending for better learning is not as easy, especially when there are so many variables affecting this relation (such as different school contexts, teachers with different skills and aptitudes, pupils of varying backgrounds and capabilities, and the range of possible specific academic outcomes to name but a few). Although findings suggest that it is difficult to bring about change especially to improve the learning of disadvantaged pupils, there are still areas which offer better prospects than others. This is what the *Toolkit* seeks to provide.

The *Toolkit* is an online live resource which is updated regularly as new evidence becomes available. Currently, there are 34 topics covered for ages between four and 18 years. Each topic provides information in terms of the average impact on academic achievement, strength of the available evidence and relevant costs. Since the development of the first versions of the Toolkit it has received increased interest from practitioners, researchers and the Department for Education in England which has reinforced the use of the *Toolkit* as a source for teachers' decisions on the pupil premium (DfE, 2014).

Another version of the *Toolkit* was recently developed which is focused on the Early Years, which mainly covers evidence available about approaches to support the learning of young children age under five years and is comprised of 12 topics. The early years are crucial for later academic success and overall educational development. Many studies demonstrate that early intervention can prevent or decrease a number of academic or psychological weaknesses (Anderson et al., 2003; Camilli, Vargas, Ryan & Barnett, 2010). In addition the pupil premium in England was extended to include younger children so; it became evident that a Toolkit focusing on the early years would be useful to practitioners.

In the sections to follow we will discuss the methodology used for the development of the Toolkit as well as some representative findings derived from both the Sutton Trust-EEF Teaching and Learning Toolkit and the Early Years Toolkit.

2. Methodology

The next sections will present the methods used to synthesize the different areas of the *Toolkit*. More specifically, a presentation of our systematic review process, the measures used to depict findings, and our inclusion/exclusion criteria will be discussed.

2.1. Meta-analysis & Effect sizes

Although first used in education and psychology (Glass, 1976), meta-analysis has perhaps been more widely applied in clinical research (Chalmers et al. 2002). It has been proven to be a very useful tool, since it integrates findings from a large number of single studies, aiming to provide greater security about findings by combining studies of impact. By using meta-analysis researchers are able to produce "a pooled estimate of effect" producing more secure findings. This is of particular importance especially if we consider the number of studies in existing

educational research. By using meta-analysis researchers are able to aggregate different studies and compare the impact of diverse approaches.

On the other hand, there are a number of criticisms arguing against the use of meta-analysis, with the main one regarding its validity (Sharp, 1997). More specifically, there is the view that quantitative synthesis might not compare similar things (commonly known as the “apples and oranges comparison”). But as Glass (2000) also argued, in studying fruit, comparing apples and oranges is sensible. The key issue is how these comparisons are being made and what inferences are drawn from this. How are the studies different and in which respects? Is it the nature of the research questions asked, or the definitions used which explains variation in findings? Clarity in terminology and classification of studies is therefore vital.

There are useful steps that can be taken in order to minimize any potential biases during systematic review for the eligibility criteria for the studies found, careful observation of the moderators included in the meta-analysis and evaluation of potential publication biases. Despite the limitations, we are confident that the use of meta-analysis or quantitative synthesis is still the best available method to answer cumulative and comparative questions about a variety of educational approaches and outcomes.

Following the initial choice of focusing on meta-analytical studies while we perform the systematic review, an overall estimate of impact is made using effect sizes. Effect sizes are widely used in evaluation research and considered valuable indicators of the relations under investigation, since they do not just indicate if an intervention works, but measure the magnitude of the observed effect. The effect size used in the Toolkit is the standardized mean difference between two groups (Coe, 2002). There are different ways of calculating the effect sizes but practically these differences are not usually large in terms of the final effect size. In terms of statistical significance we also tend to report confidence intervals along with the effect size; where when the confidence interval includes zero statistical significance has not been reached. In order to have a better picture of the magnitude of the effect size Cohen (2013) has developed a general guideline suggesting that an effect of 0.2 is considered small, 0.5 medium and 0.8 as large. It is worth mentioning that on average in educational research an effect size of around 0.4 is the overall average (Sipe and Curlette, 1997; Hattie and Timperley, 2007; Hattie, 2009), but we need to consider that a smaller effect size can be beneficial depending on other variables such as cost-effectiveness. In order to make findings more accessible and understandable to school teachers and educational practitioners we have converted the effect sizes in both Toolkits to months of additional gains in attainment; an element that has been proven to be very popular increasing the Toolkits’ overall accessibility and user-friendly character. More details and examples will be provided in later sections. Overall, the focus of the Toolkit is to provide information from the different research areas included not about “what works” but about “what has worked”. We think this is an important point in terms of understanding the inferences from the evidence. It is not possible to guarantee generalisability, however this does not mean that the information is not valuable. By emphasizing that; it shows what has been (or has not been) effective, across the studies included, on average, it can help practitioners decide how to improve outcomes for the pupils for which they are responsible.

2.2. Eligibility Criteria

So as to ensure quality in a systematic review we have applied specific eligibility criteria. The systematic review involves a structured search using a number of search engines or gateways (for example: FirstSearch, Web of Science, Jstor, ERIC, ProQuest Dissertations, Google Scholar) as well as a search of relevant peer-reviewed journals specialising in reviews (for example: Educational Research Review and Educational Research). Additionally, a hand-search of the references of the meta-analyses is performed and included when studies can be retrieved. Once the studies are retrieved we check them against our inclusion/exclusion criteria. These include;

- Inclusion of studies with academic outcomes

- Adequate information for the calculation of effect sizes
- School and pre-school children
- Experimental or quasi-experimental design

Once the final list of studies has been collected and reviewed; we extract relevant information from each included study. In cases where there are no meta-analyses on certain topics under investigation we collect either correlational or large-scale studies estimating a broader effect for that area. One example that can be drawn from the *Toolkit* is the topic of “Performance Pay” where findings are based on one correlational study and a single empirical study. In similar cases we report an evidence rating indicating that the evidence is very limited. Similar quality ratings will be discussed providing examples in the findings section.

2.3. *Cost effectiveness estimates*

As mentioned earlier, the development of a cost-effectiveness estimate is important when teachers are weighing their options as to which approach they will use. We estimate the costs of an intervention having a class of 25 students. Depending on the approach additional training or materials are also being calculated and added to the overall cost. Table 1 shows the scale used for the costing:

Table 1. Cost-effectiveness estimates

£	Very low: up to about £2000/year/class of 25 pupils, or less than £80/pupil/year
££	Low: £2,001-£5,000/year/class of 25 pupils, or up to about £170/pupil/year
£££	Moderate: £5,001 to £18,000/year/class of 25 pupils, or up to about £700 /pupil/year
££££	High: £18,001 to £30,000/year/class of 25 pupils, or up to £1,200/pupil
£££££	Very high: over £30,000/year/class of 25 pupils, or over £1,200/pupil. By 2014/5, the Pupil Premium is projected to rise to approximately £1,200/pupil

As is evident from Table 1, depending on the nature of the approach costs may vary from very low to very high. This is another variable that plays a significant role in the teachers’ choice of adopting a specific intervention to help disadvantaged learners. Another variable that follows a similar rationale is the one regarding the evidence rating represented by a padlock symbol and ranging from one to five (see below).

3. Toolkit areas and findings

The Sutton Trust-EEF Teaching and Learning Toolkit is currently comprised of 34 different themes each of which has a user friendly overview which includes a definition, an estimate of impact in terms of average months gain, issues for practitioners to consider (such as what might improve implementation), and an overview of the security of the evidence with an estimate cost effectiveness. An appendix lists the references used, and provides more detail about the effect sizes by study, strength of evidence, cost assumptions and suggested programmes which have with evidence of impact. The Early Years version of the *Toolkit* has 12 areas following the same structure and format.

3.1. *Sutton Trust-EEF Teaching and Learning Toolkit & Early Years Toolkit*

The included approaches range in effect from -4 to +8 months additional gain in attainment. Results however need to be interpreted considering the other variables outlined above. For example “extending school time” may

have a relatively low impact on attainment (+2 months) but it is cheap to administer compared with one-to-one tutoring which has more extensive evidence, but which is also more expensive. Therefore, teachers might choose a less costly intervention with lower impact on attainment, rather than an intervention which is more expensive to administer. Table 2 shows a selection of three of *Toolkit's* 34 topics.

Table 2. Findings from the Sutton Trust-EEF Teaching and Learning Toolkit

Toolkit Topic	Cost	Evidence	Impact
Behaviour interventions	£££	🔒🔒🔒🔒	+4 months
Extending school time	£££	🔒🔒	+2 months
One-to-one Tuition	££££	🔒🔒🔒🔒	+5 months

This is the first level of the online version of the *Toolkit* that one gets on the first page as a quick overview. Once a selection of a specific intervention is made then more detailed information is available for the user. The information presents the intervention in detail and suggests ways that the approach can be applied along with some programmes (if available) that have been included based on our selection criteria. Figure 1 shows each topic along with its average effect size.

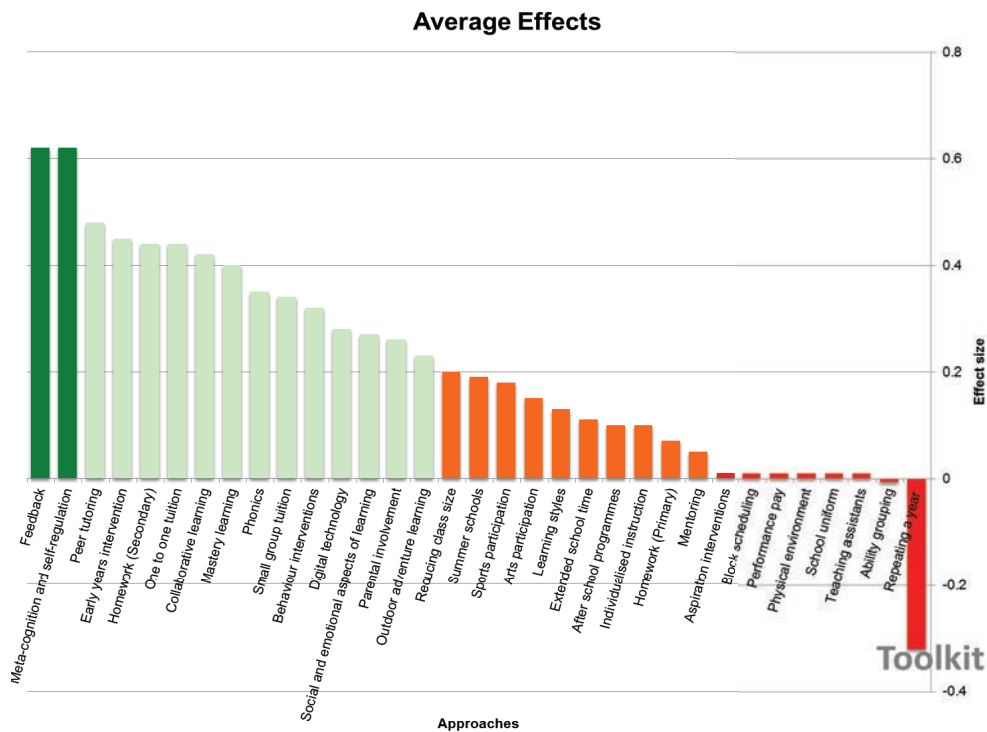


Figure 1 Average Effect sizes across the Toolkit




These findings provide a rich evidence base about different school-based interventions which aims to inform practitioners' decisions about supporting pupils from disadvantaged backgrounds. More detailed examination of the categories separately, such as by conducting more comparable meta-analyses and by undertaking further investigation of what causes variation in effect sizes across studies is a future aim of the current research. At a first

glance we can observe that approaches such as feedback and meta-cognition have the largest impact on attainment, on average, whereas repeating a school year and ability grouping tend to affect pupils’ attainment levels negatively.

3.2. Early Years Toolkit

The Early Years Toolkit has been developed with the same broad criteria but focuses on younger ages and also involves some approaches that are not included in the other Toolkit (such as play-based learning). The Early Years Toolkit emerged from a gap in the literature of high quality interventions focusing on the early years and responded to a policy change in England to extend the pupil premium eligibility to 3-5 year olds. Furthermore, teachers and other practitioners communicated the need for a *Toolkit* only for this early age group. Table 3 presents some headline findings from the Early Years Toolkit. The included approaches range from 0 to +7 months additional gain of academic performance.

Table 3. Findings from three of the Early Years toolkit topics

Toolkit Topic	Cost	Evidence	Impact
Extra Hours	£££££		+3 months
Parental Engagement	£££		+5 months
Earlier starting age	£££££		+6 months

Having a general overview, based on the areas covered by the Early Years Toolkit we found self-regulation strategies to have the highest impact (+7 months) whereas the physical environment to have no impact at all (0 months). Like it was mentioned previously this early years version of the *Toolkit* is relatively new and will be enriched and updated as new evidence becomes available.

4. Conclusions

Why is the *Toolkit* different from similar ‘What Works’ summaries of educational interventions? The *Toolkit* combines characteristics that together create a unique comparative evidence base. It focuses on rigorous quantitative studies of intervention research, provides an estimate of impact translated into months learning progress, and offers a cost-estimate of the interventions presented providing clear comparative evidence. We believe that the practical significance of an effect depends on its relative costs and benefits. Only when presented with the full picture we can make informed decisions, especially in an area like education. Overall, the aim of the *Toolkit* is to provide evidence of what has worked; therefore suggesting what is likely to work in the future. There are also limitations surrounding the *Toolkit* mainly including the limitations that surround the use of meta-analysis as mentioned earlier in this paper. Further research and a more in depth analysis of the available findings is necessary in order to examine more carefully possible relations and variations between and within strands. Another aim is to communicate and introduce the *Toolkit* outside the borders of the United Kingdom. A first attempt has been already successful with an Australian version of the *Toolkit* including Australian studies for contextual and implementation evidence (for more information please visit: <http://australia.teachingandlearningtoolkit.net.au/>). This paper has provided a brief overview of our ongoing research covering different categories of school-based interventions aiming to improve pupils’ attainment. We therefore suggest that educational research should be research-based and directed towards the child’s well-being both in an educational and psychological level.

Acknowledgements

The authors would like to acknowledge the Sutton Trust and the Education Endowment Foundation who supported the creation of the Toolkit and its further development.

References

- Australian Teaching and Learning Toolkit. <http://australia.teachingandlearningtoolkit.net.au/>
- Anderson, L. M., Shinn, C., Fullilove, M. T., Scrimshaw, S. C., Fielding, J. E., Normand, J., & Task Force on Community Preventive Services. (2003). The effectiveness of early childhood development programs: A systematic review. *American journal of preventive medicine*, 24(3), 32-46.
- Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *The Teachers College Record*, 112(3).
- Chalmers I, Hedges LV, Cooper H (2002). A brief history of research synthesis. *Evaluation and the Health Professions* 25: 12-37.
- 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>
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press Inc., New York.
- Department for Education (DfE), 2014. Research priorities and questions. Pupil Premium: raising achievement of disadvantaged pupils. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/288078/Pupil_Premium_Research_priorities_and_questions.pdf
- Department for Education (DfE), 2015. Policy Paper 2010 to 2015 government policy; education of disadvantaged children. <https://www.gov.uk/government/publications/2010-to-2015-government-policy-education-of-disadvantaged-children/2010-to-2015-government-policy-education-of-disadvantaged-children>
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational researcher*, 3-8.
- Glass, G. V. (2000). Meta-analysis at 25. <http://www.gvglass.info/papers/meta25.html>
- Hanushek, E., and Woessman, L. (2010). The economics of international differences in educational achievement. *NBER Working Paper 15949*, April 2010, p18.
- Hattie, J. (2009). *Visible learning: A synthesis of 800+ meta-analyses on achievement*. Abingdon: Routledge.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1), 81-112.
- Sharpe, D. (1997). Of apples and oranges, file drawers and garbage: why validity issues in meta-analysis will not go away. *Clinical psychology review*, 17(8), 881-901.
- 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
- 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.