



Education
Endowment
Foundation

Review: EEF Implementation and Process Evaluation (IPE) Quality Pilot

August 2021

Bronwen Maxwell, Anna Stevens, Sean Demack, Mike Coldwell, Claire Wolstenholme, Sarah Reaney-Wood, Bernadette Stiehl (Sheffield Institute of Education, Sheffield Hallam University)

Hugues Lortie-Forgues (University of York)



The Education Endowment Foundation (EEF) is an independent grant-making charity dedicated to breaking the link between family income and educational achievement, ensuring that children from all backgrounds can fulfil their potential and make the most of their talents.




The EEF aims to raise the attainment of children facing disadvantage by:

- identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;
- evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale; and
- encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus Trust (now part of Impetus - Private Equity Foundation) and received a founding £125m grant from the Department for Education.

Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

For more information about the EEF or this report please contact:

-  Jonathan Kay
Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP
-  0207 802 1653
-  jonathan.kay@eefoundation.org.uk
-  www.educationendowmentfoundation.org.uk

Contents

About the reviewer	4
Acknowledgements	4
Executive summary	5
Study objective.....	5
The IPE quality measure	5
Quality of EEF trial IPEs	5
Introduction	7
Background and study objective	7
Ethics and data protection.....	7
Project team	7
Development and operationalisation of the IPE quality measure	8
Design rationale.....	8
The piloted IPE quality measure	9
IPE quality findings	11
Conclusion	13
Potential and limitations of the IPE quality measure	13
Quality of EEF trial IPEs	13
References	15

List of tables and figures

Table 1: Core Sheffield Hallam University project team for the IPE quality pilot.....	7
Table 2: Data and inference quality (Tashakkori and Teddlie, 2003, p.694)	8
Table 3: IPE quality measure	9
Table 4: Intersection of overall IPE quality and padlock rating —mean padlock rating for levels of IPE quality	12
Table 5: Intersection of padlock categories and overall IPE quality—mean IPE quality for levels of padlock rating	12
Figure 1: Distribution of IPE quality categorisations by variable and overall	11

About the reviewer

The 2019/2020 Review of EEF Projects was conducted by a team from the Sheffield Institute of Education (SloE), including Sean Demack, Bronwen Maxwell, Mike Coldwell, Anna Stevens, Claire Wolstenholme, Sarah Reaney-Wood, and Bernadette Stiel, together with a senior statistician from the University of York, Hugues Lortie-Forgues. This is one of three publications from the review. The lead evaluator for this report was Bronwen Maxwell.

The Sheffield Institute of Education (SloE) at Sheffield Hallam University is a leading provider of initial and continuing teacher education—undergraduate, post-graduate, and doctoral education programmes—and has a long-established track record in educational research, evaluation, and knowledge exchange. Key areas of research and evaluation expertise span curriculum and pedagogy, policy and professional learning, and diversity and social justice. The SloE has extensive experience of evaluation methodologies and undertaking large-scale evaluations for a range of funders, including the Education Endowment Foundation (EEF).

Contact details:

Principal investigators: Sean Demack and Bronwen Maxwell

Sheffield Institute of Education Research and Knowledge Exchange Centre

Sheffield Hallam University

Room 10101, Arundel Building

City Campus

Howard St

Sheffield

S1 1WB

Tel: 0114 225 6066

Email: SloECDARE@shu.ac.uk

Acknowledgements

Lisa Clarkson, Lyn Pickerel, Rosie Smith, Noreen Drury, Charlotte Rowley, Hannah Joyce, Kellie Cook, Coding team

Executive summary

Study objective

The study objective was to develop an Implementation and Process Evaluation (IPE) quality measure and pilot it using data coded in the Review of EEF Projects (Demack et al., 2021) from the 79 EEF trial reports that incorporated IPEs and had been published up to January 2019.

The IPE quality measure

This first attempt to develop a measure of IPE quality in EEF trials draws on Tashakkori and Teddlie's (2003) conceptual framework for assessing the quality of mixed-methods studies, which incorporates the components of data quality (validity or trustworthiness and reliability or dependability) and inference quality (that is, design quality—a combination of methodological rigour and interpretative rigour). The IPE quality measure has been designed with five dimensions: sufficiency of data sources, data collection methods, sampling, analysis, and conduct. Qualitative 'high', 'medium', and 'low' quality criteria have been developed for each dimension separately to capture data and inference quality. Rules for combining the quality gradings on each dimension to create an overall grading of high, medium, or low quality for the IPE were also developed.

The IPE quality measure has potential, with further refinement, for use either to support evaluators in IPE design and reporting or as a quality assessment measure. Should the EEF wish to use the IPE quality measure for assessment purposes, evaluators will require more precise guidance on reporting so a fair assessment can be made.

The IPE quality measure could be further developed by:

- incorporating specific criteria related to the extent to which the IPE tested the theory of change and gathered and analysed data to test—
 - the hypothesised causal mechanisms,
 - the quality of combining methods, and
 - the quality of integrating impact and IPE data and findings (a more recent focus in EEF guidance for evaluators);
- a stronger emphasis on the measurement of inference rigour; and
- (for dimensions that include more than one criterion for assigning a high, medium, or low categorisation) developing a scoring system for the dimension.

Future use of the tool will require full inter-rater reliability checks to reduce bias.

Quality of EEF trial IPEs

While there are limitations with the measure, it is reasonable to claim that there is significant variation in the quality of IPEs in EEF trials. Overall, 24 evaluations (30%) were identified as having a high quality IPE, 38 (48%) as having a medium quality IPE, and 17 (22%) as having a low quality IPE.

Three dimensions of IPE quality that were most commonly identified as high quality were the sufficiency of data sources, data collection methods, and IPE conduct. Sampling methods were most likely to be classed as medium quality and analysis methods were most likely to be classed as low. In a significant number of reports, IPE analysis methods were weakly specified or absent.

IPE quality increased over time from 2014 to 2019, reflecting the EEF's increasing focus on IPE design, but there remains substantial variation in reporting of both IPE methods and findings across reports with no clear trends observed with respect to variance around the mean over time. This increase in quality aligns with the observed increase in mean EEF

'padlock' rating—an indication of the security of the findings—over time indicating increasing quality, overall, in EEF trial designs.

Introduction

Background and study objective

The EEF's mission is to break the link between family income and educational achievement. This is achieved through summarising the best available evidence in plain language, generating new evidence of 'what works' to improve teaching and learning, and supporting teachers and school leaders to use research evidence to maximise the benefit for young people. Evidence of what works to raise attainment is primarily generated through randomised controlled trials commissioned by the EEF. These trials include a quantitative impact evaluation to estimate the effect on pupils' attainment (and, in some cases, other outcomes) and an implementation and process evaluation (IPE).

The EEF first published an introductory handbook on the conduct of IPEs of interventions in education settings for evaluators in 2014 (Humphrey et al., 2016b). This drew on a synthesis of existing evidence (Humphrey et al., 2016a). The handbook refers to an IPE as:

'the generation and analysis of data to examine how an intervention is put into practice, how it operates to achieve its intended outcomes, and the factors that influence these processes' (p.6).

As further guidance published by the EEF in 2019 states, IPEs are important 'to help explain why an intervention has or has not been successful, what factors have contributed to this result, and what lessons we can learn about educational practice and research' (EEF, 2019, p.1). Building on Humphrey's work and EEF guidance issued in 2017, the 2019 guidance sets out the key principles for EEF evaluators planning, conducting, and reporting IPEs. It aims to increase the quality and transparency of IPEs and incorporates a stronger focus on the integration of impact and IPE analyses.

As part of a wider review of EEF projects (Demack et al., 2021) commissioned by the EEF and undertaken in 2019 and 2020, an objective was set to develop an IPE quality measure and pilot it using data coded in the main review from the 82 EEF trial reports that had been published at the time of the review. The 2019 EEF IPE guidance was not available at the time of the development of the IPE quality measure.

Ethics and data protection

The project received ethical approval from the Faculty of Social Sciences and Humanities at Sheffield Hallam University. All data used in the quantitative analyses in the main Review of EEF projects (Demack et al., 2021) that has been extracted and further analysed for this report has been coded from publicly available sources (the EEF trial reports). No personal data was held for the purposes of compiling this report.

Project team

Table 1: Core Sheffield Hallam University project team for the IPE quality pilot

Team member	Title	Role/responsibilities
Professor Bronwen Maxwell	Head of Commissioned Research	Principal investigator IPE measure design and interpretation of pilot findings
Anna Stevens	Research Fellow	Extraction of quantitative data from main review and pilot analyses
Sean Demack	Principal Research Fellow	Additional analyses

The full project team for the EEF review that generated the quantitative data used in this report is set out in Demack et al. (2021).

Development and operationalisation of the IPE quality measure

Design rationale

The EEF IPE Introductory Handbook (Humphreys et al., 2016), which sets out the EEF's expectations for the design of high-quality IPEs in trials, was the starting point for developing an IPE quality measure.¹ Given the recommendation that IPEs should routinely adopt a mixed-methods design, a brief review of the literature on assessing quality in mixed-methods studies was also undertaken (see O'Caithain, 2015, for an overview). Measuring quality in mixed-methods studies presents certain challenges. Firstly, traditionally different approaches and terminology have been applied in quantitative and qualitative studies. For example, the concepts of validity and reliability, which are used to assess the quality of quantitative studies, are frequently regarded as inappropriate for assessing the quality of qualitative studies where concepts such as trustworthiness, dependability, and authenticity may be applied. Secondly, within both traditions measures of quality do not always cover (or are not applied in ways that cover) all aspects of the study that influence the robustness of knowledge claims. To address both of these issues, Tashakkori and Teddlie (2003) propose that the assessment of quality in mixed-methods studies should distinguish between *data quality* (validity or trustworthiness, reliability or dependability) and *inference quality* (that is, design quality—a combination of methodological rigour and interpretative rigour) and that attention is paid to both these constructs of quality. The key components of data quality and inference quality are set out in Table 2.

Table 2: Data and inference quality (Tashakkori and Teddlie, 2003, p.694)

Component being evaluated	Definition; question asked	Components or aspects of evaluation	Evaluation question	Evaluation and improvement strategy
Data quality	Do the data, records, observations, etc. meet the minimum criteria to be acceptable and trustworthy?	Validity, trustworthiness	Did we indeed capture the phenomenon or attribute that we intended to (or we believe we captured)?	Consistency within aspects of the <i>same</i> measurement or observation procedure; method (data collection procedure) triangulation.
	Does the data adequately represent the theoretical phenomena or the attributes under study?	Reliability, dependability	Did we <i>accurately</i> capture or represent the phenomenon or attribute under investigation?	Consistency between different procedures for measurement and observation of the same phenomenon or attribute; audit trail, data triangulation.
Inference quality	Does the inference meet the minimum criteria to be defensible and credible?	Design quality	Were the procedures implemented with quality and rigour? Is there 'within-design' consistency?	Was the method of study appropriate for answering the research question(s)? Was the method capable of capturing the answers, effects, and relationships? Were the components of the design (measurement, sampling etc.) implemented adequately?
		Interpretive rigour	Are the results or findings interpreted in a defensible manner? Is there: <ul style="list-style-type: none"> • cross-inference consistency? • theoretical consistency? • interpretive agreement? • interpretive distinctiveness? 	Does the inference follow the findings? Are the interpretations consistent with the theory and state of knowledge in the field? Are the inferences consistent with each other? Do the global inferences adequately incorporate the inferences made from the QUAL and QUAN strands of the study?

¹ The IPE Guidance issued in 2019 was not available at the time of developing the IPE quality measure for this pilot.

This framework influenced the construction of our pilot measure of IPE quality, although there are significant limitations in assessing interpretive rigour in our measure. These and other limitations are considered further at the end of this section.

A key challenge was developing a measure that could be operationalised within the boundaries of this review. Three issues were central to the final construction of the measure and the associated codes. First, there needed to be sufficient information in most of the trial reports included in the review to support reliable coding decisions. Prior to the main coding stage this requirement was tested through a rapid review of ten trial reports. Second, the code descriptions needed to be sufficiently precise to permit reliable coding that can be applied within the time constraints of the coding process for this review. Third, the code descriptors need to be applicable to both quantitative and qualitative methods and data.

Coding of the IPE quality variables followed the same procedure and quality checks as for the main review of EEF projects (Demack et al., 2021). It became apparent that there was inconsistency across the wider coding team and that for a number of reports some coders recorded that they were not confident enough to make judgements on specific codes. This is not surprising as accurate coding requires an in-depth understanding of IPE design. To address this issue, all the IPE variables for all trial reports were checked and where necessary re-coded by one senior researcher. While this approach provided a high level of consistency in coding it also points to a further potential limitation of wider use of the measure.

The piloted IPE quality measure

The IPE quality of an evaluation was assessed across five dimensions: sufficiency of data sources, data collection methods, sampling, analysis, and conduct. The coding frame is presented in Table 3.

Table 3: IPE quality measure

Variables and codes	Coding descriptors
Sufficiency of data sources	
1. High	Data is collected from all the groups that are necessary to answer the RQs, for example (as appropriate), leaders, teachers, pupils, delivery partners, and others connected to the intervention. Where the focus of the intervention is pupil change, this includes data collected directly from pupils or through observation of pupils engaging in the intervention. Data is also collected from the control group to the extent necessary to establish the 'business as usual' condition.
2. Medium	Some gaps in data collection from groups that are necessary to answer the RQs, for example (as appropriate), leaders, teachers, pupils, delivery partners, and others connected to the intervention or insufficient data is collected from the control group to establish the 'business as usual' condition.
3. Low	Significant gaps in data collection from groups that are necessary to answer the RQs, for example (as appropriate), leaders, teachers, pupils, delivery partners, and others connected to the intervention or no data is collected from the comparison group.
Quality of data collection methods	
1. High	All methods of collecting data are clearly specified and valid—they measure what they are supposed to measure.
2. Medium	Methods of collecting data are variably specified or variably valid—they do not all measure what they are supposed to measure.
3. Low	Methods of collecting data are poorly specified or lack validity—most do not measure what they are supposed to measure.
Quality of data sampling	
1. High	Sampling approach is clear, justified, and appropriate in relation to all methods used. For qualitative work, the sample does not need to be statistically representative but to be categorised as 'high' it would require a sample that is random or purposive rather than a convenience sample.
2. Medium	Sampling approach is largely clear, reasonably well justified, and appropriate to the methods used but does not fully meet the criteria for 'high'.
3. Low	Sampling approach is unclear or is poorly justified or not appropriate to the methods used.
Quality of analysis methods	
1. High	Methods of analysis are clearly set out and appropriate in relation to the type/s of data and to answer the research questions.
2. Medium	Methods of analysis are variably set out or vary in appropriateness in relation to the type/s of data and to answer the research questions.
3. Low	Methods of analysis are largely missing or are inappropriate in relation to the type/s of data and to answer the research questions.
IPE conduct	

1. High	Intended data collection and analysis methods are followed or any changes to methods are justified and appropriate.
2. Medium	Intended data collection and analysis methods are not always followed or changes to methods are not always clearly justified or are not always appropriate.
3. Low	There is low adherence to intended data collection methods or it is unclear whether intended data collection and analysis methods are followed, or any changes to data collection or analysis methods are generally not justified or not appropriate.

For reporting purposes, an overall measure of IPE quality for each trial report was constructed by assigning a value of three to high scores, two to medium scores, and one to low scores across the five IPE dimensions. From this, a total mean IPE quality score was calculated. These preliminary overall gradings were then compared to the grading for the data sufficiency variable. If the grading for the data sufficiency was lower than the preliminary overall IPE quality grading, the overall grading was reduced—it could not exceed the data sufficiency category. This reduction in quality was applied because if data is not collected from all the groups that are necessary to answer the research questions the overall quality of the IPE cannot be considered to be high. Five trial reports were adjusted from high to medium overall IPE quality based on this criterion. The revised list of high quality IPE reports was then reviewed to identify any that had a low score on any of the remaining four dimensions (data collection methods, sampling, data analysis methods, and IPE conduct). Where a low score was found, the high grading was adjusted to a medium grading. This resulted in the re-grading of one report. It is important to note that this scoring logic has not been subject to testing, so the overall IPE quality findings reported below should be treated with some caution.

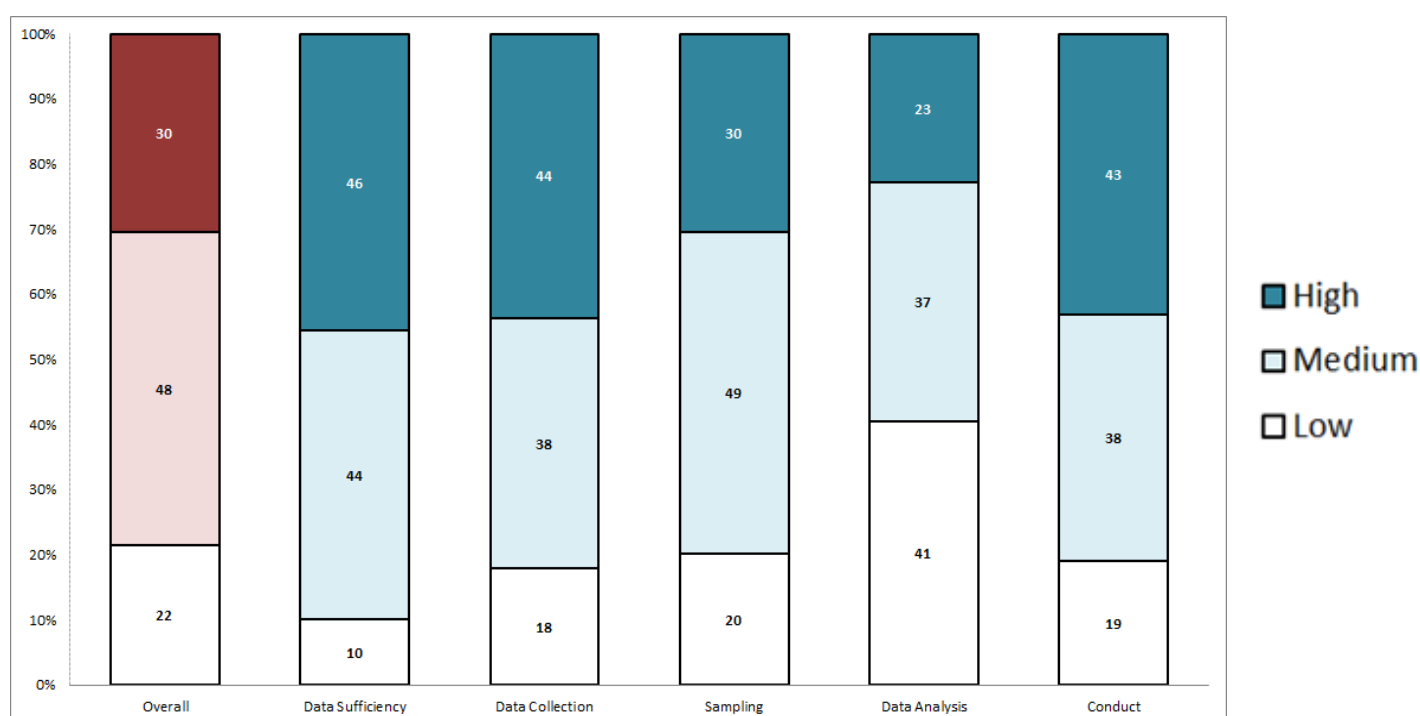
IPE quality findings

Three of the 82 evaluations included in the main Review of EEF Projects did not have an IPE, so the findings are based on 79 evaluations.

Three dimensions of IPE quality that were most commonly identified as high quality were the sufficiency of data sources (36 evaluations, 46%), data collection methods, and IPE conduct (both in 34 evaluations, 41%). For projects that were not coded as high for sufficiency of data sources, the main reasons were the absence of data directly from pupils or from observation of pupils engaged in the intervention (for trials focused on pupil change) or data from the control group to establish 'business as usual'. Sampling methods were most likely to be classed as medium (39 evaluations, 48%) and analysis methods were most likely to be classed as low (32 evaluations, 39%). The low grading for quality of analysis methods was found in a significant number of reports because analysis methods were either not specified or were only weakly specified, each of which is a criterion for a low grading, meaning it is not possible to ascertain whether the methods were appropriate. These findings require careful interpretation as categorisation may simply reflect an absence of an explanation in a report—for example, about the approach to sampling, data collection, or analysis—rather than the approach being inappropriate and undermining the claims that are made.

Overall, 24 evaluations (30%) were identified as having a high-quality IPE, 38 evaluations (48%) a medium quality IPE, and 17 evaluations (22%) a low-quality IPE. The distribution of IPE quality grades across the individual dimensions and the overall IPE quality grade are displayed in Figure 1.

Figure 1: Distribution of IPE quality categorisations by variable and overall



Treating the three-point ordinal overall IPE quality variable as a scale variable (1: low; 2: medium; 3: high), the mean IPE quality across the six years of the review was 2.2. The mean IPE quality was found to increase over time from a mean of 1.6 in 2014 to a mean of 2.5 in 2018 with the three trials published in 2019 all classed as having a high-quality IPE (a mean of 3.0),² reflecting the EEF's increasing focus on IPE design and the development of detailed guidance in 2017. This increase in quality aligns with the observed increase in mean EEF padlock rating over time indicating an increasing quality, overall, in EEF trial designs. A positive correlation (Pearson $r = 0.40$; Spearman $\rho = 0.41$) was observed between IPE quality and EEF padlock rating.³ Further detail on the association was gained by intersecting the

² While the mean IPE quality score was observed to increase over time, no clear trends were observed with respect to variance around the mean. Standard deviations ranged between 0.544 (in 2016) and 0.789 (in 2017).

³ Number of evaluations=79; the three evaluations without an overall IPE quality rating were excluded in these analyses.

mean overall IPE quality category across padlock categories and mean padlocks across overall IPE quality (Tables 4 and 5).

Table 4: Intersection of overall IPE quality and padlock rating —mean padlock rating for levels of IPE quality

	Number of evaluations or trials	Mean padlock rating (SD)
Overall IPE quality		Eta² = 0.16
High	24	3.7 (1.13)
Medium	38	3.1 (1.13)
Low	17	2.2 (1.30)
No IPE quality data	3	3.0 (1.00)
ALL	82	3.1 (1.25)

Table 5: Intersection of padlock categories and overall IPE quality—mean IPE quality for levels of padlock rating

	Number of evaluations or trials	Mean IPE quality rating (SD)
Trial quality (EEF padlocks)		Eta² = 0.20
0	3	1.0 (0.00)
1	7	2.1 (0.69)
2	12	1.6 (0.79)
3	27	1.9 (0.77)
4	24	2.3 (0.81)
5	9	2.6 (0.53)

As well as examining the relationship between IPE and trial security for the 79 evaluations in the review with IPEs, intersecting the two factors could be used to identify 'exemplar' evaluations. For example, there are five five-padlock trials and 11 four-padlock trials that also had a high-quality IPE. A single trial (Chess in Primary Schools) scored the highest across *all* IPE quality dimensions *and* was awarded a five-padlock trial security rating.

Conclusion

Potential and limitations of the IPE quality measure

This was the first attempt to develop an IPE quality measure for EEF trials. The measure developed is underpinned by Tashakkori and Teddlie's (2003) frame for assessing the quality of mixed-methods studies and the components of data quality and inference quality that comprise the frame (Table 2) and incorporates criteria for 'high', 'medium', and 'low' quality across five dimensions: sufficiency of data sources, data collection methods, sampling, analysis, and conduct (Table 3).

Designing the pilot measure created a number of challenges: these included the tension between designing a comprehensive measure that is applicable across a range of mixed-methods approaches and producing a tool that is practical to use (see O'Caithain, 2015, for a fuller discussion).

While the tool does provide a helpful starting point for further development and was found to be relatively easy to administer when applied by a knowledgeable researcher, key limitations of the tool in its current format include:

- the use of categorical codes that include more than one criterion;
- insufficient focus on quality in relation to combining methods;
- some limited measurement of inference rigour; and
- it does not currently include evaluators' engagement with prior evidence and theory and the extent to which the IPE gathered and analysed data to test the causal mechanisms.

Regarding the latter point on evaluator engagement: this was an initial intention but within the scope of this main review of EEF projects we were unable to develop measures that were sufficiently valid or reliable to bring into a quality measure (see the Theory and Evidence theme in the section on Presenting the Explanatory Variables in Demack et al., 2021).

More generally, it is important to note that any IPE quality measure will, to some extent, always include subjective judgements that can be influenced by the assessors' own perspectives on the contested area of what constitutes quality in educational research. While it was beyond the scope of this review to undertake a comprehensive inter-rater reliability check, this is strongly recommended for any future IPE quality assessments to reduce bias. In addition, consideration should be given to using at least some binary variables—for example, data collection from the control group could be operationalised in this way instead as one factor contributing to the data sufficiency category.

The revised IPE guidance (EEF, 2019) published after the design of the measure piloted in this study raises further questions as to how best the measure can be developed. The emphasis on integrating the IPE with the impact evaluation needs to be taken into consideration so, for example, it may be appropriate to look at combined impact evaluation and IPE inference quality rather than inference quality solely in relation to the IPE.

Quality of EEF trial IPEs

Significant variation in the level of detail provided on IPE methods and findings in the evaluation reports also limited the validity and reliability of the measure when applied to the 79 reported EEF trials with an IPE. Three dimensions of IPE quality that were most commonly identified as high quality were the sufficiency of data sources, data collection methods, and IPE conduct. Sampling methods were most likely to be classed as medium quality and analysis methods were most likely to be classed as low. In a significant number of reports, IPE sampling and analysis methods were weakly specified or absent. This may mean that the actual quality of the IPEs may have been higher than it was possible to ascertain from the write-up.

Nonetheless, it is possible to draw the conclusion that there is significant variation in the quality of IPEs in EEF trials. Of the 79 trials reviewed, 24 evaluations (30%) were identified as having a high-quality IPE, 38 evaluations (48%) a medium quality IPE, and 17 evaluations (22%) a low quality IPE. Tentatively, IPE quality was found to increase over time from 2014 to 2018 with the three trials published in 2019 all classed as having a high-quality IPE, most likely reflecting the EEF's increasing focus on IPE design. This increase in quality aligns with the observed increase in mean EEF padlock rating over time indicating increasing quality, overall, in EEF trial designs.

Looking forward, if a measure of IPE quality is to be adopted, evaluators will require more precise guidance on reporting. Alternatively, rather than being used as an assessment tool, the measure could be used as a checklist to support evaluators in IPE design and reporting.

References

- Demack, S., Maxwell, B., Coldwell, M., Stevens, A., Wolstenholme, C., Reaney-Wood, S., Stiell, B. and Lortie-Forgues, H. (2021), 'Review of EEF Projects':
https://educationendowmentfoundation.org.uk/public/files/Publications/Review_of_EEF_Projects.pdf
- EEF (2019) 'Implementation and Process Evaluation Guidance for EEF Evaluations', London: Education Endowment Foundation:
https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_guidance.pdf
- Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. (2016a) 'Implementation and Process Evaluation (IPE) for Interventions in Education Settings: A Synthesis of the Literature', London: Education Endowment Foundation.
- Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. (2016b) 'Implementation and Process Evaluation (IPE) for Interventions in Education Settings: An Introductory Handbook', London: Education Endowment Foundation.
- O'Caithain, A. (2015) 'Assessing the Quality of Mixed Methods Research: Toward a Comprehensive Framework', in Tashakkori, A. and Teddlie, C. (eds), *SAGE Handbook of Mixed Methods in Social & Behavioral Research*:
<https://methods.sagepub.com/book/sage-handbook-of-mixed-methods-social-behavioral-research-2e/n21.xml>
- Tashakkori, A. and Teddlie, C. (2003) 'The Past and Future of Mixed Methods Design: From Data Triangulation to Mixed Methods Designs', in Tashakkori, A. and Teddlie, C. (eds), *SAGE Handbook of Mixed Methods in Social & Behavioral Research*, Thousand Oaks, CA: Sage (pp. 671–702).

You may re-use this document/publication (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence v3.0.

To view this licence, visit <https://nationalarchives.gov.uk/doc/open-government-licence/version/3> or email: psi@nationalarchives.gsi.gov.uk

Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned. The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

This document is available for download at <https://educationendowmentfoundation.org.uk>



The Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP

<https://educationendowmentfoundation.org.uk>

 [@EducEndowFoundn](https://twitter.com/EducEndowFoundn)

 [Facebook.com/EducEndowFoundn](https://www.facebook.com/EducEndowFoundn)