Standards and Procedures for *Evidence 4 Impact* (E4I)

October 2017

What is E4I?

Evidence 4 Impact (E4I) is an independent tool that provides school leaders and teachers with reliable, easy-to-use information on educational interventions/programmes that meet solid standards of evidence. It can guide their decisions about what interventions to consider implementing in their school. For interventions that are currently available to schools in the UK it includes links to the research evidence and to the providers. You can find it here: http://www.evidence4impact.org.uk

Many educational interventions claim to be **evidence-based** or **supported by research**, but it can be difficult to assess these claims or compare the evidence for different interventions. E4I aims to overcome this difficulty by providing a simple evidence rating system, along with concise evidence summaries, which enable you to make a well-informed judgement on the extent to which an intervention's effectiveness is proven. This document explains how these ratings are decided.

A positive evidence rating on E4I is not a guarantee that an intervention or programme will work in your school but it is a good indication of what should be most likely to work, given appropriate conditions. The tool is continually updated as more research evidence emerges.

Central to E4I is a database of interventions that are available to be implemented in the UK. The database can be easily searched for the impact of an intervention on a range of outcomes (primary or secondary reading, writing, maths, science or social-emotional). The results can then be filtered by key stage, subject area and targeted group, so that they are tailored to the specific needs of your class or school. Each intervention review includes:

- a rating of its effectiveness
- an overview of the intervention or programme
- a summary of the evidence of its effectiveness
- a Union Jack symbol if the intervention has been evaluated in the UK and has shown some level of effectiveness
- a link to the provider of the intervention in the UK
- a link to review(s) or studies of the intervention
- where available, indicative costs for UK schools to implement the intervention

E4I is updated regularly to include new interventions or to take new research into account. If you have developed or are using an intervention or programme which is not included in the E4I database but you think should be, please let us know.

KEY STAGES

Key Stage 1

TARGETTED GROUPS

Struggling readers

PRACTICES

One-to-one tuition

COST

£2658 per pupil

READING RECOVERY

STRONG
FOR PRIMARY READING

🇬🇧 EVALUATED IN THE UK

❓ How are these ratings generated?

PROGRAMME DESCRIPTION

Reading Recovery® is a short-term early intervention tutoring programme aimed at struggling readers in KS1 (the lowest-achieving 20%). The goals of Reading Recovery® are to promote literacy skills, reduce the number of KS1 pupils who are struggling to read, and prevent long-term reading difficulties.

The intervention can last up to 20 weeks but pupils who progress more quickly may finish within 12 weeks. The focus of each lesson is to comprehend messages in reading and construct messages in writing, learning how to attend to detail without losing focus on meaning. The lessons include assessment, reading known stories, reading a story that was read once the day before, writing a story, working with a cut-up sentence, and reading a new book. The lesson series finishes when the child is able to read and write without help, at the appropriate level for their age.

STAFFING REQUIREMENTS

Pupils participating in Reading Recovery receive daily 30-minute one-to-one lessons with a specially trained Reading Recovery teacher, which supplement their normal reading classes.

EVIDENCE SUMMARY

The Best Evidence Encyclopaedia (2009) rated Reading Recovery as having strong evidence of effectiveness for struggling readers but only limited effectiveness for pupils with English as an Additional Language.

The What Works Clearinghouse (2008) found Reading Recovery to have positive effects on alphabetics and general reading achievement, and potentially positive effects on fluency and comprehension.

The studies found a mean effect size of +0.28 for primary reading.

	NO. OF STUDIES	NO. OF STUDENTS	AVERAGE EFFECT SIZE
STRONG FOR PRIMARY READING	7	1,280	+0.28

KEY RESEARCH

🔗 http://www.bestevidence.org.uk/programmes/reading_recovery.html

🔗 https://ies.ed.gov/ncee/wwc/EvidenceSnapshot/420

E4I evidence standards

Interventions listed on the E4I website are given a rating to show how well their effectiveness is supported by high quality evaluations. E4I has the following categories: Strong, Moderate, Limited, No Impact and Not Evaluated. An explanation of how we arrived at a rating and what this means in practice is presented in the table below.

Rating	Level of evidence	What does this mean?	What should an educator do?
Strong	At least one randomised study with a collective sample size of 500 students (analysed at the individual level) or 30 classes/schools (analysed at the class/school level), and a sample-size-weighted effect size of at least +0.20	Has been shown to work in well-controlled studies.	This intervention has a good chance of improving your pupils' outcomes if it is implemented as designed.
Moderate	At least one randomised or matched study with a sample size of 300 students (analysed at the individual level) or 20 classes/schools (analysed at the class/school level) with a mean effect size of at least +0.10.	There is a moderate level of evidence supporting the intervention.	If there are no interventions with strong evidence on the outcomes that you are targeting, then interventions in this category would be worth using.
Limited	At least one randomised or matched study with a collective sample size of 150 students (analysed at the individual level) or 10 classes/schools (analysed at the class/school level) with a mean effect size of at least +0.05.	Some indication of impact but little evidence supporting the intervention.	If there are no interventions with moderate or strong evidence on the outcomes that you are targeting, you might use an intervention in this category and carefully assess pupil progress.
No Impact	The studies meet the criteria for Limited or better but the results showed a sample-size-weighted mean effect size less than +0.05.	Insufficient indication of positive effects of the intervention.	Look for an alternative intervention that has evidence of effectiveness or pilot the intervention and evaluate its effectiveness.
Not Evaluated	No studies meet the criteria for inclusion so the effectiveness of the intervention cannot be determined at this time.	This intervention has not been evaluated in a robust study.	You should look for an intervention that has evidence of effectiveness or pilot the intervention and evaluate its effectiveness.

Within each category we will present interventions according to an algorithm that emphasises the following, in order of importance:

1. Weighted mean effect size, across all qualifying studies.

2. Number and quality of studies.

3. Collective sample size across all qualifying studies.

Evaluated in the UK

If an intervention has been evaluated in the UK and found to reach a rating of Limited or above in the UK evaluation(s), a Union Jack symbol will indicate that on the ratings page. This is to provide added confidence that the intervention will be effective if implemented with UK pupils.

Procedures

The following procedures are carried out by the reviewers of the Best Evidence Encyclopaedia. If the results of a systematic review of an intervention do not align with the EEF evaluation findings we will work together to find a solution.

Finding eligible studies

Electronic searches of educational databases (eg, JSTOR, ERIC, EBSCO, Psych INFO, Dissertation Abstracts International), using appropriate keywords depending on the subject and age. Tables of contents of relevant journals are scanned for potential studies, as well as reference lists for additional studies. Searches of developers' websites. Studies submitted by developers, researchers, and others are accepted for initial screening. The intention is to accept for initial screening every study that could possibly meet inclusion standards.

Initial eligibility screen

1. Studies must be of interventions that are currently available for implementation in the UK. This is initially determined by a review of the intervention providers' website (if it exists) and other information suggesting an active dissemination effort. Where this is unclear, we contact providers or researchers linked to the intervention to find out about its current availability in the UK.
2. Studies must have assessed achievement or non-cognitive outcomes for children between the ages of 3 and 18.
3. Studies have to have been carried out from 1990 to the present or from 2000 to the present if they evaluated technology approaches.
4. Studies must have compared experimental groups to control groups. Either random assignment to conditions or matched, quasi-experimental assignment based on pre-specified schools, classes, or students had to be used. After-the-fact (post hoc) matching is not acceptable, and comparisons to norming groups, pre-post comparisons, or other non-experimental comparisons are not accepted.
5. Control groups should have been taught essentially the same content using "business as usual" or alternative teaching methods. Comparisons of two equally innovative approaches, without a control group representing ordinary practice, are not accepted.
6. Studies have to provide pretest data to establish initial equivalence. On achievement measures, the average pretest difference should not exceed 25% of a standard deviation. Studies must have established equivalence before the experiment, and also equivalence at pretest for the remaining sample after attrition at the end of the study.
7. Studies' dependent variable(s) must include a quantitative measure of the outcome. The measure could be a standardised test or a test created by test developers not involved with the research, but tests made by the developers or researchers of the intervention themselves are not acceptable. Also, tests that are aligned with content taught in the experimental but not the control group are not acceptable. For example, a study teaching electricity could not use a test of electricity if the control group is not also teaching electricity, even if the test is standardised. Tests administered individually by students' own teachers or others with a potential stake in the outcome are not accepted.
8. Study durations have to be at least 12 weeks, from intervention inception to post-test.
9. Studies must be conducted in 'real-world settings' not in laboratories or other artificial settings. If conducted in a school setting, studies must have at least 2 teachers and 30 students per treatment. Effects from small studies of the same intervention may be pooled.
10. Studies could have taken place anywhere, as long as the intervention being evaluated is available to schools in the UK and the report is available in English.
11. From pretest to post-test, attrition should be similar between experimental and control groups. Studies with differential attrition of more than 15 percentage points will be rejected. Also, if attrition causes the pretests of the final sample to differ by more than 25% of a standard deviation, the study

will be excluded.

12. Studies must have used a form of an intervention that could in principle be replicated. Studies that provided exceptional, non-replicable resources, such as placing a graduate student in each class to help teachers every day, will not be included.

13. EEF evaluations should have at least three padlocks on the EEF findings security rating to be included in E4I.

Summarising study outcomes

Clustering

Students go to schools and classes in clusters, and clustering must be accounted for statistically. (Studies in which students are assigned individually, such as tutoring studies, are an exception.) The requirement to account for clustering means that only large studies, typically involving 40-50 schools or classrooms, have sufficient power to detect effect sizes of +0.20, a typical target.

To obtain a rating in the Strong or Moderate category, studies in which assignment and/or treatment took place at the cluster level must use multi-level modelling (MLM) or other means of accounting for clustering. However, studies in which pupils participate in the intervention in clusters but were analysed at the student level will be considered correlational (as they are), and therefore qualify as Limited if the effects are at that level or above. This will keep the numbers of qualifying interventions high, while reserving the Strong and Moderate categories for large experiments analysed at the appropriate cluster level that are likely to replicate in practice because they have already proven successful at scale.

Calculating effect sizes

Ordinarily, effect sizes (ES) should be calculated as the experimental–control difference in means (adjusted for covariates) (XE-XC) divided by the unadjusted pooled post-test standard deviation (SD) (or that of the control group if the pooled SD is not available):

$\underline{ES = XE-XC}$

   Pooled SD

Standard deviations already adjusted for pretests or other covariates may not be used as the denominator of the effect size formula. SDs of gain scores may not be used. Only unadjusted SDs are acceptable. If reviews have already been completed we will use whatever the reviewer used if it does not make a substantive difference.

Pooling effect sizes

Effect sizes will be pooled at the study level and the programme level. If there are separate measures reported, we will combine outcomes for measures of the same subject. A single effect size for each subject (eg, science) will be calculated for each study, and then effect sizes will be averaged across studies, weighted by sample size using an inverse variance procedure. We will generally base the effectiveness rating on the outcomes for the total sample included in the studies. Where relevant, we will provide information about effects on subgroups.

Additional information

We will seek to obtain from each study the following information by treatment condition:

•	Initial, final sample sizes

•	Initial, final number of schools

•	Initial, final number of teachers

•	Percent free school meals

•	Percent English as an Additional Language (EAL)

•	Percent Special Educational Needs and Disabilities (SEND)

•	Type of location (eg, urban/rural)

Other information

A team of people from the IEE, EEF, and Johns Hopkins University will review the E4I procedures and ratings regularly.

E4I is a growing resource, which will develop as more educational interventions are rigorously evaluated. As findings from EEF-funded projects and other high quality evaluations are produced, these will be added to the E4I ratings. We welcome suggestions for additional programmes and evaluations to be included in E4I. Please contact the IEE at iee@york.ac.uk.

Useful websites

Best Evidence Encyclopaedia

www.bestevidence.org.uk

Blueprints for Healthy Youth Development
www.colorado.edu/cspv/blueprints

Child Trends
www.childtrends.org

Early Intervention Foundation
www.eif.org.uk

Education Endowment Foundation
https://educationendowmentfoundation.org.uk/resources/teaching-learning-toolkit

EPPI (The Evidence for Policy and Practice Information and Co-ordinating Centre (part of the Social Science Research Unit, Institute of Education, University of London)
www.eppi.ioe.ac.uk/cms/

The Campbell Collaboration
www.campbellcollaboration.org

Investing in Children
www.investinginchildren.eu

Promising Practices Network
www.promisingpractices.net/programs.asp

Social Programs that Work
www.evidencebasedprograms.org

What Works Clearinghouse
www.whatworks.ed.gov

Contact us

We welcome suggestions for additional programmes and evaluations to be included in E4I. Please contact the IEE at iee@york.ac.uk.