Centre of Excellence for Development Impact and Learning

March 2022

Design Paper 14

Using meta-analysis to explore the transferability of education mid-range theories to Cameroon, Chad, Nigeria and Niger (S-144): research design paper

Jonathan Kay, Steve Higgins, Emma Dobson and Patrick Okwen

LIDC London International Development Centre

UKaid from the British people

Oxford Policy Management

About CEDIL

The Centre of Excellence for Development Impact and Learning (CEDIL) is an academic consortium supported by the UK government through UK Aid. The mission of the Centre is to test innovative methodologies in evaluation and evidence synthesis and promote evidence-informed development. CEDIL-supported projects fall into three programmes of work: evaluating complex interventions, enhancing evidence transferability, and increasing evidence use.

CEDIL design paper

CEDIL design papers present the protocols of studies supported by the CEDIL programme. They describe the objectives, the methodology and the implementation plan of impact evaluations, evidence syntheses, and methodological papers supported by CEDIL.

About this design paper

This design paper was submitted to CEDIL by the "Using meta-analysis to explore the transferability of education mid-range theories to Cameroon, Chad, Nigeria and Niger" Project S.144 team.

Contact Jonathan Kay at Jonathan.kay@eefoundation.org.uk

Suggested citation: Kay, J., Higgins, S., Dobson, E. and Okwen, P. (2022) 'Using meta-analysis to explore the transferability of education mid-range theories to Cameroon, Chad, Nigeria and Niger (S-144): research design paper', CEDIL Design Paper 14. Centre of Excellence for Development Impact and Learning (CEDIL), London and Oxford.

This project was funded by the Centre of Excellence for Development Impact and Learning (CEDIL), supported by UK aid from the UK Government. The views expressed in this research project paper do not necessarily reflect the UK Government's official policies or CEDIL.

Using meta-analysis to explore the transferability of education mid-range theories to Cameroon, Chad, Nigeria and Niger (S-144): research design paper

Introduction

This project will expand the Education Endowment Foundation's (EEF) existing evidence synthesis project to include education evidence from low- and middle-income countries (LMICs) in partnership with Durham University and Effective Basic Services (eBASE) Africa.

The aim of the project is to create an evidence portal with approaches and evidence that are relevant to the context of the Chad Basin (Cameroon, Chad, Niger and Nigeria). Evidence will be synthesised and translated into an accessible resource that is used by policymakers and practitioners in the region.

The project will also use evidence synthesis to empirically investigate the whether country income moderates the impact of approaches across a wide range of educational theories and approaches.

Policy relevance

Policymaker and practitioner facing outputs

The intended outputs of this project are targeted for the use of policymakers and practitioners in the Chad Basin. The Teaching and Learning Toolkit is used by approximately (64%) school leaders in England.[1] We have already had preliminary conversations with the Cameroon Ministry of Education in which policymakers have expressed an interest in an evidence portal summarising evidence relevant to Cameroon.

Alongside the actual evidence portal, active dissemination will take place to policymakers and practitioners. Evidence will be translated into teacher-facing courses that give advice on best practice and policymaker engagement sessions will take place.

Transferability of evidence.

Whether interventions or approaches can be adapted from high income contexts has clear implications for policymakers in LMICs. This project will begin to map which underlying theories may be promising. For example, if study location does not moderate

Innovation

Evidence portals are a recent innovation in the move towards evidence based practice and policymaking (White 2019). This project is unique in several key aspects:

- It will be the first education evidence portal that is translated for French language speakers

[1] Based on 2019 teacher polling by NFER, found here.

- It will be the first education evidence portal that presents magnitude and security of impacts for approaches within LMICs
- As part of the project, we will create the only database of studies that contains full details of context, methodology and impact between high, middle and low-income countries

The final point is crucial for beginning to explore the transferability of evidence empirically within education. Current databases or evidence portals, such as the What Works Clearinghouse or 3ie evidence repository, focus on a specific country, region or income level. This silo-ing of research means that - aside from replication trials of specific interventions between countries - there has been little effort to systematically explore the transferability of educational theories (for example, metacognition) between high and low income contexts.

Previous translations of evidence portals – for example, the Australian version of the Teaching and Learning Toolkit, have re-contextualised evidence by conducting rapid evidence reviews of local research and displaying this research alongside the global estimates of impact. This is the first re-contextualisation of an evidence portal that will empirically examine the transferability of the global evidence to a specific context.

Technical design

The project comprises of several distinct phases:

1. Stakeholder engagement and mapping of relevant strands
2. Updating relevant Toolkit strands to identify the transferability of practices to LMICs
3. Adding new topics of approaches that are widely used in LMICs

Meetings with stakeholders (see separate paper on stakeholder engagement) are used to identify which of the existing strands in the Teaching and Learning Toolkit could be applicable to the Chad Basin. For example, whether the underlying theory of change actually be applied within a LMIC context. The consultation is also used to identify new topics that are not currently part of the Teaching and Learning Toolkit.

The second two pieces of work will happen simultaneously. Searches of grey literature will be conducted to identify additional studies within existing strands that take place in LMICs. For example, adding studies from the 3ie evidence repository or from the JPAL list of evaluations.

Systematic searches will be conducted for the new topics identified through stakeholder engagement (see below for more details on search strategy).

Identified studies will then be added to the database. The methods between existing strands and new strands will be consistent to ensure comparability. Protocols for new strands with details on search terms, strand specific data extraction and LMIC specific data extraction will be published on the EEF website.

Inclusion criteria for the EEF Evidence Database

The inclusion criteria aim to identify relevant educational evidence for schools and policy makers interested in school-based education, consistent with the mission of the EEF, which is dedicated to breaking the link between family income and educational achievement. Specifically, the EEF aims to:

- raise the attainment of 3-18 year-olds, particularly those facing disadvantage;
- develop their essential life skills; and
- prepare young people for the world of work and further study.

A PICOS and SPIDER analysis (Methley et al. 2014) were used to define the database scope:

PICOS	***SPIDER***	***Database scope***	***Explanation and examples***
Population	Sample	Early years and school age learners from 3-18 learning in their first language.	The focus is on educational settings. This can include out of school interventions, such as summer schools or after school clubs, where the aim is to improve academic learning; or where the impact of the activity is evaluated in terms of its educational benefit (e.g. scouts or guides or an 'Outward Bound' course). Higher education settings (degree level) are excluded. Studies of second language learners (L2) studying subjects other than an additional language are excluded[2].
Intervention	Phenomenon of interest	Educational intervention or approach, including named or clearly defined programmes and recognisable approaches classifiable according to the Toolkit strand definitions (e.g. peer tutoring or small group teaching). The intervention or approach is undertaken in a normal educational setting or environment for the learners involved, such as a nursery or school or a typical setting (e.g. an outdoor field centre or museum).	The focus is on the ecological validity of the research. The intervention or approach should have a duration of at least one week or a minimum of five hours of activity time in terms of learners' experience. This excludes laboratory studies or atypical environments used to test theoretical rather than educational questions.
Comparison	Design	A valid comparison between those receiving the educational intervention or approach and those not receiving it[3].	The aim is to provide an estimate of impact based on a counterfactual. Studies would be excluded where this is no control for maturation (e.g. single subject studies or single cohort designs

[2] A study of Spanish speaking students learning mathematics in English would be excluded. A study of Spanish speaking students learning French in a Spanish medium school would be included.

[3] Specific design features are identified through coding so that these can be investigated as moderators.

			with pre- and post-test only for the intervention or approach).
Outcome(s)	Evaluation	Assessment of educational or cognitive achievement which reports quantitative results from testing of attainment or learning outcomes such as by standardised tests or other appropriate curriculum assessments or school examinations or appropriate cognitive measures.	The focus is on educational achievement in schools or other educational settings. The availability of non-cognitive outcomes is recorded, but these are not extracted because of the challenge of commensurability.
Study design	Research type	Designs where a quantitative estimate of the impact of the intervention or approach on the educational attainment of the sample can be calculated or estimated in the form of an effect size (standardised mean difference) based on a counterfactual comparison.	A standardised mean difference of the impact of the intervention or approach must be reported or must be calculable[4], such as from randomised controlled trials, quasi-experimental studies, regression continuity designs and natural experiments with a valid comparison. In addition, the standard error of this effect must be reported, calculable or estimable.

This analysis was used to create specific inclusion and exclusion criteria[5].

Inclusion criteria	**Excluded**
The majority of the sample (>50%) on which the analysis is based are learners or pupils aged between 3-18 (further education or junior college students are be included where their study is for school level qualifications).	The majority of the sample are: post-secondary education; in higher education; adults; infants under 3; other students over 18.
Evaluates the impact of an educational intervention or approach, including named or clearly defined programmes and recognisable approaches classifiable according to the Toolkit strand definitions (see Appendix B).	Intervention or approach is not classifiable applicable to the current Toolkit strand definitions (see Appendix B).

[4] This includes other measures of impact such as correlational and categorical effect sizes where these result from a counterfactual comparison and where they can meaningfully be converted to a standardised mean difference (Borenstein, Hedges, Higgins, & Rothstein, 2005).

[5] Sample size is not included in these criteria. This is because we intend to undertake an analysis of the relationship between sample size and effect size based on the existing evidence of an inverse relationship in education (e.g. Slavin and Smith, 2009) and other fields (e.g. Kühberger et al. 2014 and Button et al. 2009). This has implications for meta-analysis as methods for publication bias and the use of a random effects model both assume sample size and effect size are independent.

The intervention or approach is undertaken in a normal educational setting or environment for the learners involved, such as a nursery or school or a typical setting (e.g. an outdoor field centre or museum).	Laboratory studies Specially created environments (both physical and virtual) designed for theoretical research questions, rather than educational benefit.
A valid counterfactual comparison between those receiving the educational intervention or approach and those not receiving it.	Single group and single subject designs where there is no control for maturation or growth.
Assessment of educational or cognitive achievement which reports quantitative results from testing of attainment or learning outcomes such as by standardised tests or other appropriate curriculum assessments or school examinations or appropriate cognitive measures.	Attitudinal, affective or motivational outcomes.
A quantitative estimate of the impact of the intervention or approach on the educational attainment of the sample involved in the intervention or approach can be calculated or estimated in the form of an effect size (standardised mean difference) with its standard error based on a counterfactual comparison.	Purely qualitative outcomes Studies where an effect size (standardised mean difference) and standard error cannot be identified, calculated or estimated with reasonable precision.[6]

Search strategy for new strands

For topics identified through our stakeholder engagement, new systematic searches will be undertaken. These sources will be used (gateways and databases):

- First search
  - Article First
  - ECO
  - Papers First
  - World Cat Dissertations
- EBSCO
  - BEI
  - Education Abstracts
  - Education Administration Abstracts
  - ERIC
  - PsycArticles
  - PsycINFO
- Taylor and Francis
  - Educational Research Abstracts Online
- ProQuest

[6] Such as by using the conversions available in programs like Comprehensive Meta-Analysis, or David B. Wilson's online conversion tool: https://campbellcollaboration.org/escalc/html/EffectSizeCalculator-Home.php .

    - ProQuest Dissertations and theses (Global)
- Elsevier
    - Science Direct
- Thomson Reuters
    - Web of Science

In addition, informal searching for 'grey' literature (reports and unpublished studies) is undertaken using 3ie evidence repository, Google, Google Scholar and Microsoft Academic.

Our approach does not use citation searching, 'pearl growing' (Schlosser et al. 2006) or expert nomination, though we used these techniques to ensure the adequacy of search terms (Papaioannou, 2010). Our rationale for this is that the use of such approaches on their own, without subsequently adapting the search criteria are likely to increase the risk of publication bias (Higgins, 2018). Where we identify includable studies from non-systematic approaches we aim to refine our search criteria and to run additional searches to find other similar studies retrieved with the amended search strings.

An example of the search strings and recording process for Teaching Assistants is included in Appendix C.

Description of methods used in the included studies

The inclusion criteria aim to identify studies with a valid counterfactual comparison between those receiving the educational intervention or approach and those not receiving it. True experimental (randomised) and quasi-experimental studies (both prospective and retrospective) designs are therefore included if they feature two educational conditions addressing the central theme of each Toolkit strand (e.g. peer tutoring compared with no peer tutoring or studies contrasting reduced class size with normal or usual sized classes). Other designs such as interrupted time series or regression discontinuity are included where they similarly provide an estimate of the effect of the intervention or approach. Design features are coded to allow for exploratory analysis. The different counterfactual conditions are:

- an active control (i.e. there is control for novelty such as with another introduced new intervention or 'treatment');
- business as usual (i.e. comparison group having their usual learning experience);
- no equivalent teaching (i.e. additional learning time, where the control or comparison group have no typical educational experience, such as in a Summer School intervention or a Before or After school club)[7].

Identifying the primary outcome from a study for the Toolkit database

Identifying the best single outcome from a study is not always straightforward as the study aims are not necessarily the same as the Toolkit aims. The key principles adopted to support identification are:

- *A good test of the impact of the intervention for the Toolkit*
  The main issue to consider is the alignment of the study with the EEF Toolkit in terms of the research design and research questions. This review is seeking for the best estimate of the difference between pupils experiencing the intervention or approach with the most appropriate counterfactual condition (those not experiencing the intervention or approach).
- *An appropriate measure of educational attainment*

[7] At this stage we are not including studies which directly compare two interventions, without a control or comparison group. These studies would be valuable to include if we can identify sufficient studies for other systematic comparisons such as in a network meta-analysis (Lumley, 2002).

The next issue is the identification of which specific curriculum or cognitive outcome is most appropriate. In general, the focus is on outcomes which are good indicators of overall educational attainment, such as reading comprehension or a standardised test of mathematics. Standardised tests or national tests and examinations tend to be better overall indicators of educational performance than researcher-designed measures or teacher-designed class tests (e.g. Sammons et al. 1995; Tymms, 1999).

- *As direct and fair a measure as possible*
  Simple outcomes rather than combined ones across subjects are usually preferable (so reading *or* mathematics rather an overall score that combines both). This is not always straightforward. In a pedagogical intervention where the focus is on general strategies and is taught across several curriculum subject it can be difficult to decide which is the primary outcome for the Toolkit. Peer-tutoring delivered in reading and mathematics may have one designated as the primary and another as the secondary or they may be combined and the average reported. It may be appropriate to combine them when they are equally valid possible outcomes. In this case the separate scores for each subject would also need to be recorded so that subject specific meta-analyses can be conducted.

The research literature distinguishes between treatment inherent and treatment independent measures (e.g. Slavin and Madden, 2011). In practice they can be hard to separate. Criterion-referenced measures can be particularly problematic here. In a spaced-learning intervention in history, a school history knowledge test is only fair if the control group were also being taught the same topic in history. Another example might be a phonics intervention where the intervention group are taught letter sounds and compared with a business-as-usual control. Here a letter recognition test may not provide a fair comparison as it is likely to over-estimate the impact on reading (as opposed to impact on letter recognition). In fact, such a measure might be a better measure of implementation fidelity. On the other hand, evaluating the impact of teaching number fact recall with a standardised test of mathematics may similarly under-estimate effects if number forms only a limited part of the standardised test.

Some examples are provided below to exemplify the issues:

1) In a study of self-regulated learning in writing for primary school pupils, a researcher-designed writing test and standardised test of mathematics were both used at the beginning of the intervention. The standardised test of mathematics was used to investigate the far transfer effects of SRL. The primary outcome here would be the writing results. Coders would be expected to note that there was additional outcome data available, but that this was to assess far transfer.
2) In an evaluation of a phonics intervention a series of outcome measures were used including letter recognition, reading fluency and a picture vocabulary test. There was no overall assessment of reading comprehension. Reading fluency was designated as the primary outcome as the researchers had used a subscale of a standardised reading test. In this case the different reading measures could be combined to provide an overall measure of reading capability. This presents a complex challenge as the letter recognition test is arguably treatment inherent so may over-estimate the impact of the intervention on reading outcomes. It is difficult to see why there might be a direct impact from a phonics intervention on vocabulary. This measure is often used a broader indicator of literacy knowledge. Over a short timescale this might not capture the direct impact of the intervention
3) In a thinking skills intervention taught in secondary schools through separate lessons the primary outcome reported in the study was the impact on Raven's matrices (a standardised test of reasoning) assessed after one term at the end of the intervention. Impact was also assessed on English and mathematics using the school's end of year exam results (two terms later). In this case the cognitive outcome could be used as the primary outcome and the school examination results were added as secondary measures. The tension here is between what may seem a more

treatment inherent measure (reasoning) set against the more ecologically valid, but unstandardised, indirect and delayed curriculum measures.

These considerations have resulted in a flow diagram to aid coders in identifying the primary outcome (see Appendix G) where the aim is to identify in each study the most comparable effect size for the Toolkit, but which also takes into account the nature of the particular intervention. Additional secondary outcomes (such as alternative measures of attainment), or equivalent measure in different subjects (where applicable for the intervention) are also identified and extracted.

As part of the development of a coding tool designed for data extraction of LMICs, additional outcome measures will be added – for example, focusing on teacher or pupil attendance.

Statistical independence of findings from a single study

There are a number of threats to the validity of findings related to statistical dependence. These are:

1) use of data from the same participants for different outcomes;
2) reporting multiple outcomes of the same type; and
3) aggregating outcomes of different types for the same sample of participants.

We identify one primary outcome for the Toolkit strand from each study (see Appendix G). This is usually, but not necessarily the primary outcome of the study[8]. Other equivalent academic and cognitive outcomes are recorded as secondary outcomes. Where it is not possible to identify a single preferred outcome (such as a reading intervention where a standardised test of reading comprehension is not reported), comparable outcomes are combined to produce one overall effect for the study (such as word reading, reading fluency or decoding skills).

Details of study coding categories

Coding is undertaken with the three current data extraction tools (see Appendix D, E and F). An additional data extraction tool will be designed during the first quarter of the project, which will be designed in consultation with stakeholders to capture potential barriers to implementation or outcome measures that are specific to LMICs. Methodological quality can be assessed using features such as design, the unit of assignment/analysis, attrition reported and method of effect size estimation (Cooper, Hedges, & Valentine, 2009)[9].

- EEF main data extraction (v 1.0 June 2019), used for all studies: Appendix D;
- Strand specific (additional codes for each Toolkit strand, such as information about tutors and tutees in peer tutoring, or groups size in small group – used for studies in each strand: Appendix E;
- EEF Toolkit effect size data extraction (v 1.0 June 2019), used for all studies: Appendix F.

[8] In some cases, the active control may be the 'intervention' we are interested in for the Toolkit, for example where teaching assistant support was provided as a comparison condition to a particular intervention. In other cases, the primary outcome for a reading intervention, for example, may be specified in the study as letter recognition, but we would identify reading comprehension as the primary outcome for the Toolkit, so as to identify as comparable outcomes as possible (Higgins, 2018).

[9] We have not selected a specific quality appraisal or risk of bias tool as the evidence is limited about the validity of these tools (in medicine at least: Hartling et al. 2009; Katikireddi et al. 2009) and the choice of tool has a direct impact on the outcomes of a meta-analysis (Voss & Rehfuess, 2013). We intend to undertake methodological exploration of the relationship between features of study quality and risk of bias in the development of the database.

The main coding tool was developed based on a comparison of available and relevant alternative coding frameworks (e.g. EPPI Centre Education guidelines (version 0.97/2003), Lipsey and Wilson (2001), IES/WWC[10], 3iE[11]).

Demographic study features include learners' age, socio-economic background and attainment level, as well as subject matter studied. Substantive features across studies will be used to explore variation in terms of pedagogical codes such as, treatment duration, provision of professional development for teachers and training for students, depending on approach. All these study features will be subsequently analysed as moderators for their potential relationship with outcome effects.

All coding activities (i.e., abstract screening, full-text review, study features coding, as well as effect size extraction) will be carried out by a team of reviewers, each working independently but discussing and resolving queries, when necessary eliciting a third opinion from the core project team. All coders receive training and have to achieve an agreed level of reliability to be included in the coding team. A 10% sample of studies (per coder and per strand) are double coded to assess reliability rates[12].

Statistical procedures and conventions

The database aims to include and summarize quantifiable school attainment outcomes from primary empirical studies which meet the inclusion criteria and match the Toolkit themes. The key metric used is the Standardised Mean Difference (d-index) or effect size. A summary table of the characteristics of included studies will be reported for each meta-analysis.

For studies that report descriptive statistics for continuous measures of pupil attainment outcomes, the post-intervention mean of the control group will be subtracted from the post-intervention mean of the intervention group and the resulting difference will be divided by the pooled standard deviation, adjusted for sample size (Hedges' g). An accompanying standard error (representing the 95% confidence interval) will also be recorded[13]. Where ever possible the descriptive outcome statistics (N, means and standard deviations for control and intervention groups) will be collected, even where the study report reports an effect size and accompanying standard error, or where an effect size can be calculated from other inferential statistics.

All effect sizes will be coded either as resulting from a post-test or gain comparison. These effect sizes will be meta-analysed separately as they may represent different metrics (such as when the intervention affects the relative spread of the intervention group (Xiao et al. 2017)). For studies where there is substantial baseline imbalance[14] a gain score effect size may be selected (such as in quasi-experimental designs or natural experiments).

Outcome data are, however, likely to be reported in a variety of formats. For studies that report inferential statistics such as t, F, or p-values only, the appropriate conversion formula will be applied to calculate the *d*-index as the effect size estimate (Lipsey & Wilson, 2001; Hedges, Shymansky, & Woodworth, 1989; Hedges & Olkin, 1985). To ensure appropriate corrections for the small sample size bias, all *d*-indices will be converted to the unbiased Hedges' *g* statistic.

---

[10] https://ies.ed.gov/ncee/wwc/StudyReviewGuide

[11] http://www.3ieimpact.org/en/

[12] Our initial assessments of reliability look at percentage agreement. We intend to undertake an analysis of coding which considers the code difficulty alongside coder reliability (Stemler, 2004).

[13] The standard error can be calculated from the confidence intervals or estimated from p-values: https://handbook-5-1.cochrane.org/chapter_7/7_7_7_2_obtaining_standard_errors_from_confidence_intervals_and.htm

[14] Chance imbalance is likely to occur in randomized studies (the smaller the study the greater the risk) and can usually be dealt with through an analysis which takes account of baseline measurements. Theoretically, if sampling of randomized studies in a meta-analysis is unbiased any imbalance is likely to even out with a large number of studies. We intend to undertake further exploration of the differences between post-test only, post-test adjusted and gain estimates of effect to identify the advantages and disadvantages of the different approaches.

Within Study Synthesis

This review focuses on academic attainment outcomes. It is likely that in some studies there will be several measures of the same or similar outcomes from the same sample of learners. When this happens, we will select the most representative measure (see above). When no single outcome is judged to be appropriate in relation to the design of the study and the Toolkit strand it is included, we will average effects deriving from similar or complementary measures of school attainment. If the same group of participants is used more than once (such as the same control group compared with two different treatment groups, each applicable to the Toolkit) the sample size and associated standard error will be reduced proportionally so as it contributed fairly to the overall average.

Toolkit Strand Synthesis

Initial study analysis and data checking will be undertaken in EPPI Reviewer 4.0 (Thomas et al. 2010), the main software used for the review. The meta-analysis functions allow for complex meta-analysis analysis to be undertaken and are based on the 'metafor' package in R (Viechtbauer, 2010). Independent effect sizes will be aggregated across studies for each Toolkit strand using a random effects model (Borenstein, Hedges, Higgins, & Rothstein, 2010), as the assumptions for applying a fixed effect model will not be met (i.e. conceptual similarity of the interventions and approaches in each strand or a sample constituting the complete population of relevant studies). The results from a random effects model analysis also perhaps best represent the overall effect of a collection of educational interventions and approaches on learning across different age groups, school subjects and educational contexts.

A complete dataset for each of the 34 Toolkit strands will then be exported for further sensitivity exploratory analysis as they become available. A series of analyses will be undertaken to check aggregation of effect sizes across studies, sensitivity analyses (see below) and to replicate moderator analyses, using *Comprehensive MetaAnalysis 3.0* (Borenstein, Hedges, Higgins, & Rothstein, 2005).

A random effects model will be adopted for each meta-analysis and the heterogeneity of the distribution of the effect sizes assessed using Q[15] and I^2[16] (Higgins et al. 2009). A pre-specified set of coded study features will be further explored through moderator variable analysis under a mixed effects model, as potential sources of systematic variation (see Appendices D-F for these variables).

As part of the CEDIL-funded project, we will be conducting moderator analysis that tries to answer questions about the transferability of evidence. We will conduct moderator analysis by country, region and income level.

Sensitivity analysis

To assess potential bias associated with individual out-of-range calculated effect sizes which may potentially distort the overall interpretation of the findings, a sensitivity analysis will be undertaken (Hedges & Olkin, 1985). This is intended to determine whether the removal of a particular effect size increases the fit of the remaining effect sizes in a homogeneous distribution while not substantially affecting the interpretation of the recalculated mean effect size. Various approaches to identifying potential outliers will be used, including visual examination of data organized into a forest plots and also performing "one study removed" (Borenstein et al. 2000 - for a more exploratory approach see Baker & Jackson, 2008). Identified outliers will be examined with the potential to remove them from

[15] Cochran's Q is calculated as the weighted sum of squared differences between individual study effects and the pooled effect across the studies, with the weights being those used in the pooling method Borenstein et al. 2008).

[16] I^2 describes the percentage variation across studies that is due to heterogeneity rather than chance (Higgins & Thompson, 2002; Higgins et al., 2003). I^2 is an intuitive and simple expression of the inconsistency of studies' results.

the final dataset. Potential sources of bias, such as study design, type of treatment, publication source, missing data, sample size, or attrition, will be carefully examined through the corresponding moderator variable analyses.

Publication Bias

Relying on available and published studies may bias or inflate the overall intervention effect, particularly in education with a relatively large proportion of smaller studies. To evaluate potential publication bias across the database, we will review the associations between publication type and the pooled effect (i.e. journal article, dissertation or thesis, technical report, book or book chapter, conference paper, and other). Thesis completion is not usually influenced by the size of the effect, unlike journal articles.

Other methods for assessing publication bias will be explored, such as a visual inspection of a funnel plot or Duval & Tweedie's (2000) trim and fill routine available in Comprehensive Meta-Analysis (CMA) (Borenstein et al., 2005). Becker (2005) and Banks et al. (2012), however, recommend the discontinuation of the use of the failsafe N to assess publication bias, as the results are often inconsistent with the results from other publication bias methods. In education all of the methods to detect publication bias are problematic due to the negative association between sample size and effect size (e.g. Slavin and Smith, 2009).

References

Baker, R., & Jackson, D. (2008). A new approach to outliers in meta-analysis. *Health Care Management Science*, *11*(2), 121-131. https://doi.org/10.1007/s10729-007-9041-8

Banks, G. C., Kepes, S., & Banks, K. P. (2012). Publication bias: The antagonist of meta-analytic reviews and effective policymaking. *Educational Evaluation and Policy Analysis*, *34*(3), 259-277. https://doi.org/10.3102/0162373712446144

Becker, B. J. (2005). Failsafe N or file-drawer number. In H. R. Rothstein, A. J. Sutton, M. Borenstein (Eds) *Publication bias in meta-analysis: Prevention, assessment and adjustments*, 111-125. London: John Wiley and Sons

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2010). *Comprehensive meta-analysis version 3*. Englewood, NJ: Biostat.

Borenstein M., Hedges L.V., Higgins, J.P.T., Rothstein, H.R. (2009) Subgroup analyses. In: *Introduction to Meta-Analysis*.London: John Wiley & Sons, Ltd, pages 59-86;

Deeks, J.J., Douglas, A.G., Bradburn, M.J. (2001) Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Davey Smith G; Altman DG *Systematic Reviews in Health Care: Meta-analysis in Context.* London: BMJ Publishing Group.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*(2), 97-111. http://dx.doi.org/10.1002/jrsm.12

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience,* 14(5), 365. https://doi.org/10.1038/nrn3475

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455-463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

EEF (2018) *Sutton Trust-EEF Teaching and Learning Toolkit & EEF Early Years Toolkit Technical appendix and process manual (Working document v.01) July 2018* London: Education Endowment Foundation

Elliott, J. H., Turner, T., Clavisi, O., Thomas, J., Higgins, J. P., Mavergames, C., & Gruen, R. L. (2014). Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. *PLoS medicine*, *11*(2), e1001603. . https://doi.org/10.1371/ journal.pmed.1001603

Hartling, L., Ospina, M., Liang, Y., Dryden, D. M., Hooton, N., Seida, J. K., & Klassen, T. P. (2009). Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *British Medical Journal*, 339, b4012. https://doi.org/10.1136/bmj.b4012

Hedges, L., & Olkin, I. (1985). *Statistical models for meta-analysis*. New York: Academic Press.

Higgins, J.P.T. & Thompson, S.G. (2002) Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*;21:1539-1558. https://doi.org/10.1002/sim.1186

Higgins, J.P.T., Thompson, S.G., Deeks, J.J. & Altman, D.G. (2003) Measuring inconsistency in meta-analyses. *British Medical Journal*;327:557-560. https://doi.org/10.1136/bmj.327.7414.557

Higgins, S. (2018) *Improving Learning: Meta-analysis of Intervention Research in Education* Cambridge: Cambridge University Press.

Higgins, S., Kokotsaki, D. & Coe, R. (2011) *Toolkit of Strategies to Improve Learning: Summary for Schools Spending the Pupil Premium: Technical Appendices* London: Sutton Trust.

Katikireddi, S. V., Egan, M., & Petticrew, M. (2015). How do systematic reviews incorporate risk of bias assessments into the synthesis of evidence? A methodological study. *Journal of Epidemiology and Community Health*, 69(2), 189-195. http://dx.doi.org/10.1136/jech-2014-204711

Kühberger, A., Fritz, A., & Scherndl, T. (2014). Publication bias in psychology: a diagnosis based on the correlation between effect size and sample size. *PloS one*, *9*(9), e105825. https://doi.org/10.1371/journal.pone.0105825

Lumley, T. (2002). Network meta-analysis for indirect treatment comparisons. *Statistics in medicine*, 21(16), 2313-2324. https://doi.org/10.1002/sim.1201

Methley, A. M., Campbell, S., Chew-Graham, C., McNally, R., & Cheraghi-Sohi, S. (2014). PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. *BMC Health Services Research*, *14*(1), 579. https://doi.org/10.1186/s12913-014-0579-0

O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M., & Ananiadou, S. (2015). Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic reviews*, 4(1), 5.

Papaioannou, D., Sutton, A., Carroll, C., Booth, A., & Wong, R. (2010). Literature searching for social science systematic reviews: consideration of a range of search techniques. *Health Information & Libraries Journal*, *27*(2), 114-122. https://doi.org/10.1186/2046-4053-4-5

Sammons, P., Nuttall, D., Cuttance, P., & Thomas, S. (1995). Continuity of school effects: A longitudinal analysis of primary and secondary school effects on GCSE performance. *School Effectiveness and School Improvement,* 6(4), 285-307. https://doi.org/10.1080/0924345950060401

Schlosser, R. W., Wendt, O., Bhavnani, S., & Nail-Chiwetalu, B. (2006). Use of information-seeking strategies for developing systematic reviews and engaging in evidence-based practice: the application of traditional and comprehensive Pearl Growing. A review. *International Journal of Language & Communication Disorders*, *41*(5), 567-582. . https://doi.org/10.1080/13682820600742190

Slavin, R., & Madden, N. A. (2011). Measures inherent to treatments in program effectiveness reviews. *Journal of Research on Educational Effectiveness*, *4*(4), 370-380. https://doi.org/10.1080/19345747.2011.558986

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. https://doi.org/10.3102/0162373709352369

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, 9(4), 1-19. Available online: http://PAREonline.net/getvn.asp?v=9&n=4.

Thomas, J., Brunton, J., & Graziosi, S. (2010) *EPPI-Reviewer 4: software for research synthesis*. London: EPPI Centre, Institute of Education.

Tymms, P. (1999). Baseline assessment, value-added and the prediction of reading. *Journal of Research in Reading*, 22(1), 27-36. https://doi.org/10.1111/1467-9817.00066

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software,* 36(3), 1-48. https://cran.r-project.org/web/packages/metafor/vignettes/metafor.pdf

Voss, P. H., & Rehfuess, E. A. (2013). Quality appraisal in systematic reviews of public health interventions: an empirical study on the impact of choice of tool on meta-analysis. *Journal of Epidemiology and Community Health*, 67(1), 98-104. http://dx.doi.org/10.1136/jech-2011-200940

Wallace, B. C., Trikalinos, T. A., Lau, J., Brodley, C., & Schmid, C. H. (2010). Semi-automated screening of biomedical citations for systematic reviews. *BMC bioinformatics*, *11*(1), 55. https://doi.org/10.1186/1471-2105-11-55

Wilson, D. B., & Lipsey, M. W. (2001). *Practical meta-analysis*. Thousand Oaks CA, US: Sage.

White, H. The twenty-first century experimenting society: the four waves of the evidence revolution. *Palgrave Commun* **5,** 47 (2019). https://doi.org/10.1057/s41599-019-0253-6

Xiao, Z., Higgins, S., & Kasim, A. (2017). An Empirical Unraveling of Lord's Paradox. *The Journal of Experimental Education,* 1-16. https://doi.org/10.1080/00220973.2017.1380591

Appendix A: Stakeholder engagement mapping

<table>
<tr><td rowspan="3">Power</td><td>High</td><td>• Department of cooperation at the ministries of basic and secondary education in Cameroon
• Ministère de L'Education Secondaires (Cameroun)
• Parliament (all)
• Senate (Cameroon and Nigeria)
• Micro-finance institutions
• Banks
• Ministry of Social Affairs (Cameroon, Niger, Chad)
• Traditional Leaders
• British High Commission (Cameroon)</td><td>• Ministère de L'Education de Base (Cameroun, Niger, and Chad)
• Ministère de L'Education Secondaires (Chad and Niger)
• Prime Minister's Office
• Development Agencies (UN Women, UNICEF, World Bank, UNESCO)
• National and regional pedagogy inspectors (both basic and secondary education)
• Academia
• State Ministry of education (Nigeria)
• AEN, eBASE Africa
• CDBPS
• Media: CRTV, print press
• HP Foundation (Niger and Nigeria)
• AFIDEP (Niger, Nigeria, and Cameroon)</td></tr>
<tr><td>Low</td><td>• Informal sector learning structures
• Schools in conflict regions</td><td>• Teachers (private and public schools; primary and secondary schools; TVET and general education; formal sector education
• MBOSCUDA (indigenous group advocacy CSO)
• CAGEAD (CSO)
• Local farmers and women's groups</td></tr>
<tr><td></td><td>Low</td><td>High</td></tr>
<tr><td></td><td colspan="2">Interest</td></tr>
</table>

Appendix B: Toolkit strand definitions for current strands

Note: this does not include strands identified in stakeholder engagement with practitioners and policymakers in Cameroon, Chad, Niger and Nigeria

Strand	Definition	Notes
Arts participation	Arts participation is defined as involvement in artistic and creative activities, such as dance, drama, music, painting, and sculpture, either as an additional part of the curriculum or as extra-curricular activities. Participation may be organised as regular weekly or monthly activities or more intensive programmes such as summer schools or residential courses. Arts education and participation include a broad range of subjects including the traditional fine arts (e.g. visual arts, music, dance, performing arts, theatre and dance), modern dance and movement, poetry and creative writing, as well as teaching strategies which explicitly include arts elements such as drama-based pedagogy.	Distinguished from *Sports participation* Attainment in literacy and numeracy remain the primary outcome for these other academic subjects. This reflects the way they have been evaluated in the educational research literature. This should be interpreted as the additional potential of these subjects, which, of course, have value in and of themselves as part of the school curriculum.
Aspiration interventions	Aspiration is what an individual hopes will happen in the future. A key indicator might be a child's reported desire to continue with education post-16. A number of approaches to raising aspirations have been tried across three broad areas: 1. Interventions that focus on parents and families; 2. Interventions that focus on teaching practice; 3. Out-of-school interventions or extra-curricular activities, sometimes involving peers and mentors.	Distinguished from *Social and Emotional Learning* and from *Metacognition and Self-regulation*
Behaviour interventions	Behaviour interventions seek to improve attainment by reducing challenging behaviour, including aggression, violence, bullying, substance abuse and general anti-social activities. Three broad categories of behaviour interventions can be identified: 1. school-level approaches to developing a positive school ethos or improving discipline which also aim to support greater engagement in learning;	Reduction of negative behaviours Distinguished from *Social and Emotional Learning* and from *Metacognition and Self-regulation*

	2. universal programmes which seek to improve behaviour and generally take place in the classroom; and 3. more specialised programmes which are targeted at students with either behavioural issues or behaviour and academic problems.	
Block scheduling	Block scheduling and timetabling changes refer to alterations to lessons within the existing length of the school day, rather than approaches which seek to extend the school day or the school year.	Changes to timetabling and lesson length Distinguished from *Extending school time*
Built environment	Changing the built environment of the educational setting; either by moving to a new school building, or seeking to improve the design, air quality, acoustics, light, or temperature of an existing building or learning space.	
Collaborative learning	Collaborative or cooperative learning can be defined as learning tasks or activities where students work together in a group small enough for everyone to participate on a collective task that has been clearly assigned. Each student can then achieve his or her learning goal if and only if the other group members achieve theirs. Cooperative learning has been used to promote better achievement, improved intergroup relations, acceptance of mainstreamed classmates, enhanced self-esteem and positive attitudes.	Includes both collaboration (same task, single outcome) and co-operation (separate tasks, joint outcome). Excludes an asymmetrical teaching role such as in peer tutoring or mentoring. Distinguished from *Peer tutoring*
Digital technology	Digital technology is mainly associated with computer-assisted strategies to support learning within schools. Approaches in this area are very varied, but a simple split can be made between: 1. Programmes for students, where learners use technology in problem solving or more open-ended learning, and 2. Technology for teachers such as interactive whiteboards or learning platforms.	A very wide range of technologies and approaches are included involving teachers' use, pupils' use and the use of technology in a direct teaching or tutoring role.
Early years intervention	Early years or early childhood interventions are approaches that aim to ensure that young children have educationally based pre-school or nursery experiences which prepare for	

	school and academic success, usually through additional nursery or pre-school provision.	
Extending school time	Extending school time is defined here as a school resource that can be used to improve learning. More specifically, research has focused on three main approaches: 1. extending the length of the school year; 2. extending the length of the school day; and, 3. providing additional time for targeted groups of pupils either before or after school. This summary focuses on extending core school time and the use of targeted before and after school programmes, particularly to support disadvantaged or low attaining pupils.	Distinguished from *Homework* by a clear teaching component, also usually takes place in the school setting
Feedback	Feedback is information given to the learner and/or teacher about the learner's performance relative to learning goals or outcomes. It should aim to (and be capable of) producing improvement in students' learning. Feedback redirects or refocuses either the teacher's or the learner's actions to achieve a goal, by aligning effort and activity with an outcome. It can be about the output of the activity, the process of the activity, the student's management of their learning or self-regulation, or them as individuals. This feedback can be verbal or written, or can be given through tests or via digital technology. It can come from a teacher or someone taking a teaching role, or from peers.	
Homework (primary /secondary)	Homework refers to tasks given to pupils by their teachers to be completed outside of usual lessons. Common homework activities in primary schools tend to be reading or practising spellings and number facts, but may also include extended activities to develop inquiry skills or more directed and focused work such as revision for tests. Common homework activities in secondary schools are completing tasks assigned in lessons, such as practicing further questions or problems in mathematics, or preparing for tasks in future lessons. It may also include routine course work or revising for	Distinguished from *Extending School Time* and *Parental Engagement/Involvement*

	tests and examinations. In some models of 'flipped learning', pupils prepare at home for classroom discussion and application tasks. It also includes activities such as 'homework clubs' where pupils are given the opportunity to complete their assigned tasks in school, usually at the end of the school day.	
Individualised instruction	Individualised instruction can be defined as a teaching system where students work at their own pace under the guidance of personalised and tailored activities whilst at school. Various models of individualised instruction have been tried over the years in education, particularly in subjects like mathematics where pupils can have individual sets of activities which they complete, often largely independently. Computer-based and online approaches have also been developed, with more recent 'intelligent tutoring' systems designed to give more tailored feedback and challenge.	Distinguished from *One to One tutiton*
Learning styles	A 'learning style' is an individual's unique approach to learning based on their strengths, weaknesses, and personal preferences, often in relation to different modes of information (visual, auditory, tactile, etc.) or in relation to its organisation (e.g. abstract, concrete; wholist, serialist, etc.). Different models in the literature describe these on a continuum from fixed to malleable according to how they conceptualise a particular 'style'. The idea underpinning learning styles is that individuals all have a particular approach to or style of learning. The theory is that learning will therefore be more effective or more efficient if pupils are taught using the specific style or approach that has been identified as their learning style. For example, pupils categorised as having a 'listening' learning style, could be taught more through storytelling and discussion and less through traditional written exercises.	

Mastery learning	Mastery learning breaks subject matter and learning content into units with clearly specified objectives which are pursued until they are achieved. Learners work through each block of content in a series of sequential steps. Students must demonstrate a high level of success on tests, typically at about the 80% level, before progressing to new content. Mastery learning can be contrasted with other approaches which require pupils to move through the curriculum at a pre-determined pace. Teachers seek to avoid unnecessary repetition by regularly assessing knowledge and skills. Those who do not reach the required level are provided with additional tuition, peer support, small group discussions, or homework so that they can reach the expected level.	
Mentoring	Mentoring in education aims to develop young people's strengths by pairing them with an older volunteer, sometimes from a similar background, who can act as a positive role model. It is often characterised as aiming to build confidence, or to develop resilience and character, rather than directly focusing on teaching or tutoring specific skills. Mentors typically build relationships with young people by meeting with them one-to-one for about an hour or so a week either at school, at the end of the school day, or at weekends. Activities vary from programme to programme, sometimes including direct academic support with homework or other school tasks. Mentoring has increasingly been offered to young people who are hard to reach or deemed to be at risk of educational failure or exclusion.	Distinguished from *Aspirations* and *One to one tuition*
Metacognition and self-regulation	Metacognition and self-regulation approaches (sometimes known as 'learning to learn' approaches) aim to help learners think about their own learning more explicitly. This is usually by teaching pupils specific strategies to set goals, and monitor and evaluate their own academic development. Self-regulation means	Related to *Reading comprehension strategies*

	managing one's own motivation towards learning. The intention is often to give pupils a repertoire of strategies to choose from during learning activities	
One to one tuition	One to one tuition is where a teacher, teaching assistant or other adult gives a pupil intensive individual support. It may be undertaken outside of normal lessons as additional teaching, for example as part of extending school time or summer schools, or as a replacement for other lessons by withdrawing the pupil for extra teaching.	Distinguished from *Mentoring* and *Peer tutoring*
Oral language interventions	Oral language interventions emphasise the importance of spoken language and verbal interaction in the classroom. They are based on the idea that comprehension and reading skills benefit from explicit discussion of either the content or processes of learning, or both. Oral language approaches include targeted reading aloud and discussing books with young children and explicitly extending pupils' spoken vocabulary as well as the use of structured questioning to develop reading comprehension. All of the approaches reviewed in this section support learners' articulation of ideas and spoken expression, such as Thinking Together or Philosophy for Children. Oral language interventions therefore have some similarity to approaches based on Meta-Cognition, which make talk about learning explicit in classrooms, and to Collaborative Learning approaches, which promote pupils' talk and interaction in groups.	
Outdoor adventure learning	Outdoor adventure learning typically involves outdoor experiences, such as climbing or mountaineering; survival, ropes or assault courses; or outdoor sports, such as orienteering, sailing and canoeing. These can be organised as intensive residential courses or	Distinguished from *Sports participation* by focus on individual skills and resilience

	shorter courses run in schools or local outdoor centres.	
Parental Involvement or Engagement	Parental Involvement covers the active engagement of parents in supporting their children's learning at school. This includes programmes focused on parents and their skills (such as improving literacy or IT skills), general approaches to encourage parents to support their children to read or do mathematics, and more intensive programmes for families in crisis.	Distinguished from *Homework* often involved in *Early years intervention* and can be related to *Behaviour*
Peer tutoring	Peer tutoring includes a range of approaches in which learners work in pairs or small groups to provide each other with explicit teaching support. In cross-age tutoring, an older learner takes the tutoring role and is paired with a younger tutee or tutees. Peer assisted learning is a structured approach for mathematics and reading with sessions of 25-35 minutes two or three times a week. In reciprocal peer tutoring, learners alternate between the role of tutor and tutee. The common characteristic is that learners take on responsibility for aspects of teaching and for evaluating their success. Peer assessment involves the peer tutor providing feedback to children relating to their performance and can have different forms such as reinforcing or correcting aspects of learning.	Distinguished from *One to one tuition* where the tutors are professional educators or older volunteers and collaborative learning in that one of those involved has an explicit teaching role
Performance pay	Performance pay schemes aim to create a direct link between teacher pay or bonuses, and the performance of their class in order to incentivise better teaching and so improve outcomes. A distinction can be drawn between awards, where improved performance leads to a higher permanent salary, and payment by results, where teachers get a bonus for higher test scores. Approaches differ in how performance is measured and how closely those measures are linked to outcomes for learners. In some schemes, students' test outcomes are the sole factor used to determine performance pay awards. In others, performance judgements can also include	

	information from lesson observations or feedback from pupils, or be left to the discretion of the headteacher.	
Phonics	Phonics is an approach to teaching reading, and some aspects of writing, by developing learners' phonemic awareness. This involves the skills of hearing, identifying and using phonemes or sound patterns in English. The aim is to systematically teach learners the relationship between these sounds and the written spelling patterns, or graphemes, which represent them. Phonics emphasises the skills of decoding new words by sounding them out and combining or 'blending' the sound-spelling patterns.	
Reading comprehension strategies	Reading comprehension approaches to improving reading focus on learners' understanding of the text. They teach a range of techniques that enable pupils to comprehend the meaning of what is written, such as inferring the meaning from context, summarising or identifying key points, using graphic or semantic organisers, developing questioning strategies, and monitoring their own comprehension and identifying difficulties themselves	Related to *Metacognition and self-regulation* and may overlap when comprehension strategies are taught explicitly to include aspects of metacognition and/or self-regulation.
Reducing class size	As the size of a class or teaching group gets smaller it is suggested that the range of approaches a teacher can employ and the amount of attention each student will receive will increase, improving outcomes for pupils.	
Repeating a year	Pupils who do not reach a given standard of learning at the end of a year are required to repeat the year by joining a class of younger students the following academic year. This is also known as "grade retention", "non-promotion" or "failing a grade". For students at secondary school level, repeating a year is usually limited to the particular subject or classes that a student has not passed. Repeating a year is very rare in the UK. Repeating a year is relatively common in the	Staying at the same educational level for academic reasons, usually failure to progress sufficiently.

	USA where the No Child Left Behind Act (2002) recommended that students be required to demonstrate a set standard of achievement before progressing to the next grade level. Students can also be required to repeat a year in some European countries including Spain, France and Germany. In some countries, such as Finland, pupils can repeat a year in exceptional circumstances, but this decision is made collectively by teachers, parents and the student rather than on the basis of end of year testing.	
School uniform	Schools identify clothing considered appropriate for pupils to wear in school, and usually specify style and colour. Schools vary as to how strictly a uniform policy is enforced.	
Setting or streaming (and Within-class attainment grouping)	The terms 'setting' and 'streaming' are used to describe a variety of approaches by which pupils with similar levels of current attainment are consistently grouped together for lessons. • 'Setting' usually involves grouping pupils in a given year group into classes for specific subjects, such as mathematics and English, but not across the whole curriculum. • 'Streaming' (also known as 'tracking' in some countries) usually involves grouping pupils into classes for all or most of their lessons, so that a pupil is in the same group regardless of the subject being taught. Pupils in different sets or streams sometimes follow a different curriculum, particularly when different national tests, different examination levels or different types of academic and vocational qualifications are available. Within-class attainment grouping involves organising pupils within their usual class for specific activities or topics, such as literacy. Pupils with similar levels of current attainment are grouped together, for example, on specific tables, but all pupils are taught by their usual	These are separate entries in the Toolkit but are combined here as the search terms overlap.

	teacher and support staff, and they usually all follow the same curriculum.	
Small group tuition	Small group tuition is defined as one teacher or professional educator working with two to five pupils together in a group. This arrangement enables the teacher to focus exclusively on a small number of learners, usually in a separate classroom or working area. Intensive tuition in small groups is often provided to support lower attaining learners or those who are falling behind, but it can also be used as a more general strategy to ensure effective progress, or to teach challenging topics or skills	Distinguished from *Collaborative learning* and *One to one tuition*
Social and emotional learning	Interventions which target social and emotional learning (SEL) seek to improve pupils' interaction with others and self-management of emotions, rather than focusing directly on the academic or cognitive elements of learning. SEL interventions might focus on the ways in which students work with (and alongside) their peers, teachers, family or community. Three broad categories of SEL interventions can be identified: universal programmes which generally take place in the classroom; more specialised programmes which are targeted at students with particular social or emotional needs; and school-level approaches to developing a positive school ethos, which also aim to support greater engagement in learning.	Distinguished from *Behaviour interventions*
Sports participation	Sports participation interventions engage pupils in sports as a means to increasing educational engagement and attainment. This might be through organised after school activities or a programme organised by a local sporting club or association. Sometimes sporting activity is used as a means to encourage young people to engage in additional learning activities, such as football training at a local football club combined with study skills, ICT, literacy, or mathematics lessons.	Attainment in literacy and numeracy remain the primary outcome for this strand. This reflects the way such approaches have been evaluated in the educational research literature. This should be interpreted as the additional potential of sports participation which, of course, has value in and of itself as part of the school curriculum

		as well as health and well-being outcomes.
Summer schools	Summer schools are lessons or classes during the summer holidays or the longest school vacation. They are often designed as catch-up programmes, although some do not have an academic focus and concentrate on sports or other non-academic activities. Others have a specific aim, such as supporting pupils at the transition from primary to secondary school or preparing high-attaining pupils for university.	Distinguished from *Extending school time*
Teaching assistants	Teaching assistants (also known as TAs or classroom support assistants) are adults who support teachers in the classroom. Teaching assistants' duties can vary widely from school to school, ranging from providing administrative and classroom support to providing targeted academic support to individual pupils or small groups.	Also known as teachers' aides, auxiliary teachers, paraprofessionals and other cognate terms.

Appendix C: Example of single search strings

Platform	Date	Database	Search String	Hits	Duplicates Removed in EPPI	Total	Notes
First Search		Article First	(kw: school OR kw: early w years OR kw: kindergarten OR kw: pre-primary) and (kw: assistant w teacher+ OR kw: teaching w assistant+ OR kw: TA OR kw: classroom w assistant+ OR kw: classroom w support OR kw: support w assistant+ OR kw: paraprofessional+ OR kw: paraeducator+ OR kw: education+ w paraprofessional+ OR kw: teacher+ w aide OR kw: auxiliary w teacher+ OR kw: education* w assistant OR kw: learning w support w assistant+ OR kw: LSA) and (kw: attainment OR kw: achievement OR kw: impact OR kw: performance OR kw: intervention)				
		ECO					
		Papers First					
First Search		World Cat Dissertations	(kw: school* OR ((kw: early and kw: years)) OR kw: kindergarten OR kw: pre-primary) and (((ti: assistant and ti: teacher*)) OR ((ti: teach* and ti: assistant*)) OR ti: TA OR ((ti: classroom and ti: assistant*)) OR (ti: classroom and ti: support) OR ((ti: support and ti: assistant*)) OR ti: paraprofessional* OR ti: paraeducator* OR ((ti: education* and ti: paraprofessional)) OR ((ti: teach* and ti: aide*)) OR (ti: auxiliary and ti: teach*) OR ((ti: learning and ti: support and ti: assistant*)) OR ti: LSA OR (ti: education* and ti: assistant)) and (kw: attainment OR kw: achievement OR kw: impact OR kw: performance OR kw: intervention)				
EBSCO		BEI	AB (school OR “early years” OR kindergarten OR “pre-primary”) AND AB (“assistant teacher*” OR “teach* assistant*” OR “classroom assistan*” OR “classroom support” OR “support assistant*” OR paraprofessional OR paraeducator OR “education* paraprofessional*” OR “teach* aide*” OR “auxiliary teach*” OR “learning support assistant*” OR “education* assistant*”) AND AB (attainment OR achievement OR impact OR performance OR intervention)				
		Education Abstracts					
		Education Administration Abstracts					
		ERIC					
		PsycArticles					
		PsycINFO		400	145	255	

These search strings have been developed over time with the help of experts in systematic review, information specialists (such as library staff) and though iteration (identifying studies which should be captured then refining the search terms to capture these and other similar studies).

EEF Toolkit main data extraction [Standard]

Public version of the main data extraction tool used to code studies included in the Education Endowment Foundation's database of studies for the Toolkit.

- Section 1 What is the publication type? [Not selectable (no checkbox)]
  - Journal article [Selectable (show checkbox)]
    A report published in a peer-reviewed journal with an ISSN.
  - Dissertation or thesis [Selectable (show checkbox)]
    A report of a study in a dissertation or thesis submitted as all or part of the assessment for a higher degree.
  - Technical report [Selectable (show checkbox)]
    An unpublished report, technical report or document providing details of a research study or studies without an ISSN or ISBN. (EEF evaluation reports are classified as technical reports.)
  - Book or book chapter [Selectable (show checkbox)]
    A report of a research study published in a book or book chapter with an ISBN
  - Conference paper [Selectable (show checkbox)]
    A report of a study presented at a research conference and subsequently made more widely available.
    NB Peer-reviewed conference proceedings with an ISBN should still be classified as a conference paper.
  - Other (Please specify) [Selectable (show checkbox)]
    A report not classifiable according to the categories above (e.g. a website). Please add further details in the notes field.
- Section 2 What is the research design and which methods were used? [Not selectable (no checkbox)]
  - What is the intervention name? [Selectable (show checkbox)]
    Provide the name of the intervention, programme or approach as given in the report.
  - How is the intervention described? [Selectable (show checkbox)]
    Brief summary of the intervention as provided in the report(s). Please include the rationale for impact on learning if given.
  - What are the intervention objectives? [Selectable (show checkbox)]
    Please provide the specific objectives or aims of the intervention, programme or approach as provided in the report
  - Is there more than one treatment group? [Not selectable (no checkbox)]
    Does the research design include more than one arm or contrast so that more

than one estimate of the impact of the intervention or approach can be made from a different comparison group or version of the intervention?

- Yes (Please specify) [Selectable (show checkbox)]
  Highlight in the text (or use the info box) to describe the design and specify the other interventions or comparisons relative to the main intervention group.
- No [Selectable (show checkbox)]
- Not specified or N/A [Selectable (show checkbox)]

- How were participants assigned? [Not selectable (no checkbox)]
  How were the participants assigned or allocated to their group (i.e. treatment and control)?
  - Random (please specify) [Selectable (show checkbox)]
    Select this code where the report describes the participants' allocation to their group as random or pseudo-random (computer generated). Please highlight in the text or add information to the info box about the randomisation details.
  - Non-random, but matched [Selectable (show checkbox)]
    No randomisation, but matched at allocation prospectively to balance on attainment (or on attainment and other variables).
  - Non-random, not matched prior to treatment [Selectable (show checkbox)]
    No random allocation and not matched prior to treatment. The nature and extent of any group differences in attainment at baseline is described and then accounted for in the analysis of impact (retrospective matching).
  - Unclear [Selectable (show checkbox)]
    Please only select this code if there are no details about control and intervention allocation or if the information is so unclear as to prevent a reasonable inference.
  - Not assigned - naturally occurring sample [Not selectable (no checkbox)]
    This is where researchers take advantage of a situation where a comparison can be made between groups from changes that either are planned or have already happened which will give and estimate of the impact of the intervention or approach of interest.
    - Retrospective Quasi Experimental Design (QED) [Selectable (show checkbox)]
      Where an experiment is created from a naturally occuring situation and two groups (or more) are compared to give an estimate of impact.
    - Regression discontinuity [Selectable (show checkbox)]
      This is a type is a quasi-experimental pretest-posttest design that identifies the causal effects of an intervention or approach by assigning a cutoff or threshold above or below which an intervention is assigned (e.g. policy change where smaller classses are introduced in a district or a test is used to allocate students to additional support). By comparing results close to but either side of the threshold, it is possible to estimate effect.
- What was the level of assignment? [Not selectable (no checkbox)]
  At which level was the assignment to intervention and control group conducted?

- Individual [Selectable (show checkbox)]
  The assignment was at the level of the individual student or pupil. No account was taken of class or school. All of the individual participants were included as a single group for allocation or randomisation.
- Class [Selectable (show checkbox)]
  The class or usual teaching group of the students was the level at which the intervention or approach was allocated. Intact classes were allocated or assigned to the intervention or approach (taking no account of school).
- School - cluster [Selectable (show checkbox)]
  The school was the level of assignment and all pupils in a single school are allocated to the same grouping (i.e. a single school would not include both intervention and control).
- School - multi-site [Selectable (show checkbox)]
  The school is the level of assignment, but each school contains both intervention and control groups. The design allows a within school comparison to be made.
- Region or district [Selectable (show checkbox)]
  The region or district is the level at which the assignment is made.
- Not provided/ not available [Selectable (show checkbox)]
  A description of the level of allocation is not provided or available in the report.
- Not applicable [Selectable (show checkbox)]

- How realistic was the study? [Not selectable (no checkbox)]
  Was the intervention implemented under "real world" conditions? Factors to consider in assessing the 'ecological validity' include where the intervention took place (usual educational setting for educational approaches of this kind) and who taught or led the intervention with the pupils (e.g. did it involve usual teachers or other education professionals).
  - High ecological validity [Selectable (show checkbox)]
    Select this code where the intervention or approach seems realistic for schools or teachers to adopt.
    Any adaptations to enable the research to be conducted do not appear to affect the validity of the findings and implications for schools. Studies which take place in schools and are taught by the usual teachers or staff have high ecological validity.
  - Low ecological validity [Selectable (show checkbox)]
    Select this code where the intervention or approach does not seems realistic or practical for schools or teachers to adopt. Studies which take place in laboratory settings and are only taught by researchers have low ecological validity.
  - Unclear [Selectable (show checkbox)]
    Select this code where there are no details about where the intervention took place or who was responsible for its delivery and it is not possible to infer

sufficient details to make a judgement about the ecological validity of the study.

- Section 3 Where did the study take place? [Not selectable (no checkbox)]
  - In which country/countries was the study carried out? (Select ALL that apply) [Not selectable (no checkbox)]
    Countries which are recognised as sovereign states by the United Nations. If you think there is a country missing please ask!
    - UK (Select all that apply) [Selectable (show checkbox)]
      - England [Selectable (show checkbox)]
      - Northern Ireland [Selectable (show checkbox)]
      - Scotland [Selectable (show checkbox)]
      - Wales [Selectable (show checkbox)]
    - USA [Selectable (show checkbox)]
    - Afghanistan [Selectable (show checkbox)]
    - Albania [Selectable (show checkbox)]
    - Argentina [Selectable (show checkbox)]
    - Angola [Selectable (show checkbox)]
    - Armenia [Selectable (show checkbox)]
    - Austria [Selectable (show checkbox)]
    - Australia [Selectable (show checkbox)]
    - Azerbaijan [Selectable (show checkbox)]
    - Bahamas, The [Selectable (show checkbox)]
    - Bahrain [Selectable (show checkbox)]
    - Bangladesh [Selectable (show checkbox)]
    - Belarus [Selectable (show checkbox)]
    - Barbados [Selectable (show checkbox)]
    - Belize [Selectable (show checkbox)]
    - Belgium [Selectable (show checkbox)]
    - Benin [Selectable (show checkbox)]
    - Bhutan [Selectable (show checkbox)]
    - Bosnia and Herzegovina [Selectable (show checkbox)]
    - Botswana [Selectable (show checkbox)]
    - Brazil [Selectable (show checkbox)]
    - Bolivia [Selectable (show checkbox)]
    - Brunei Darussalam [Selectable (show checkbox)]
    - Burkina Faso [Selectable (show checkbox)]
    - Bulgaria [Selectable (show checkbox)]
    - Cabo Verde [Selectable (show checkbox)]
    - Cambodia [Selectable (show checkbox)]
    - Canada [Selectable (show checkbox)]
    - Cameroon [Selectable (show checkbox)]
    - Central African Republic [Selectable (show checkbox)]

- Chad [Selectable (show checkbox)]
- Chile [Selectable (show checkbox)]
- Colombia [Selectable (show checkbox)]
- Congo [Selectable (show checkbox)]
- Costa Rica [Selectable (show checkbox)]
- Côte d'Ivoire / Ivory Coast [Selectable (show checkbox)]
- Croatia [Selectable (show checkbox)]
- China [Selectable (show checkbox)]
  If just Hong Kong, use Hong King code only, NOT China
- Cuba [Selectable (show checkbox)]
- Cyprus [Selectable (show checkbox)]
- Denmark [Selectable (show checkbox)]
- Czech Republic [Selectable (show checkbox)]
- Dominican Republic [Selectable (show checkbox)]
- Egypt [Selectable (show checkbox)]
- Ecuador [Selectable (show checkbox)]
- El Salvador [Selectable (show checkbox)]
- Equatorial Guinea [Selectable (show checkbox)]
- Estonia [Selectable (show checkbox)]
- Eritrea [Selectable (show checkbox)]
- Ethiopia [Selectable (show checkbox)]
- Finland [Selectable (show checkbox)]
- Fiji [Selectable (show checkbox)]
- France [Selectable (show checkbox)]
- Gabon [Selectable (show checkbox)]
- Georgia [Selectable (show checkbox)]
- Gambia, The [Selectable (show checkbox)]
- Germany [Selectable (show checkbox)]
- Greece [Selectable (show checkbox)]
- Ghana [Selectable (show checkbox)]
- Guatemala [Selectable (show checkbox)]
- Grenada [Selectable (show checkbox)]
- Guinea-Bissau [Selectable (show checkbox)]
- Guinea [Selectable (show checkbox)]
- Guyana [Selectable (show checkbox)]
- Haiti [Selectable (show checkbox)]
- Honduras [Selectable (show checkbox)]
- Hong Kong (see China) [Selectable (show checkbox)]
- Hungary [Selectable (show checkbox)]
- Iceland [Selectable (show checkbox)]
- Indonesia [Selectable (show checkbox)]

- India [Selectable (show checkbox)]
- Iran [Selectable (show checkbox)]
- Iraq [Selectable (show checkbox)]
- Ireland [Selectable (show checkbox)]
- Italy [Selectable (show checkbox)]
- Israel [Selectable (show checkbox)]
- Jamaica [Selectable (show checkbox)]
- Japan [Selectable (show checkbox)]
- Jordan [Selectable (show checkbox)]
- Kenya [Selectable (show checkbox)]
- Kazakhstan [Selectable (show checkbox)]
- Kuwait [Selectable (show checkbox)]
- Kiribati [Selectable (show checkbox)]
- Lao (or Laos) [Selectable (show checkbox)]
  Lao People's Democratic Republic
- Kyrgyzstan [Selectable (show checkbox)]
- Latvia [Selectable (show checkbox)]
- Lebanon [Selectable (show checkbox)]
- Liberia [Selectable (show checkbox)]
- Lesotho [Selectable (show checkbox)]
- Libya [Selectable (show checkbox)]
- Liechtenstein [Selectable (show checkbox)]
- Luxembourg [Selectable (show checkbox)]
- Lithuania [Selectable (show checkbox)]
- Madagascar [Selectable (show checkbox)]
- Macedonia [Selectable (show checkbox)]
- Malaysia [Selectable (show checkbox)]
- Malawi [Selectable (show checkbox)]
- Mali [Selectable (show checkbox)]
- Maldives [Selectable (show checkbox)]
- Malta [Selectable (show checkbox)]
- Marshall Islands [Selectable (show checkbox)]
- Mauritania [Selectable (show checkbox)]
- Mauritius [Selectable (show checkbox)]
- Micronesia [Selectable (show checkbox)]
- Mexico [Selectable (show checkbox)]
- Moldova [Selectable (show checkbox)]
- Mongolia [Selectable (show checkbox)]
- Mozambique [Selectable (show checkbox)]
- Namibia [Selectable (show checkbox)]
- Myanmar (Burma) [Selectable (show checkbox)]

- Nepal [Selectable (show checkbox)]
- Nauru [Selectable (show checkbox)]
- The Netherlands [Selectable (show checkbox)]
- New Zealand [Selectable (show checkbox)]
- Nicaragua [Selectable (show checkbox)]
- Nigeria [Selectable (show checkbox)]
- Niger [Selectable (show checkbox)]
- Pakistan [Selectable (show checkbox)]
- Norway [Selectable (show checkbox)]
- Palau [Selectable (show checkbox)]
- Panama [Selectable (show checkbox)]
- Papua New Guinea [Selectable (show checkbox)]
- Peru [Selectable (show checkbox)]
- Philippines [Selectable (show checkbox)]
- Poland [Selectable (show checkbox)]
- Puerto Rico (US dependency) [Selectable (show checkbox)]
- Portugal [Selectable (show checkbox)]
- Qatar [Selectable (show checkbox)]
- Romania [Selectable (show checkbox)]
- Rwanda [Selectable (show checkbox)]
- Russia [Selectable (show checkbox)]
- Saint Kitts and Nevis [Selectable (show checkbox)]
- Saint Lucia [Selectable (show checkbox)]
- Saint Vincent and the Grenadines [Selectable (show checkbox)]
- San Marino [Selectable (show checkbox)]
- Samoa [Selectable (show checkbox)]
- Saudi Arabia [Selectable (show checkbox)]
- São Tomé and Príncipe [Selectable (show checkbox)]
- Serbia [Selectable (show checkbox)]
- Senegal [Selectable (show checkbox)]
- Seychelles [Selectable (show checkbox)]
- Sierra Leone [Selectable (show checkbox)]
- Slovakia [Selectable (show checkbox)]
- Singapore [Selectable (show checkbox)]
- Slovenia [Selectable (show checkbox)]
- Solomon Islands [Selectable (show checkbox)]
- South Africa [Selectable (show checkbox)]
- Somalia [Selectable (show checkbox)]
- South Korea / Republic of Korea [Selectable (show checkbox)]
- South Sudan [Selectable (show checkbox)]
- Sri Lanka [Selectable (show checkbox)]

- Spain [Selectable (show checkbox)]
- Sudan [Selectable (show checkbox)]
- Suriname [Selectable (show checkbox)]
- Swaziland / Eswatini [Selectable (show checkbox)]
- Sweden [Selectable (show checkbox)]
- Switzerland [Selectable (show checkbox)]
- Taiwan [Selectable (show checkbox)]
- Syria [Selectable (show checkbox)]
- Tanzania [Selectable (show checkbox)]
- Tajikistan [Selectable (show checkbox)]
- Thailand [Selectable (show checkbox)]
- Timor-Leste [Selectable (show checkbox)]
- Togo [Selectable (show checkbox)]
- Tonga [Selectable (show checkbox)]
- Tunisia [Selectable (show checkbox)]
- Trinidad and Tobago [Selectable (show checkbox)]
- Turkey [Selectable (show checkbox)]
- Turkmenistan [Selectable (show checkbox)]
- Tuvalu [Selectable (show checkbox)]
- Ukraine [Selectable (show checkbox)]
- Uganda [Selectable (show checkbox)]
- United Arab Emirates [Selectable (show checkbox)]
- Uruguay [Selectable (show checkbox)]
- Uzbekistan [Selectable (show checkbox)]
- Vanuatu [Selectable (show checkbox)]
- Venezuela [Selectable (show checkbox)]
- Vietnam [Selectable (show checkbox)]
- West Indies (Use for Caribbean colonial dependencies) [Selectable (show checkbox)]
  Cayman Islands (United Kingdom)
  Anguilla (United Kingdom)
  Antigua and Barbuda
  Aruba (Netherlands)
  Bonaire (Netherlands)
  British Virgin Islands (United Kingdom)
  Curaçao (Netherlands)
  Guadeloupe (France)
  Martinique (France)
  Montserrat (United Kingdom)
  Nueva Esparta (Venezuela)
  Saba (Netherlands)
  Saint Barthélemy (France)

Saint-Martin (France)
Sint Eustatius (Netherlands)
Sint Maarten (Netherlands)
United States Virgin Islands (United States)
Federal Dependencies of Venezuela (Venezuela)
Turks and Caicos Islands (United Kingdom)

- Yemen [Selectable (show checkbox)]
- Zambia [Selectable (show checkbox)]
- Zimbabwe [Selectable (show checkbox)]

- Is there more specific information about the location? [Not selectable (no checkbox)]
  Further information on where the study took part (e.g. city, district, urban, suburban, rural etc.) as provided by the study.
  - Specific to the location or place [Selectable (show checkbox)]
    Information about the specific place where the research was undertaken (e.g. name of the city, state, city or region)
  - Information about the type of location [Selectable (show checkbox)]
    Information about what kind of location (e.g. urban, rural, suburban).
  - No information provided [Selectable (show checkbox)]
    Please use this code if there is no further information about the specific location (place name) or the type of location (e.g. urban/ rural).
- What is the educational setting (Select ALL that apply) [Not selectable (no checkbox)]
  What is the type of educational setting that the students attend which is the focus of the intervention or approach?
  - Nursery school/pre-school [Selectable (show checkbox)]
    A separate nursery school or pre-school setting or a nursery or early years class in a primary school.
    The focus is on the type of setting or educational provision.
  - Primary/elementary school [Selectable (show checkbox)]
    A school for children of normal school age (depending on the jurisdiction). The focus is on the type of school or setting. Pupils will typically be between the ages of 5 and 11.
  - Middle school [Selectable (show checkbox)]
    An intermediate school provided in some jurisdictions for pupils between their primary (or elementary) and secondary educational stages.
  - Secondary/High school [Selectable (show checkbox)]
    A school for older pupils, after primary or elementary education (and after middle school where provided). Pupils will usually be between the ages of 11 and 18.
  - Residential/boarding school [Selectable (show checkbox)]
    A school where pupils reside as well as study; boarding either by week or over a term.

        - Independent/private school [Selectable (show checkbox)]
        - Home [Selectable (show checkbox)]
        - Further education/junior or community college [Selectable (show checkbox)]
          A formal educational setting for older secondary pupils. Students will usually be 16 or older, but still studying for school-level, vocational or professional qualifications (i.e. not higher education or leading to a Bachelor's degree)
        - Other educational setting (please specify) [Selectable (show checkbox)]
          An educational setting which cannot be classified under one of the other definitions. Please provide details of the educational setting as given in the study (e.g. field centre, museum classroom, concert or rehearsal hall, public theatre, workplace training, etc.)
        - Outdoor adventure setting [Selectable (show checkbox)]
          Educational activities taking place outdoors such as Outward Bound courses, sailing and kayaking or canoeing, camping, climbing or courses based at an outdoor education centre.
          All studies classified under the Toolkit strand 'Outdoor adventure learning' should be included.
          Field studies centres where the activities focus solely on school subjects like Geography or Biology should not be included (please use 'Other' for these and specify the type of setting).
        - No information provided [Selectable (show checkbox)]
- Section 4 What is the sample of the study? [Not selectable (no checkbox)]
    - What is the overall sample analysed? [Selectable (show checkbox)]
      What is the total number of participants in the data analysed (both intervention and control/comparison)? Please add additional details in the notes.
    - What is the gender of the students? [Not selectable (no checkbox)]
      Please indicate the gender of the total sample.
        - Female only [Selectable (show checkbox)]
        - Male only [Selectable (show checkbox)]
        - Mixed gender [Selectable (show checkbox)]
          Provide the percentage or number of female pupils in the study. Please highlight the section or add details of where this can be found in the report.
        - No information provided [Selectable (show checkbox)]
    - What is the age of the students? (Select ALL that apply) [Not selectable (no checkbox)]
      Please provide additional information if available (e.g. grade level(s), mean age, or mean and standard deviation).
        - 3 [Selectable (show checkbox)]
        - 4 [Selectable (show checkbox)]
        - 5 [Selectable (show checkbox)]
        - 6 [Selectable (show checkbox)]
        - 7 [Selectable (show checkbox)]

        - 8 [Selectable (show checkbox)]
        - 9 [Selectable (show checkbox)]
        - 10 [Selectable (show checkbox)]
        - 11 [Selectable (show checkbox)]
        - 12 [Selectable (show checkbox)]
        - 13 [Selectable (show checkbox)]
        - 14 [Selectable (show checkbox)]
        - 15 [Selectable (show checkbox)]
        - 16 [Selectable (show checkbox)]
        - 17 [Selectable (show checkbox)]
        - 18 [Selectable (show checkbox)]
        - No information provided [Selectable (show checkbox)]
    - What is the proportion of low SES/FSM students in the sample? [Not selectable (no checkbox)]
      What proportion of the students in the study are receiving free school meals (FSM) or reduced price lunches or are identified as being from a low socio-economic status? If possible, record this as a percentage. Please highlight or add further details as reported in the study.
        - FSM or low SES student percentage [Selectable (show checkbox)]
          Please add the percentage of pupils in the sample who are are receiving free school meals (FSM) or reduced price lunches or are identified as being from a low socio-economic status background.
        - Further information about FSM or SES in the study sample. [Selectable (show checkbox)]
          Please highlight any details provided in the study about the socio-economic status of the students involved in the research (such as eligibility for free or reduced price school meals or lunches).
        - No SES/FSM information provided [Selectable (show checkbox)]
          Select this option if there is no information about the socio-economic status of the students involved in the research (such as eligibility for free or reduced price school meals or lunches).
- Section 5 What was involved in the intervention? [Not selectable (no checkbox)]
  Details about the intervention, approach or policy being evaluated.
    - What type of organisation was responsible for providing the intervention? [Not selectable (no checkbox)]
      Please indicate what kind of organisation was responsible for the provision or management and organisation of the intervention?
        - School or group of schools [Selectable (show checkbox)]
        - Charity or voluntary organisation [Selectable (show checkbox)]
        - University/ researcher design [Selectable (show checkbox)]
        - Local education authority or district [Selectable (show checkbox)]
          Local education authority or district (government or public funding)

    - Private or commercial company [Selectable (show checkbox)]
    - Other (please provide details) [Selectable (show checkbox)]
- Was training for the intervention provided? [Not selectable (no checkbox)]
  Was training provided to the delivery team as part of the preparation and support for the intervention? If so, who provided it?
    - Yes (Please specify) [Selectable (show checkbox)]
      Please highlight the text or add details to the info box as provided in the report.
    - No [Selectable (show checkbox)]
    - Unclear/ Not specified [Selectable (show checkbox)]
- Who is the focus of the intervention? (Select ALL that apply) [Not selectable (no checkbox)]
  Who is the main focus of the intervention study? Although the interest of the Toolkit is on student outcomes, the focus of behavioural change may be on others in educational settings, such as teachers or parents. NB All interventions must report outcomes on student's attainment.
    - Students [Selectable (show checkbox)]
      The main focus of the intervention is on the behaviours, interactions or activities of the students or pupils. Others may be involved (such as in training to deliver or implement a new approach), but the main aim is to change students' activities, behaviours and interactions to improve educational outcomes.
    - Teachers [Selectable (show checkbox)]
      The main focus of the intervention is on the teachers and their behaviours, interactions and activities. Although the final outcome may be to improve students' attainment, the focus and study aims focus on the teachers as a clear or explicit part of the rationale.
    - Teaching assistants [Selectable (show checkbox)]
      The focus of the intervention includes teaching assistants or teacher's aides (and/or other para-professionals) and their behaviours, interactions and activities. Although the final outcome may be to improve students' attainment, the focus and study aims involve teaching assistants as part of the process.
    - Other education practitioners [Selectable (show checkbox)]
    - Non-teaching staff [Selectable (show checkbox)]
      The main focus of the intervention is on the non-teaching staff in schools and their behaviours, interactions and activities. This includes all staff who would not normally have a teaching role (e.g. administrative staff, lunchtime supervisors, facilities management etc.). Although the final outcome may be to improve students' attainment, the focus and study aims include the non-teaching staff as part of the rationale.
    - Senior management [Selectable (show checkbox)]
      The main focus of the intervention is on the senior management in schools (e.g. headteachers, deputy head teachers, heads of department) and their

behaviours, interactions and activities. Although the final outcome may be to improve students' attainment, the focus and study aims include the senior management as part of the rationale.

    - Parents [Selectable (show checkbox)]
      Parents or carers of students in the educational settings involved are involved because of their parental or caring responsibilities.
    - Other (Please specify) [Selectable (show checkbox)]
- What is the intervention teaching approach? (Select ALL that apply) [Not selectable (no checkbox)]
  What was the main teaching or learning approach used for an intervention session?
    - Large group/class teaching (+6) [Selectable (show checkbox)]
      A large group (more than 6 students) with a teacher or supporter of the intervention, typically in a classroom setting.
    - Small group/intensive support (3-5) [Selectable (show checkbox)]
      Intensive small group provision by a teacher, teaching assistant or other supporter of the intervention in small group setting (3 - 5 participants in a group), sometimes in a separate teaching space or classroom.
    - Paired learning [Selectable (show checkbox)]
      Two pupils either working together, or peer teaching each other
    - One to one [Selectable (show checkbox)]
      One to one instruction where the teacher is not a peer, but a teacher, teaching assistant, volunteer or other education professional.
    - Student alone (self-administered) [Selectable (show checkbox)]
      Pupils or students working through study materials independently and/or unsupervised.
    - Other (Explain in notes) [Selectable (show checkbox)]
- Were any of the following involved in the intervention or approach? [Not selectable (no checkbox)]
    - Digital technology [Not selectable (no checkbox)]
      The main approach depends on the use of digital technology (e.g. tablets, laptops, software, internet) by pupils or teachers (e.g. interactive whiteboards).
        - Yes [Selectable (show checkbox)]
        - No [Selectable (show checkbox)]
    - Parents or community volunteers [Not selectable (no checkbox)]
      Parents or community volunteers working with their children (or other pupils).
        - Yes [Selectable (show checkbox)]
        - No [Selectable (show checkbox)]

- When did the intervention take place? (Select ALL that apply) [Not selectable (no checkbox)]
  When was the intervention delivered?
  - During regular school hours [Selectable (show checkbox)]
    The intervention or approach takes place completely or mainly during regular school hours.
  - Before/after school [Selectable (show checkbox)]
    The intervention or approach takes place completely or mainly before or immediately after normal school hours. This should mainly apply to activities taking place on school or normal educational settings.
  - Evenings and/or weekends [Selectable (show checkbox)]
    Where the intervention or approach takes place during evenings or weekends. Activities which take place immediately after school and at school (or in the same educational setting) should not be included.
  - Summer/ holiday period [Selectable (show checkbox)]
    Where the educational activity takes place as additional time in what would normally be a holiday period (e.g. summer holidays or other vacation times).
  - Other (please specify) [Selectable (show checkbox)]
  - Unclear/ not specified [Selectable (show checkbox)]
    Use this code where there are no details provided of when the intervention was delivered and where the information provided does not allow a reasonable inference to be made about timing.
    The usual inference for most interventions where the timing is not specified will be 'During regular school hours'. If this inference cannot reasonably be made please indicate in the notes the details in the report which produce the ambiguity or lack of clarity.
- Who was responsible for the teaching at the point of delivery? (Select ALL that apply) [Not selectable (no checkbox)]
  Please provide details (e.g. staff involved, training level provided, number/ proportions of staff).
  This should focus on the experience of pupils, rather than any initial training and support.
  - Research staff [Selectable (show checkbox)]
    Select this code where the intervention or approach was delivered largely or exclusively by researchers or the research team.
  - Class teachers [Selectable (show checkbox)]
    Select this code when the intervention or approach was taught or delivered by professional teachers as part of their usual teaching or wider professional activity.
  - Teaching assistants [Selectable (show checkbox)]
    Select this code where the majority of the teaching or delivery of the intervention is undertaken by teaching assistants (or teacher's aides, para-professionals, auxiliary teachers, nursery nurses in early years settings and

other cognate terms). These will be staff usually employed by a school, but without a full teaching qualification.

- Other school staff [Selectable (show checkbox)]
  Staff employed by the school, but neither teachers nor teaching assistants (or those in similar paid roles). It includes administrative staff, lunch-time supervisors, facilities staff.
- External teachers [Selectable (show checkbox)]
  Teachers or other professional educational staff hired or employed by the research team or the delivery organisation.
- Parents/carers [Selectable (show checkbox)]
  Parents or carers whose main relationship with the intervention is through their parental or caring responsibilities. This includes where parents working with their own children, or working with other children in the school or educational setting that their own children attend.
- Lay persons/volunteers [Selectable (show checkbox)]
  Adults (over 18 years) involved as volunteers or undertaking unpaid work who provide the majority of the support to pupils or lead in the delivery of the intervention to students.
- Peers [Selectable (show checkbox)]
  Other students or pupils at the same school or educational setting as the intervention group; or at another local school (e.g. secondary students tutoring pupils at their own or their peers' primary schools). Peers will normally be of similar age and socio-economic or cultural background. University students tutoring primary school pupils would not be classified as 'peers'.
- Digital technology [Selectable (show checkbox)]
  Include digital technology where the technology has a role in the educational activity, such as where automated feedback or marking is provided, or where it provides an explicit teaching role (intelligent tutoring or the use of explanatory videos) or where differentiated activities are offered or allocated automatically to learners. Incidental use of technology which is usually involved in the normal teaching and learning activities of the intervention group should not be included as this has already been recorded.
- Unclear/not specified [Selectable (show checkbox)]
  Use this code where there are no details provided of who or how the intervention was delivered or where the information provided does not allow a reasonable inference to be made.

- What was the duration of the intervention? (Please add to info box and specify units) [Selectable (show checkbox)]
  Duration of the intervention or approach (from beginning to end). Please specify units (e.g. months, weeks, days). This may differ from the duration of the research project or evaluation which could involved pre- and post-testing periods.

- What was the frequency of the intervention? [Selectable (show checkbox)]
  What is the frequency of the intervention (as delivered)? e.g. daily, twice weekly, weekly monthly.
- What is the length of intervention sessions? [Selectable (show checkbox)]
  What is the length in minutes of a typical session?
- Are implementation details and/or fidelity details provided? [Not selectable (no checkbox)]
  Are details provided about how successfully the intervention was implemented or taken up? Please indicate what type of information by selecting the appropriate checkbox and higlighting relevant text in the report.
  - Qualitative [Selectable (show checkbox)]
    Please select if qualitative details about the intervention or approach are provided, such as describing and issues or challenges about implementation, or comments on the training and/ or implementation process.
  - Quantitative [Selectable (show checkbox)]
    Please select if quantitative details about implementation are provided, such as number of schools or teachers trained, or number of sessions attended.
  - No implementation details provided. [Selectable (show checkbox)]
    No details about the implementation process are provided.
- Are the costs reported? [Not selectable (no checkbox)]
  Are there any financial costs or details reported?
  - Yes (Please add details) [Selectable (show checkbox)]
    If this option is selected, please add details as provide in the report(s).
  - No [Selectable (show checkbox)]
- Who undertook the outcome evaluation? [Not selectable (no checkbox)]
  Here we are interested in how independent the evalution was.
  - The developer [Selectable (show checkbox)]
    This is the usual option and should be selected unless the information is unclear or confusing. This is where the researcher or developer evaluated their own programme or approach.
  - A different organization paid by developer [Selectable (show checkbox)]
    The development team is different from the evaluation team but it is commissioned directly by the developer or researcher who developed the intervention approaches.
  - An organization commissioned independently to evaluate [Selectable (show checkbox)]
    The research team is different from the evaluation team and commissioned independently (e.g. EEF reports).
  - Unclear/not stated [Selectable (show checkbox)]
    There is insufficient information about the status of the evaluation research to indicate or infer how independent the evaluation is.

        - Is this an EEF evaluation? [Selectable (show checkbox)]
          If the evaluation was funded by the Education Endowment Foundation please select.
- Section 6 What kind of primary outcomes are reported? [Not selectable (no checkbox)]
    - What kind of tests were used? (Select ALL that apply) [Not selectable (no checkbox)]
      What type(s) of test(s) were used to measure the intervention outcomes on learning at pupil/student level?
        - Standardised test (Please specify) [Selectable (show checkbox)]
          A standardised test is administered and scored in a consistent way. The properties of the test are established through piloting on a group to determine the mean and spread of the scores for a particular target group. Standardised tests are usually named and the properties published.
          Please add the name of the test(s) used, a brief description and any details reported.
        - Researcher developed test (Please add details) [Selectable (show checkbox)]
          A test developed or designed for a specific research project. Please add any details as provided in the report(s).
        - School-developed test (Please add details) [Selectable (show checkbox)]
          A test or examination developed and used by a school or schools involved in the research as part of their usual assessment approach. Please add any details as provided in the report(s).
        - National test or examination (Please specify) [Selectable (show checkbox)]
          A test or examination used in regional or national evaluations of student and school performance. These may be optional or compulsory, but are organised and/ or administered by the regional or national education administration in a particular jurisdiction.
        - International tests (Please specify) [Selectable (show checkbox)]
          Tests used for international comparisons of student performance (e.g. PISA, TIMMS, PIRLS etc.). Please specify the name of the test.
    - Curriculum subjects tested (Select ALL that apply) [Not selectable (no checkbox)]
      If the outcomes relate to the subjects of the school curriculum outcomes, record which subjects are included.
        - Literacy (first language) [Not selectable (no checkbox)]
          Aspects of literacy including speaking and listening, reading and writing. Include study of literature when this is first language study.
            - Reading comprehension [Selectable (show checkbox)]
              This may include aspects such as main idea identification and passage comprehension. When a test provides different outcomes, e.g. TOWRE (Test of Word Reading Efficacy) provides word attack, word identification, & passage comprehension, choose passage comprehension as main outcome.

        - Decoding/phonics [Selectable (show checkbox)]
          These measures gave a focus on recognizing letters and making the correct sounds associated with the letters or letter combinations. They made be referred to as phonological or phonemic awareness.
        - Spelling [Selectable (show checkbox)]
          Where the focus is on the correct spelling of words.
        - Reading other [Selectable (show checkbox)]
          e.g. phonics, reading fluency, vocabulary comprehension (receptive vocabulary)
          When a test provides different outcomes, e.g. TOWRE (Test of Word Reading Efficacy) provides word attack, word identification, & passage comprehension, choose passage comprehension as main outcome
        - Speaking and listening/Oral language [Selectable (show checkbox)]
          Speaking and listening or oral language and communication outcomes, including vocabulary use (productive spoken vocabulary).
        - Writing [Selectable (show checkbox)]
          A test of written language including quality, quantity and written vocabulary (range).
    - Mathematics [Selectable (show checkbox)]
      All aspects of mathematics including number and numerical operations, shape and space (geometry), algebra, data-handling etc.
    - Science [Selectable (show checkbox)]
      All general science subjects including physics, chemistry, biology as well as specific subjects such as ecology or astronomy.
    - Social studies [Selectable (show checkbox)]
      Either integrated social studies courses or programmes or separate curriculum areas of social studies (e.g. history, geography, civics, sociology, economics or anthropology).
    - Arts [Selectable (show checkbox)]
      Expressive and performing arts, including music, art, drama, drawing, painting, sculpture and the decorative arts.
    - Languages [Selectable (show checkbox)]
      Where the aim is to develop communicative or literacy capability in a language other than the first language or usual language of instruction in the school.
    - Other curriculum test [Selectable (show checkbox)]
      Please provide a description of the outcome as reported where it is a test of a school curriculum subject not included in the categories above (e.g. music, art, classics).

- In addition to the primary educational attainment outcome, are there other outcomes reported? [Not selectable (no checkbox)]

    - Yes [Selectable (show checkbox)]
    - No [Selectable (show checkbox)]
- If yes, which other outcomes are reported? [Not selectable (no checkbox)]
    - Cognitive outcomes measured (Please specify) [Selectable (show checkbox)]
      If non-curricular cognitive outcomes are measured, please indicate and specify the outcomes (e.g. reasoning, memory, intelligence, etc.). Include the name of the test where possible (e.g. Raven's Matrices, Stanford–Binet Intelligence Scales etc.).
    - Other types of student outcomes (Please specify) [Selectable (show checkbox)]
      e.g. attendance, measures of behaviour, health status, non-cognitive attitudes/dispositions, etc. as assessed through a test or a survey.
    - Other participants (i.e. not students) outcomes (Please specify) [Selectable (show checkbox)]
      If outcomes are measured and reported for other participants involved in the research (such as teachers or parents), please note which participants and which outcomes have been measured e.g. parental participation.

Appendix E: Examples of EEF Toolkit strand specific data extraction

Peer Tutoring v.04 Aug 2018 [Not selectable (no checkbox)]
Peer tutoring includes a range of approaches in which learners work in pairs or small groups to provide each other with explicit teaching support. In cross-age tutoring, an older learner takes the tutoring role and is paired with a younger tutee or tutees. Peer-assisted learning is a structured approach for mathematics and reading with sessions of 25-35 minutes two or three times a week. In reciprocal peer tutoring, learners alternate between the role of tutor and tutee. The common characteristic is that learners take on responsibility for aspects of teaching and for evaluating their success. Peer assessment involves the peer tutor providing feedback to children relating to their performance and can have different forms such as reinforcing or correcting aspects of learning.

Peers are defined as other students or pupils at the same school or educational setting as the intervention group; or at another local school (e.g. secondary students tutoring pupils at their own or their peers' primary schools). Peers will normally be of similar age and socio-economic or cultural background.

University students tutoring primary school pupils would not usually be classified as 'peers'.

- Who were the tutors? [Not selectable (no checkbox)]
  Describe the type of tutors involved.
    - Same-age as tutees (if yes, select one from drop down) [Not selectable (no checkbox)]

        - matched to same level of attainment [Selectable (show checkbox)]
          e.g. high attainers with high attainers
        - matched to different level of attainment [Selectable (show checkbox)]
          e.g. high attainers with lower attainers
        - not matched on attainment [Selectable (show checkbox)]
      - Cross-age (i.e. different school year from tutees) [Not selectable (no checkbox)]
        - Matched to same relative level of attainment [Selectable (show checkbox)]
          e.g. older high attainers with younger hihg attainers
        - Not matched on attainment [Selectable (show checkbox)]
      - Were the tutors... [Not selectable (no checkbox)]
        - From the same school [Selectable (show checkbox)]
          Tutors and tutees are from a different school or institution.
        - From a different school [Selectable (show checkbox)]
          Tutors and tutees are from the different schools or institutions.
- Was the teaching role alternating/reciprocal? [Not selectable (no checkbox)]
  - Yes [Selectable (show checkbox)]
  - No [Selectable (show checkbox)]
- Tutee attainment level [Not selectable (no checkbox)]
  What is the level of academic attainment of the tutees?
  - Low attaining [Selectable (show checkbox)]
  - Average [Selectable (show checkbox)]
  - High attaining [Selectable (show checkbox)]
  - Mixed [Selectable (show checkbox)]
  - Not mentioned [Selectable (show checkbox)]
- Was digital technology involved? [Not selectable (no checkbox)]
  Please select if the peer tutoring involve digital technology, either shared (such as using an iPad together) or for communication (such as video conference or chat program)?
  - Yes [Selectable (show checkbox)]
  - No [Selectable (show checkbox)]
- Was an incentive provided for the tutors and/or tutees? [Not selectable (no checkbox)]
  Was there a reward or an incentive given as part of the intervention?
  - Yes [Selectable (show checkbox)]
  - No [Selectable (show checkbox)]

Feedback v.02 Oct 2018 [Not selectable (no checkbox)]
Feedback is information given to the learner and/or the teacher about the learner's performance relative to learning goals. It should aim towards (and be capable of producing) improvement in students' learning. Feedback redirects or refocuses either the teacher's or the learner's actions to achieve a goal, by aligning effort and activity with an outcome. It can be about the learning activity itself, about the process of activity, about the student's management of their learning or self-regulation or (the least effective) about them as individuals. This feedback can be verbal, written, or can be given through tests or via digital technology. It can come from a teacher or someone taking a teaching role, or from peers or technology

- What was the source of the feedback? [Not selectable (no checkbox)]
  - Teacher [Selectable (show checkbox)]
  - Teaching assistant [Selectable (show checkbox)]
  - Volunteer [Selectable (show checkbox)]
  - Parent(s) or other relatives [Selectable (show checkbox)]
    Parent(s), carer(s) or guardian(s). Also use for other family members (such as grandparents or siblings).
  - Researcher [Selectable (show checkbox)]
  - Peer (same age/ class) [Selectable (show checkbox)]
  - Peer (group) [Selectable (show checkbox)]
    Feedback from more than one same age pupil (e.g. when feedback is formalised in collaborative learning)
  - Peer (older) [Selectable (show checkbox)]
  - Digital or automated [Selectable (show checkbox)]
    Feedback from a computer or other digital device (e.g. mobile phone, website or program) where there is some automation involved.
  - Other non-human [Selectable (show checkbox)]
    Such as from a worked example or where answers are checked after the task has been completed.
  - Self [Selectable (show checkbox)]
    Only use this code when checking or self-assessment is strategic and self-regulated (such as applying a checking algorithm or mnemonic).
  - Other (please specify) [Selectable (show checkbox)]
    Please add notes about the source for this category, as described in the study.
- Who was the feedback directed to? [Not selectable (no checkbox)]
  This will almost always be to pupils, but may be to the teacher. If to the teacher, then there should be some explicit model of further feedback to change subsequent pupil behaviours or performance.
  - Individual pupil [Selectable (show checkbox)]
  - General (group or class) [Selectable (show checkbox)]
    Where the feedback is not specific to an individual learner, please indicate.
  - Teacher [Selectable (show checkbox)]
    Only select this code when this is explicitly part of the model of feedback in the research study.
- What form did the feedback take? (Select one) [Selectable (show checkbox)]
  This focuses on how the feedback was communicated. Choose the main feedback approach if there is more than one.
  - Spoken verbal [Selectable (show checkbox)]
    Feedback provided in spoken form, this includes audio recorded comments.
  - Non-verbal [Selectable (show checkbox)]
    Where feedback was communicated physically other that with words, such a through body language, gesture or other non-verbal means, such as extended wait time.
  - Written verbal [Selectable (show checkbox)]
    Where written comments are provided, either handwritten or digitally.
  - Written, non-verbal [Selectable (show checkbox)]
    Such as tick or check marks, or with symbols or icons (this includes marked tests or test results).

- When did the feedback happen? (Select one) [Not selectable (no checkbox)]
  Choose the option which best describes the feedback timing.
  - Prior to the task [Selectable (show checkbox)]
    Sometimes described as 'feedforward', this is where pupils are primed with information before undertaking a task (e.g. students complete test and get positive, negative results regardless of actual score and then their performance on a following test is measured).
  - During the task [Selectable (show checkbox)]
    Where the feedback is contemporaneous with the task or part of the task.
  - Immediate [Selectable (show checkbox)]
    Where the feedback was provided immediately or shortly after the activity was completed (such as at the end of the task, or later the same day.
  - Delayed (short) [Selectable (show checkbox)]
    Where the feedback occurred more than one day and up to a week after the task or activity.
  - Delayed (long) [Selectable (show checkbox)]
    Where the feedback occurred more that a week after the task of activity.
- What kind of feedback was provided? [Not selectable (no checkbox)]
  - About the outcome [Selectable (show checkbox)]
    Where the feedback was about the outcome or completed task (e.g. correct or incorrect).
    - Correct [Selectable (show checkbox)]
      Where feedback was about the correct answers or responses
    - Incorrect [Selectable (show checkbox)]
      Where feedback focussed on the incorrect answers or responses.
  - About the process of the task [Selectable (show checkbox)]
    Where the feedback is about how the task or activity is currently being, or should be, undertaken (process rather than outcome).
  - About the learner's strategies or approach [Selectable (show checkbox)]
    Where the feedback was to support the learner's own regulation or control of what they were doing (i.e. metacognition and/or self-regulation) often in the form or prompts or cues.
  - About the person [Selectable (show checkbox)]
    Feedback directed at the individual or self, such as good boy or clever girl.
- What was the emotional tone of the feedback? [Not selectable (no checkbox)]
  Select the most appropriate description for the emotional tone of the feedback. Select more than one only where this is explictly part of the design, otherwise select the best overall description, based on how it is described in the study.
  - Positive [Selectable (show checkbox)]
  - Neutral [Selectable (show checkbox)]
    Where the feedback was designed or perceived to be neutral in tone.
  - Negative [Not selectable (no checkbox)]
    This is where the feedback is deliberately designed to be discouraging. It should not be used for feedback about incorrect responses or results.

Teaching assistants v.01 Aug 2018 [Not selectable (no checkbox)]
Teaching assistants (also known as TAs or classroom support assistants) are adults who support teachers in the classroom. Teaching assistants' duties can vary widely from school to school, ranging from providing administrative and classroom support to

providing targeted academic support to individual pupils or small groups.
Cognate terms:
support staff; adult support staff; teaching assistants; associate staff; classroom assistants; classroom support assistant; auxiliary teachers; teachers' aide; education paraprofessional; nursery nurse (in early years' settings)

- How are the teaching assistants described? [Not selectable (no checkbox)]
  - Teaching or classroom assistant [Selectable (show checkbox)]
    Teaching assistant is commonly used in England, whereas classroom assistant is more usual in Scotland.
  - Higher level teaching assistant [Selectable (show checkbox)]
    A higher level teaching assistant (HLTA) can take additional responsibilities such as teaching classes on their own, covering planned absences and allowing teachers time to plan and mark.
  - Teacher's aide [Selectable (show checkbox)]
    This description is commonly used in the USA.
  - Paraprofessional or paraeducator [Selectable (show checkbox)]
    This is a more generic term covering a number of paraprofessional educator roles, sometimes known as a para, para-pro, paraeducator.
  - Educational or instructional assistant [Selectable (show checkbox)]
  - Pupil support worker or student support worker [Selectable (show checkbox)]
  - Other (please specify) [Selectable (show checkbox)]
    Please add details of how the teaching assistants are described.
- What is the teaching assistants' role? [Not selectable (no checkbox)]
  - Curriculum instruction (please specify) [Selectable (show checkbox)]
    This is where the teaching assistant has a teaching or instructional role when working with an individual pupil or group of pupils. Please add details in the notes or highlight the curriculum focus (e.g. maths, reading, writing).
  - Behaviour support [Selectable (show checkbox)]
    Support rather than instruction. This is where the teaching assistant role is primarily behavioural or motivational to keep pupils engaged or on task or to provide support to access the tasks or activities.
  - Assessment [Selectable (show checkbox)]
    This is where the teaching assistant's role is primarily to undertake assessment of the pupil(s) such as through testing or observation.
  - General classroom support [Selectable (show checkbox)]
    This is where the teaching assistant role is to provide help with classroom activities such as managing groups of pupils or involvement in non-teaching activities (preparation, marking etc.).
  - Not specified/ unclear [Selectable (show checkbox)]
    This is where, although teaching assistants were involved there is little or no information about their role.
- How many pupils is the teaching assistant working with? [Not selectable (no checkbox)]
  This is to indicate the number of pupils the teaching assistant has responsibility for at any particular time.
  - One to one [Selectable (show checkbox)]
    This is where the teaching assistant works with pupils individually, either providing instruction or support.

- Small group (2-4 pupils) [Selectable (show checkbox)]
  This is where the teaching assistant is working with a small group of between two and four pupils, together, at the same time.
- Large group (5 - 12 pupils) [Selectable (show checkbox)]
  This is where the teaching assistant is working with a larger group of pupils altogether, but not the whole class.
- Whole class [Selectable (show checkbox)]
  This is where the teaching assistant is working with the whole class, all together at the same time.

EEF Toolkit effect size data extraction v 1.0 June 2018 [Standard]

Public version of the effect size data extraction tool used to code studies included in the Education Endowment Foundation's database of studies for the Toolkit.

- Section 1 What are the details of the study design? [Not selectable (no checkbox)]
  - What was the study design? [Not selectable (no checkbox)]
    What type of study design is used for the evaluation of impact?
    - Individual RCT [Selectable (show checkbox)]
      where individual participants are the unit of randomisation and no provision is made for clustering in the design
    - Cluster RCT [Selectable (show checkbox)]
      where school or class is the unit of randomisation - i.e. all pupils in same school are in same group (between school/class) and where the class or school variance can be assigned to either intervention or control
    - Multisite RCT [Selectable (show checkbox)]
      where both control and intervention pupils may be in the same class or school (within school/class) so that in the analysis the school or class level variance should be shared between intervention and control groups
    - Prospective QED [Selectable (show checkbox)]
      quasi-experimental design – allocation/matching but no randomisation
    - Retrospective QED [Selectable (show checkbox)]
      natural experiment with matching/ equivalence is reached through design/analysis
    - Interrupted time series QED [Selectable (show checkbox)]
      Same group will be treated as control and comparison e.g. ABAB
    - Regression Discontinuity with randomisation [Selectable (show checkbox)]
      Prospective regression discontinuity design where participants around the cut off are randomised to treatment or control.
    - Regression Discontinuity - not randomised [Selectable (show checkbox)]
      RD with non-random allocation (prospective matching to create equivalence)
    - Regression Continuity - naturally occurring [Selectable (show checkbox)]
      Regression Continuity design naturally occurring - retrospective matching.

      Exploits or manipulates a naturally occurring discontinuity to explore the causal effect of an educational intervention or approach. Regression

discontinuity designs elicits the causal effects of interventions by assigning a cut off or threshold above or below which an intervention is assigned

- What is the number of schools involved in the study? [Not selectable (no checkbox)]
  - What is the number of schools involved in the intervention group(s)? [Selectable (show checkbox)]
    Please provide the number of schools involved in the intervention or versions of the intervention. Please only enter numeric data in the info box.
  - What is the number of schools involved in the control or comparison group? [Selectable (show checkbox)]
    Please provide the number of schools involved in the control group. Please only enter numeric data in the info box.
  - What is the total number of schools involved? [Selectable (show checkbox)]
    Please record the total number of schools involved in the study. This will be the sum of intervention and control schools in a cluster randomised trial, but in a multisite trial, where there are control and intervention pupils in each school, it may be the same as for intervention/ control. Please only enter numeric data in the info box.
  - Not provided/ unclear / not applicable [Selectable (show checkbox)]
    Please indicate if the number of schools involved in not provided, is unclear, or not applicable (such as in a Outdoor Education study).
- What is the number of classes involved? [Selectable (show checkbox)]
  - What is the total number of classes involved in the intervention group? [Selectable (show checkbox)]
    Please provide the number of classes involved in the intervention or versions of the intervention. Please only enter numeric data in the info box.
  - What is the total number of classes involved in the control or comparison group? [Selectable (show checkbox)]
    Please provide the number of classes involved in the control group. Please only enter numeric data in the info box.
  - What is the total number of classes involved? [Selectable (show checkbox)]
    Please record the total number of classes involved in the study. Please only enter numeric data in the info box.
  - Not provided/ unclear / not applicable [Selectable (show checkbox)]
    Please indicate if the number of classes involved in not provided, is unclear, or not applicable (such as in a Outdoor Education study).
- Are details of randomisation provided? [Not selectable (no checkbox)]
  - Yes [Selectable (show checkbox)]
    Please select if details are provided about how any randomisation was undertaken. Please highlight the relevant section of the study where possible.
  - Not applicable [Selectable (show checkbox)]
    Please select if the study is not described as a randomised design (e.g. Quasi-experimental or naturally occuring experiment).

        - No / Unclear [Selectable (show checkbox)]
          Please select if the study is described as randomised but no details are provided or these details are unclear. If the details are unclear, please highlight the relevant section of the report.
- Section 2 How is the sample described? [Not selectable (no checkbox)]
  Information about the sample size, groups and comparability.
    - What is the sample size for the intervention group? [Selectable (show checkbox)]
      Record the initial or assigned sample size for the treatment group in the notes. Please enter numeric data only in the info box. This should be either the main counterfactual comparison of the intervention or approach for the Toolkit from this study, or the first reported.
    - What is the sample size for the control group? [Selectable (show checkbox)]
      Record the initial or assigned sample size for the control group in the notes. Please enter numeric data only in the info box.
    - *What is the sample size for the second intervention group? [Selectable (show checkbox)]
      *Record the initial or assigned sample size for a second or alternative treatment group in the notes (*if there is one). This should be an equally valid comparison of the intervention or approach for the Toolkit as the first intervention group reported above. Please enter numeric data only in the info box.*
    - *What is the sample size for the third intervention group? [Selectable (show checkbox)]
      *Record the initial or assigned sample size for a third or different treatment group in the notes (*if there is one). This should be an equally valid comparison of the intervention or approach for the Toolkit as the other intervention groups reported above. Please enter numeric data only in the info box.*
    - Does the study report any group differences at baseline? [Not selectable (no checkbox)]
      Is there quantitative information about the similarity of treatment and control groups at the beginning of the intervention?
        - Yes [Selectable (show checkbox)]
          Please select if there is information provided about how comparable the intervention and control groups are at the beginning of the study in terms of the analysis. Please also highlight the relevant section of the text where this is possible.
        - No/Unclear [Selectable (show checkbox)]
          Please select this option if there is no information about the baseline comparability of the groups or if this is unclear. If there is information, but it is unclear, please highlight the relevant section of the study, where this is possible.
    - Is comparability taken into account in the analysis? [Not selectable (no checkbox)]

Are covariates in treatment and control groups assessed, and, if unbalanced, controlled in adjusted analysis?
  - Yes [Selectable (show checkbox)]
  - No [Selectable (show checkbox)]
  - Unclear or details not provided [Selectable (show checkbox)]
- Is attrition or drop out reported? [Not selectable (no checkbox)]
  If the sample recruited differs from the sample analysed, are the reasons for this reported? Please include details of attrition or drop-out or any pupils excluded from the analysis.
  - Yes [Selectable (show checkbox)]
  - No [Selectable (show checkbox)]
  - Unclear (please add notes) [Selectable (show checkbox)]
    Please check this option if the amount of atttrition is unclear. Please also add notes about attrition if there is information about different groups or outcomes.
- What is the attrition in the treatment group? [Selectable (show checkbox)]
  Number of drop-outs in the intervention group as a percentage of the n of the intervention group. Please enter numeric data only in the info box
- Are the variables used for comparability reported? [Not selectable (no checkbox)]
  Does the study state which variables are used to assess the comparability of the treatment and control groups?
  - Yes [Selectable (show checkbox)]
  - No [Selectable (show checkbox)]
  - N/A [Selectable (show checkbox)]
  - If yes, which variables are used for comparability? [Not selectable (no checkbox)]
    Select the variables considered in assessment of similarity e.g. prior attainment, age, gender, SES, special educational needs, ethnicity.
    - Educational attainment [Selectable (show checkbox)]
      A measure of either direct (e.g. reading comprehension) or indirect (reasoning) educational performance or capability.
    - Gender [Selectable (show checkbox)]
    - Socio-economic status [Selectable (show checkbox)]
    - Special educational needs [Selectable (show checkbox)]
    - Other (please specify) [Selectable (show checkbox)]
- What is the total or overall percentage attrition? [Selectable (show checkbox)]
  Please report the percentage of drop-outs or overall attrition in the whole sample. This is the number of drop-outs divided by the initial sample x 100.Or you can calculate as the (initial sample minus the analysed sample) divided by the initial sample time 100. ((N-n)/N) x 100. Please add the % sign (e.g. 15.8%). For more information see:
  https://ies.ed.gov/ncee/wwc/Docs/OnlineTraining/wwc_training_m2.pdf

- Is clustering accounted for in the analysis? [Not selectable (no checkbox)]
  Does analysis take account of clustering? e.g. regression with school or cluster or MLM (multi-level modelling) or HLM (hierarchical linear modelling)?
  - Yes [Selectable (show checkbox)]
  - No [Selectable (show checkbox)]
  - Unclear [Selectable (show checkbox)]

- Section 3 Outcome details [Not selectable (no checkbox)]
  - Outcomes [Not selectable (no checkbox)]
    - Are descriptive statistics reported for the primary outcome? [Not selectable (no checkbox)]
      - Yes [Selectable (show checkbox)]
        - If yes, please add for the intervention* group [Not selectable (no checkbox)]
          *Descriptive statistics for the intervention group. *If there is more than one intervention group please add this below.*
          - Number (n) [Selectable (show checkbox)]
            What is the number for the intervention group in the data analysed for this outcome? Add numeric data only to the info box.
          - Pre-test mean [Selectable (show checkbox)]
            Please record the pre-test mean (if provided) for the intervention group for this outcome. Add numeric data only to the info box.
          - Pre-test standard deviation [Selectable (show checkbox)]
            Please record the pre-test standard deviation (if provided) for the intervention group for this outcome. Add numeric data only to the info box.
          - Post-test mean [Selectable (show checkbox)]
            Please report the post-test mean for this outcome for the intervention group (if provided) for this outcome. Add numeric data only to the info box.
          - Post test standard deviation [Selectable (show checkbox)]
            Please record the post-test standard deviation for the intervention group for this outcome (if provided). Add numeric data only to the info box.
          - Gain score mean (if reported) [Selectable (show checkbox)]
            Please add the gain score (pre-test to post test) mean for the intervention group. Add numeric data only to the info box.
          - Gain score standard deviation (if reported) [Selectable (show checkbox)]
            Please add the gain score (pre-test to post test) standard deviation for the intervention group. Add numeric data only to the info box.
          - Any other information? [Selectable (show checkbox)]
            Please add any other statistical information reported about this

outcome for the intervention group (e.g. standard error (SE)), or use to add notes about the numeric data in the categories above.

- If yes please add for the control group [Not selectable (no checkbox)]
  Descriptive statistics for the intervention group
  - Number (n) [Selectable (show checkbox)]
    What is the number for the control group in the data analysed for this outcome? Add numeric data only to the info box.
  - Pre-test mean [Selectable (show checkbox)]
    Please record the pre-test mean (if provided) for the control group for this outcome. Add numeric data only to the info box.
  - Pre-test standard deviation [Selectable (show checkbox)]
    Please record the pre-test standard deviation (if provided) for the control group for this outcome. Add numeric data only to the info box.
  - Post-test mean [Selectable (show checkbox)]
    Please report the post-test mean for this outcome for the control group (if provided) for this outcome.
  - Post test standard deviation [Selectable (show checkbox)]
    Please record the post-test standard deviation for the control group for this outcome (if provided).
  - Gain score mean (if reported) [Selectable (show checkbox)]
    Add numeric data only to the info box.
  - Gain score standard deviation (if reported) [Selectable (show checkbox)]
    Add numeric data only to the info box.
  - Any other information? [Selectable (show checkbox)]
    Please add any other statistical information reported about this outcome for the intervention group (e.g. standard error (SE)).
- If yes, please add for a second intervention* group (if needed) [Not selectable (no checkbox)]
  Descriptive statistics for a second intervention group, if needed.
  - Number (n) [Selectable (show checkbox)]
    What is the number for the intervention group in the data analysed for this outcome? Add numeric data only to the info box.
  - Pre-test mean [Selectable (show checkbox)]
    Please record the pre-test mean (if provided) for the intervention group for this outcome. Add numeric data only to the info box.
  - Pre-test standard deviation [Selectable (show checkbox)]
    Please record the pre-test standard deviation (if provided) for the intervention group for this outcome. Add numeric data only to the info box.
  - Post-test mean [Selectable (show checkbox)]
    Please report the post-test mean for this outcome for the

intervention group (if provided) for this outcome. Add numeric data only to the info box.

- Post test standard deviation [Selectable (show checkbox)]
  Please record the post-test standard deviation for the intervention group for this outcome (if provided). Add numeric data only to the info box.
- Gain score mean (if reported) [Selectable (show checkbox)]
  Please add the gain score (pre-test to post test) mean for a second intervention group (if needed). Add numeric data only to the info box.
- Gain score standard deviation (if reported) [Selectable (show checkbox)]
  Please add the gain score (pre-test to post test) standard deviation for a second intervention group (if need). Add numeric data only to the info box.
- Any other information? [Selectable (show checkbox)]
  Please add any other statistical information reported about this outcome for the intervention group (e.g. standard error (SE)), or use to add notes about the numeric data in the categories above.
- If needed, please add for the control group [Not selectable (no checkbox)]
  Descriptive statistics for the second control group (if needed and if different from the primary outcome control)
  - Number (n) [Selectable (show checkbox)]
    What is the number for the control group in the data analysed for this outcome? Add numeric data only to the info box.
  - Pre-test mean [Selectable (show checkbox)]
    Please record the pre-test mean (if provided) for the control group for this outcome. Add numeric data only to the info box.
  - Pre-test standard deviation [Selectable (show checkbox)]
    Please record the pre-test standard deviation (if provided) for the control group for this outcome. Add numeric data only to the info box.
  - Post-test mean [Selectable (show checkbox)]
    Please report the post-test mean for this outcome for the control group (if provided) for this outcome.
  - Post test standard deviation [Selectable (show checkbox)]
    Please record the post-test standard deviation for the control group for this outcome (if provided).
  - Gain score mean (if reported) [Selectable (show checkbox)]
    Please add the gain score (pre-test to post test) mean for this group (if need). Add numeric data only to the info box.

- Gain score standard deviation (if reported) [Selectable (show checkbox)]
  Please add the gain score (pre-test to post test) standard deviation for this group (if need). Add numeric data only to the info box.
- Any other information? [Selectable (show checkbox)]
  Please add any other statistical information reported about this outcome for the intervention group (e.g. standard error (SE)).

- If yes, please add for a third intervention* group (if needed) [Not selectable (no checkbox)]
  Descriptive statistics for a third intervention group, if needed.
  - Number (n) [Selectable (show checkbox)]
    What is the number for the intervention group in the data analysed for this outcome? Add numeric data only to the info box.
  - Pre-test mean [Selectable (show checkbox)]
    Please record the pre-test mean (if provided) for the intervention group for this outcome. Add numeric data only to the info box.
  - Pre-test standard deviation [Selectable (show checkbox)]
    Please record the pre-test standard deviation (if provided) for the intervention group for this outcome. Add numeric data only to the info box.
  - Post-test mean [Selectable (show checkbox)]
    Please report the post-test mean for this outcome for the intervention group (if provided) for this outcome. Add numeric data only to the info box.
  - Post test standard deviation [Selectable (show checkbox)]
    Please record the post-test standard deviation for the intervention group for this outcome (if provided). Add numeric data only to the info box.
  - Gain score mean (if reported) [Selectable (show checkbox)]
    Please report the gain score (pre-test to post-test) mean for this outcome for a third intervention group (if needed) for this outcome. Add numeric data only to the info box.
  - Gain score standard deviation (if reported) [Selectable (show checkbox)]
    Add numeric data only to the info box.
  - Any other information? [Selectable (show checkbox)]
    Please add any other statistical information reported about this outcome for the intervention group (e.g. standard error (SE)), or use to add notes about the numeric data in the categories above.
  - If needed please add for a control group [Not selectable (no checkbox)]

Descriptive statistics for a third control group (if needed and if different from the primary outcome control)

- Number (n) [Selectable (show checkbox)]
  What is the number for the control group in the data analysed for this outcome? Add numeric data only to the info box.
- Pre-test mean [Selectable (show checkbox)]
  Please record the pre-test mean (if provided) for the control group for this outcome. Add numeric data only to the info box.
- Pre-test standard deviation [Selectable (show checkbox)]
  Please record the pre-test standard deviation (if provided) for the control group for this outcome. Add numeric data only to the info box.
- Post-test mean [Selectable (show checkbox)]
  Please report the post-test mean for this outcome for the control group (if provided) for this outcome.
- Post test standard deviation [Selectable (show checkbox)]
  Please record the post-test standard deviation for the control group for this outcome (if provided).
- Gain score mean (if reported) [Selectable (show checkbox)]
  Add numeric data only to the info box.
- Gain score standard deviation (if reported) [Selectable (show checkbox)]
  Add numeric data only to the info box.
- Any other information? [Selectable (show checkbox)]
  Please add any other statistical information reported about this outcome for the intervention group (e.g. standard error (SE)).

- No [Selectable (show checkbox)]

- Is there follow up data? [Not selectable (no checkbox)]
  Please provide details of any assessment to measure long lasting effects (e.g. delayed post-test or long term follow up)
  - Yes [Selectable (show checkbox)]
  - No [Selectable (show checkbox)]
- Primary outcome [Outcome]
  Please indicate the primary outcome and enter additional data using the 'Outcomes' box.
  The primary outcome should be the outcome most relevant to the Toolkit strand(s) in terms of educational impact, such as standardised tests of reading or mathematics (for literacy or mathematics interventions) or national test or examination results. See handbook and supporting resources for further information.
- Secondary outcome(s) [Outcome]
  Please add secondary outcomes in this section where they represent a fair test of the impact of the evaluation at post test. This should not include delayed or

follow up tests, or outcomes used to check the specificity of impact (e.g. a maths test use to control for intervention effect in a literacy intervention) or checking for transfer outcomes.

- SES/FSM outcome [Outcome]
  If a separate effect is reported for low socio-economic status or free or reduced price school meals pupils please add here.

- DO NOT USE [Not selectable (no checkbox)]
  Please do not mark this section. This section is completed in the 'Outcome specific code' screen.
  - Outcome classification [Not selectable (no checkbox)]
    Outcome classifications for meta-analysis and meta-regressions. Please select all that apply
    - Sample: High achievers (select one from this group) [Outcome classification code]
      Classification of the students in the sample in relation to their level of academic attainment. Those described as high attainers or high 'ability'; usually those in the top half or the top third of the distribution (depending on classifications).
    - Sample: Low achievers [Outcome classification code]
      Classification of the students in the sample in relation to their level of academic attainment. Those described as low attainers or low 'ability'; usually those in the bottom half or the bottom third of the distribution (depending on classifications).
    - Sample: Average [Outcome classification code]
      Classification of the students in the sample in relation to their level of academic attainment. Those described as performing at or around average attainment or of average 'ability'; usually those in the middle quartiles (depending on classifications).
    - Sample: Exceptional [Outcome classification code]
      Students described as gifted and talented or of exceptional 'ability'. Usually those in the top 10 per cent of the distribution.
    - Sample: All [Outcome classification code]
      Analysis applied to normal or typical sample of pupils. The whole range of attainment or 'ability' for the educational setting was included in the intervention.
    - Test type: Standardised test (select one from this group) [Outcome classification code]
      A standardised test is administered and scored in a consistent way. The properties of the test are established through piloting on a group to determine the mean and spread of the scores for a particular target group. Standardised tests are usually named and the properties published.
    - Test type: Researcher developed test [Outcome classification code]
      A test developed or designed for a specific research project

- Test type: School-developed test [Outcome classification code]
  A test or examination developed and used by a school or schools involved in the research as part of their usual assessment approach.
- Test type: National test [Outcome classification code]
  A test or examination used in regional or national evaluations of students and school performance. These may be optional or compulsory, but are organised and/or administered by the regional or national administration in a particular jurisdiction.
- Test type: International tests [Outcome classification code]
  Tests used for international comparisons of student performance (e.g. PISA, TIMMS, PIRLS, etc.)
- Analysis: Post-test unadjusted (select one from this group) [Outcome classification code]
  A simple comparison of the differences between control and intervention groups using only the post-test data, usually from an older randomised controlled trial (RCT) or where baseline equivalence has been established.
- Analysis: Post-test adjusted for baseline attainment [Outcome classification code]
  A post-test comparison where a measure of educational attainment at pre-test is controlled for in the analysis of the impact of the intervention or approach e.g. ANCOVA, OLS regression.
- Analysis: Post-test adjusted for baseline attainment AND clustering [Outcome classification code]
  A post-test comparison where a measure of educational attainment at pre-test is controlled for in the analysis of the impact of the intervention or approach and where the estimate is adjusted for clustering at class or school level (e.g. ANCOVA, MLM, OLS regression).
- Analysis: Pre-post gain [Outcome classification code]
  Outcome assessment based on the difference between an individual's pre-test and post test scores and the range of these difference (gain score or pre-post analysis).
- Toolkit: Arts participation (select at least one Toolkit strand) [Outcome classification code]
  Arts participation is defined as involvement in artistic and creative activities, such as dance, drama, music, painting, or sculpture. It can occur either as part of the curriculum or as extra-curricular activity. Participation may be organised as regular weekly or monthly activities, or more intensive programmes such as summer schools or residential courses. Whilst these activities have educational value in themselves, this Toolkit entry focuses on the benefits of arts participation for core academic attainment.
- Toolkit: Aspiration interventions [Outcome classification code]
  By aspirations we mean the things children and young people hope to

achieve for themselves in the future. To meet their aspirations about careers, university, and further education, pupils often require good educational outcomes. Raising aspirations is therefore often believed to incentivise improved attainment.

- Toolkit: Behaviour interventions [Outcome classification code]
  Behaviour interventions seek to improve attainment by reducing challenging behaviour. This entry covers interventions aimed at reducing a variety of behaviours, from low-level disruption to general anti-social activities, aggression, violence, bullying, and substance abuse. The interventions themselves can be split into three broad categories:
  1. Approaches to developing a positive school ethos or improving discipline across the whole school which also aim to support greater engagement in learning.
  2. Universal programmes which seek to improve behaviour and generally take place in the classroom.
  3. More specialised programmes which are targeted at students with specific behavioural issues.
- Toolkit: Block scheduling [Outcome classification code]
  Block scheduling is an approach to school timetabling in secondary schools. It typically means that pupils have fewer classes (4-5) per day, for a longer period of time (70-90 minutes). The three main types of block schedules found in the research are:
  4x4 block scheduling: 4 blocks of extended (80–90 minute) classes each day, covering the same 4 subjects each day. Students take 4 subjects over 1 term, and 4 different subjects in the following term. A/B block scheduling: 3 or 4 blocks of extended (70–90 minute) classes each day, covering the same 3 or 4 subjects on alternating days. Students take 6 or 8 subjects each term. Hybrid: a hybrid of traditional models and 3/4-class-per-day approaches. Students have 5 classes per day, of between 60 and 90 minutes.
- Toolkit: Built environment [Outcome classification code]
  Changing the physical conditions or built environment of the learning setting, either by moving to a new school building or seeking to improve the structure, air quality, noise, light, or temperature of an existing building or classroom.
- Toolkit: Collaborative learning [Outcome classification code]
  A collaborative (or cooperative) learning approach involve pupils working together on activities or learning tasks in a group small enough for everyone to participate on a collective task that has been clearly assigned. Pupils in the group may work on separate tasks contributing to a common overall outcome, or work together on a shared task.
  Some collaborative learning approaches put mixed ability teams or groups to work in competition with each other in order to drive more effective

collaboration. There is a very wide range of approaches to collaborative and cooperative learning involving different kinds of organisation and tasks. Peer tutoring can also be considered as a type of collaborative learning, but in the Toolkit it is reviewed it as a separate topic.

- Toolkit: Digital technology [Outcome classification code]
  The use of digital technologies to support learning. Approaches in this area are very varied, but a simple split can be made between:
  Programmes for students, where learners use technology in problem solving or more open-ended learning, and
  Technology for teachers such as interactive whiteboards or learning platforms which may be used by the teachers, or where the technology may provide instruction more directly.
- Toolkit: Early years intervention [Outcome classification code]
  Early years or early childhood interventions are approaches that aim to ensure that young children have educationally based pre-school or nursery experiences which prepare for school and academic success, usually through additional nursery or pre-school provision. Many of the researched programmes and approaches focus on disadvantaged children. Some also offer parental support. The research summarised here looks at general or multi-component programmes and approaches.
- Toolkit: Extending school time [Outcome classification code]
  This summary focuses on extending core teaching and learning time in schools and the use of targeted before and after school programmes. Other approaches to increasing learning time are included in other sections of the Toolkit, such as Homework, Early years intervention and Summer schools.
  The research focuses on three main approaches to extending teaching and learning time in schools:
  extending the length of the school year;
  extending the length of the school day; and
  providing additional time for targeted groups of pupils, particularly disadvantaged or low-attaining pupils, either before or after school.
- Toolkit: Feedback [Outcome classification code]
  Feedback is information given to the learner and/or the teacher about the learner's performance relative to learning goals. It should aim towards (and be capable of producing) improvement in students' learning. Feedback redirects or refocuses either the teacher's or the learner's actions to achieve a goal, by aligning effort and activity with an outcome. It can be about the learning activity itself, about the process of activity, about the student's management of their learning or self-regulation or (the least effective) about them as individuals. This feedback can be verbal, written, or can be given through tests or via digital technology. It can come from a teacher or someone taking a teaching role, or from peers.

- Toolkit: Homework [Outcome classification code]
  Homework refers to tasks given to pupils by their teachers to be completed outside of usual lessons. Common homework activities in primary schools tend to be reading or practising spelling and number facts, but may also include more extended activities to develop inquiry skills or more directed and focused work such as revision for tests which is more similar to homework set in secondary schools. Other homework activities may include reading or preparing for work to be done in class, or practising and completing tasks or activities already taught or started in lessons, as well as revision for exams.
- Toolkit: Individualised instruction [Outcome classification code]
  Individualised instruction involves different tasks for each learner and support at the individual level. It is based on the idea that all learners have different needs, and that therefore an approach that is personally tailored — particularly in terms of the activities that pupils undertake and the pace at which they progress through the curriculum — will be more effective. Various models of individualised instruction have been tried over the years in education, particularly in subjects like mathematics where pupils can have individual sets of activities which they complete, often largely independently. More recently, digital technologies have been employed to facilitate individual activities and feedback.
- Toolkit: Learning styles [Outcome classification code]
  The idea underpinning learning styles is that individuals all have a particular approach to or style of learning. The theory is that learning will therefore be more effective or more efficient if pupils are taught using the specific style or approach that has been identified as their learning 'style'. For example, pupils categorised as having a 'listening' learning style, could be taught more through storytelling and discussion and less through traditional written exercises.
- Toolkit: Mastery learning [Outcome classification code]
  Mastery learning breaks subject matter and learning content into units with clearly specified objectives which are pursued until they are achieved. Learners work through each block of content in a series of sequential steps.
  Students must demonstrate a high level of success on tests, typically at about the 80% level, before progressing to new content. Mastery learning can be contrasted with other approaches which require pupils to move through the curriculum at a pre-determined pace. Teachers seek to avoid unnecessary repetition by regularly assessing knowledge and skills. Those who do not reach the required level are provided with additional tuition, peer support, small group discussions, or homework so that they can reach the expected level.

- Toolkit: Metacognition and self-regulation [Outcome classification code]
  Metacognition and self-regulation approaches aim to help pupils think about their own learning more explicitly, often by teaching them specific strategies for planning, monitoring and evaluating their learning. Interventions are usually designed to give pupils a repertoire of strategies to choose from and the skills to select the most suitable strategy for a given learning task.
  Self-regulated learning can be broken into three essential components: cognition - the mental process involved in knowing, understanding, and learning;
  metacognition - often defined as 'learning to learn'; and
  motivation - willingness to engage our metacognitive and cognitive skills.
- Toolkit: Mentoring [Outcome classification code]
  Mentoring in education involves pairing young people with an older peer or volunteer, who acts as a positive role model. In general, mentoring aims to build confidence, develop resilience and character, or raise aspirations, rather than to deliver specific academic skills or knowledge. Mentors typically build relationships with young people by meeting with them one to one for about an hour a week over a sustained period, either during school, at the end of the school day, or at weekends.
  Activities vary between different mentoring programmes, sometimes including direct academic support with homework or other school tasks. For programmes focused primarily on direct academic support see One to one tuition and Peer tutoring.
  Mentoring has increasingly been offered to young people who are deemed to be hard to reach or at risk of educational failure or exclusion.
- Toolkit: One to one tuition [Outcome classification code]
  One to one tuition involves a teacher, teaching assistant or other adult giving a pupil intensive individual support. It may happen outside of normal lessons as additional teaching – for example as part of Extending school time or a Summer school – or as a replacement for other lessons.
- Toolkit: Oral language interventions [Outcome classification code]
  Oral language interventions emphasise the importance of spoken language and verbal interaction in the classroom.
  They are based on the idea that comprehension and reading skills benefit from explicit discussion of either the content or processes of learning, or both. Oral language approaches include:
  Targeted reading aloud and discussing books with young children
  Explicitly extending pupils' spoken vocabulary
  The use of structured questioning to develop reading comprehension. All of the approaches reviewed in this section support learners' articulation of ideas and spoken expression, such as Thinking Together or Philosophy for Children. Oral language interventions therefore have some similarity to

approaches based on metacognition, which make talk about learning explicit in classrooms, and to Collaborative Learning approaches, which promote pupils' talk and interaction in groups.

- Toolkit: Outdoor adventure learning [Outcome classification code]
  Outdoor adventure learning typically involves outdoor experiences, such as climbing or mountaineering; survival, ropes or assault courses; or outdoor sports, such as orienteering, sailing and canoeing. These can be organised as intensive residential courses or shorter courses run in schools or local outdoor centers.
  Adventure education usually involves collaborative learning experiences with a high level of physical (and often emotional) challenge. Practical problem-solving, explicit reflection and discussion of thinking and emotion (see also Metacognition and self-regulation) may also be involved.
  Adventure learning interventions typically do not include a formal academic component, so this summary does not include forest schools or field trips.
- Toolkit: Parental engagement [Outcome classification code]
  We define parental engagement as the involvement of parents in supporting their children's academic learning. It includes:
  1. approaches and programmes which aim to develop parental skills such as literacy or IT skills;
  2. general approaches which encourage parents to support their children with, for example reading or homework;
  3. the involvement of parents in their children's learning activities; and
  4. more intensive programmes for families in crisis.
- Toolkit: Peer Tutoring [Outcome classification code]
  Peer tutoring includes a range of approaches in which learners work in pairs or small groups to provide each other with explicit teaching support. In cross-age tutoring, an older learner takes the tutoring role and is paired with a younger tutee or tutees. Peer-assisted learning is a structured approach for mathematics and reading with sessions of 25-35 minutes two or three times a week. In reciprocal peer tutoring, learners alternate between the role of tutor and tutee. The common characteristic is that learners take on responsibility for aspects of teaching and for evaluating their success. Peer assessment involves the peer tutor providing feedback to children relating to their performance and can have different forms such as reinforcing or correcting aspects of learning.
  Peers are defined as other students or pupils at the same school or educational setting as the intervention group; or at another local school (e.g. secondary students tutoring pupils at their own or their peers' primary schools). Peers will normally be of similar age and socio-economic or cultural background.

University students tutoring primary school pupils would not usually be classified as 'peers'.

- Toolkit: Performance pay [Outcome classification code]
  Performance pay schemes aim to create a direct link between teacher pay or bonuses, and the performance of their class in order to incentivise better teaching and so improve outcomes. A distinction can be drawn between awards, where improved performance leads to a higher permanent salary, and payment by results, where teachers get a bonus for higher test scores. Approaches differ in how performance is measured and how closely those measures are linked to outcomes for learners. In some schemes, students' test outcomes are the sole factor used to determine performance pay awards. In others, performance judgements can also include information from lesson observations or feedback from pupils, or be left to the discretion of the headteacher.
- Toolkit: Phonics [Outcome classification code]
  Phonics is an approach to teaching reading, and some aspects of writing, by developing learners' phonemic awareness. This involves the skills of hearing, identifying and using phonemes or sound patterns in English. The aim is to systematically teach learners the relationship between these sounds and the written spelling patterns, or graphemes, which represent them. Phonics emphasises the skills of decoding new words by sounding them out and combining or 'blending' the sound-spelling patterns.
- Toolkit: Reading comprehension strategies [Outcome classification code]
  Reading comprehension strategies focus on the learners' understanding of written text. Pupils are taught a range of techniques which enable them to comprehend the meaning of what they read. These can include: inferring meaning from context; summarising or identifying key points; using graphic or semantic organisers; developing questioning strategies; and monitoring their own comprehension and identifying difficulties themselves (see also 'Metacognition and self-regulation').
- Toolkit: Reducing class size [Outcome classification code]
  As the size of a class or teaching group gets smaller it is suggested that the range of approaches a teacher can employ and the amount of attention each student will receive will increase, thereby improving outcomes for pupils.
- Toolkit: Repeating a year [Outcome classification code]
  Pupils who do not reach a given standard of learning at the end of a year are required to repeat the year by joining a class of younger students the following academic year. This is also known as "grade retention", "non-promotion" or "failing a grade". For students at secondary school level, repeating a year is usually limited to the particular subject or classes that a student has not passed.
  Repeating a year is very rare in the UK but is relatively common in the

USA where the No Child Left Behind Act (2002) recommended that students be required to demonstrate a set standard of achievement before progressing to the next grade level. Students can also be required to repeat a year in some European countries including Spain, France and Germany. In some countries, such as Finland, pupils can repeat a year in exceptional circumstances, but this decision is made collectively by teachers, parents and the student rather than on the basis of end of year testing.

- Toolkit: School uniform [Outcome classification code]
  Schools identify clothing considered appropriate for pupils to wear in school, and usually specify the style and colour. Schools vary as to how strictly a uniform policy is enforced.
- Toolkit: Setting or streaming [Outcome classification code]
  Pupils with similar levels of current attainment are grouped together either for specific lessons on a regular basis (setting or regrouping), or as a whole class (streaming or tracking). The assumption is that it will be possible to teach more effectively or more efficiently with a narrower range of attainment in a class.
- Toolkit: Small Group Tuition [Outcome classification code]
  Small group tuition is defined as one teacher or professional educator working with two, three, four, or five pupils. This arrangement enables the teacher to focus exclusively on a small number of learners, usually on their own in a separate classroom or working area. Intensive tuition in small groups is often provided to support lower attaining learners or those who are falling behind, but it can also be used as a more general strategy to ensure effective progress, or to teach challenging topics or skills.
- Toolkit: Social and emotional learning [Outcome classification code]
  Interventions which target social and emotional learning (SEL) seek to improve attainment by improving the social and emotional dimensions of learning, as opposed to focusing directly on the academic or cognitive elements of learning. SEL interventions might focus on the ways in which students work with (and alongside) their peers, teachers, family or community. Three broad categories of SEL interventions can be identified:
  1. Universal programmes which generally take place in the classroom;
  2. More specialised programmes which are targeted at students with particular social or emotional problems;
  3. School-level approaches to developing a positive school ethos which also aim to support greater engagement in learning.
- Toolkit: Sports participation [Outcome classification code]
  Sports participation interventions engage pupils in sports as a means to increasing educational engagement and attainment. This might be through after-school activities or a programme organised by a local sporting club or association. Sometimes sporting activity is used to encourage young

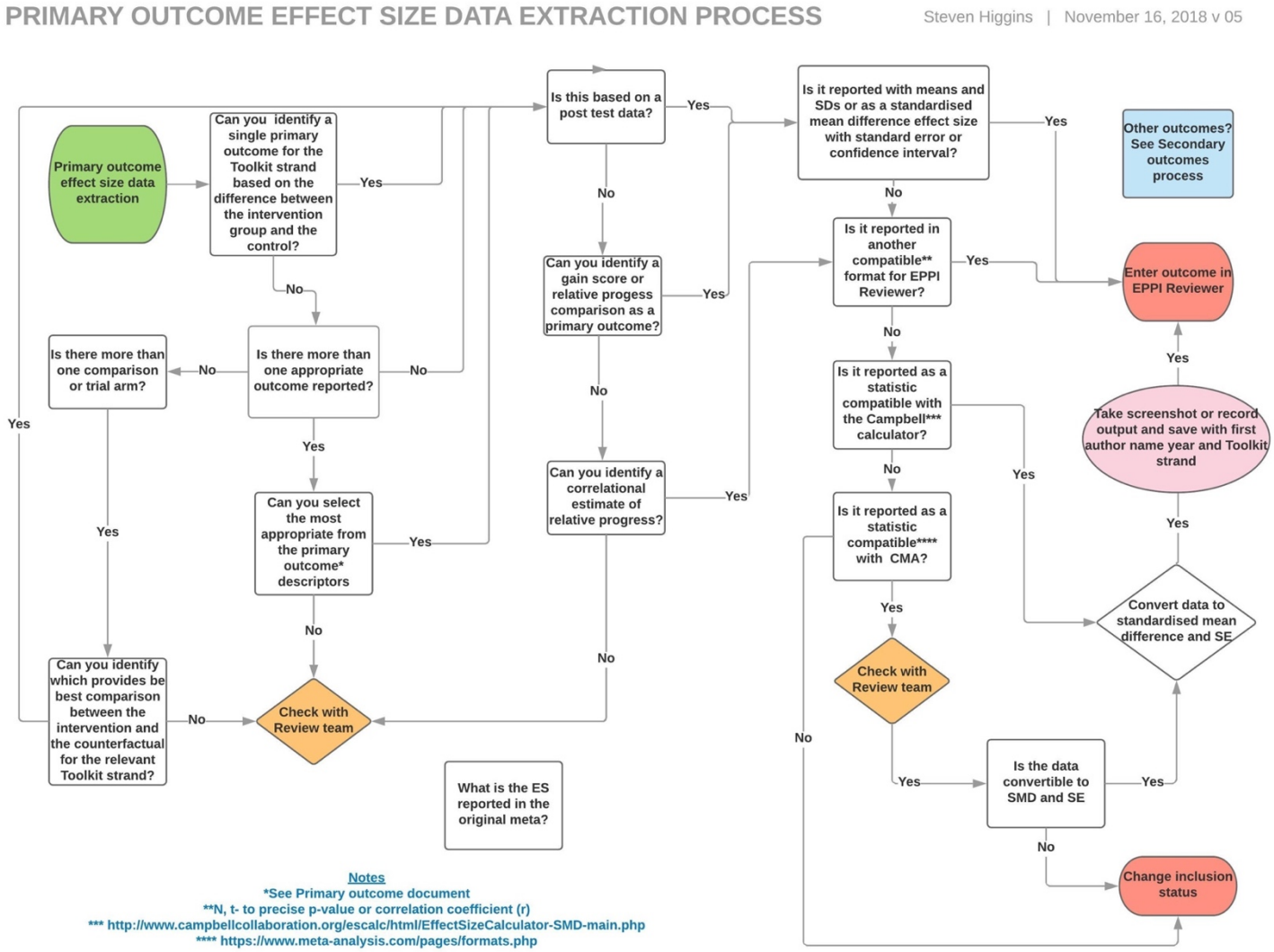
people to engage in additional learning activities, such as football training at a local football club combined with study skills, ICT, literacy or mathematics lessons.

- Toolkit: Summer schools [Outcome classification code]
  Summer schools are lessons or classes during the summer holidays, and are often designed as catch-up programmes. Some summer schools do not have an academic focus and concentrate on sports or other non-academic activities. Others may have a specific focus, such as pupils at the transition from primary to secondary school, or advanced classes to prepare high-attaining pupils for university.
- Toolkit: Teaching assistants [Outcome classification code]
  Teaching assistants (also known as TAs or classroom support assistants) are adults who support teachers in the classroom. Teaching assistants' duties can vary widely from school to school, ranging from providing administrative and classroom support to providing targeted academic support to individual pupils or small groups.
  Cognate terms: support staff; adult support staff; teaching assistants; associate staff; classroom assistants; classroom support assistant; auxiliary teachers; teacher's aide; education paraprofessional; nursery nurse (in early years' settings)

- Comparison [Not selectable (no checkbox)]
  Please do not mark this section. This section is completed in the 'Outcomes specific code' screen.
  - With active control [Comparison]
    i.e. there is control for novelty/ an introduced new treatment
  - With business as usual [Comparison]
    i.e. comparison group having usual learning experience
  - With no equivalent teaching [Comparison]
    i.e. additional learning time / no treatment, such as in a Summer School intervention or a Before or After school club
- Intervention outcome measure [Not selectable (no checkbox)]
  Type or focus of educational test used to measure the outcome of the impact of the intervention or approach.
  - Literacy: reading comprehension [Intervention]
    e.g. passage comprehension
  - Literacy: decoding/phonics [Intervention]
  - Literacy: spelling [Intervention]
  - Literacy: reading other [Intervention]
    Other reading outcomes (e.g. reading fluency, vocabulary comprehension (receptive vocabulary))
  - Literacy: speaking and listening/oral language [Intervention]
  - Literacy: writing [Intervention]
  - Mathematics [Intervention]

- Science [Intervention]
- Social Studies [Intervention]
  e.g. history, geography, economics
- Arts [Intervention]
  e.g. music, art
- Languages [Intervention]
  Second or foreign languages, based on the dominant language of instruction in the educational setting.
- Curriculum: other [Intervention]
  Other curriculum outcomes not included in the above options (please specify)
- Combined subjects [Intervention]
  Where the study combines two or more test outcomes from different subjects to provide an overall measure of educational progress (e.g. KS2 English and mathematics or multiple GCSE subjects.
- Cognitive: reasoning [Intervention]
  Tests of verbal, analogical or visual reasoning, including IQ or other 'intelligence' tests.
- Cognitive: other [Intervention]
  Other tests of cognitive performance such as working memory or perception.

Appendix G: Primary outcome identification



Appendix H: Effect size conversions

There are a number of online and software tools which allow various descriptive and inferential statistics to be converted to an effect size and associated standard error (and/or confidence interval). However, it is always important to consider what the data represents, before undertaking a conversion.

David Wilson's 'Practical Meta-Analysis Effect Size Calculator' hosted on the Campbell Collaboration site: http://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-SMD-main.php

Possible conversions in David Wilson's 'Practical Meta-Analysis Effect Size Calculator'

Standardised Mean Difference (d)
Means and standard deviations
t-test, unequal sample sizes
t-test, equal sample sizes
F-test, 2-group, unequal sample sizes
F-test, 2-group, equal sample sizes
t-test p-value, equal sample sizes
t-test p-value, unequal sample sizes
Means and standard errors
2 by 2 frequency table
Binary proportions
Point-biserial correlation, equal Ns
Point-biserial correlation, unequal Ns
Point-biserial correlation p-value, equal Ns
Point-biserial correlation p-value, unequal Ns
Phi-coefficient
Phi-coefficient p-value
Chi-square
Chi-square p-value
Frequency distribution
Frequency distribution (proportions)
Unstandardised regression coefficient
Standardised regression coefficient
Means and full sample standard deviation
Mean gains scores and gain score SDs
Mean gain scores, pre and post SDs, and paired t-tests
Mean gain scores, pre and post SDs, and pre-post r
Means and standard deviations with subgroups
F-test, 3 or more groups
Means and ANCOVA
Two-way ANOVA

Possible conversions in Comprehensive Meta-analysis (CMA)

Continuous (means)

- Unmatched groups, post data only
    - Mean, SD and sample size in each group
    - Difference in means, common SD, and sample size
    - Cohen's d (standardised by pooled within-groups SD) and sample size

    - Means, sample size, and t-value
    - Difference in means, sample size, and t-value
    - Sample size and t-value
    - Means, sample size, and p-value
    - Difference in means, sample size, and p-value
    - Sample size and p-value
- Unmatched groups, pre and post data
    - Means, SD pre and post, N, in each group, Pre/Post Corr
    - Means, SD difference, N, in each group, Pre/Post Corr
    - Means pre and post in each group, t within groups, N
    - Means pre and post in each group, p within groups, N
    - Means pre and post in each group, F for difference between changes, N
    - Mean change, SD pre and post, N, in each group, Pre/Post Corr
    - Mean change, SD difference, N, in each group, Pre/Post Corr
    - Mean change in each group, t within groups, N
    - Mean change in each group, p within groups, N
    - F for difference between changes, N
- One group (pre-post) and matched groups
    - Mean difference, SD of difference, and sample size
    - Means, SD Pre, SD Post, Pre/Post correlation, and sample size
    - Means, sample size, and paired t-value
    - Means, sample size, and paired p-value
    - Mean difference, sample size, and t-value
    - Mean difference, sample size, and p-value
    - Sample size and t-value from paired t-test
    - Sample size and p-value from paired t-test
- Computed effect sizes
    - Raw difference in means and confidence limits (independent groups)
    - Raw difference in means and standard error (independent groups)
    - Raw difference in means and variance (independent groups)
    - Cohen's d (standardised by pooled within-groups SD) and confidence limits
    - Cohen's d (standardised by pooled within-groups SD) and standard error
    - Cohen's d (standardised by pooled within-groups SD) and variance
    - Hedges' g (standardised by pooled within-groups SD) and confidence limits
    - Hedges' g (standardised by pooled within-groups SD) and standard error
    - Hedges' g (standardised by pooled within-groups SD) and variance
    - Raw mean difference and confidence limits (paired study)
    - Raw mean difference and standard error (paired study)
    - Raw mean difference and variance (paired study)
    - Cohen's d (standardised by SD of difference scores) and confidence limits
    - Cohen's d (standardised by SD of difference scores) and standard error

www.CEDILprogramme.org