EVIDENCE-BASED EDUCATION: FROM CAUSAL ASCRIPTIONS TO RELATIVE EFFECTIVENESS GENERALIZATIONS AND EFFECTIVENESS PREDICTIONS

A Research Brief by the International Science and Evidence-based Education Assessment
An Initiative of UNESCO MGIEP*

UNESCO MGIEP/RESEARCH BRIEF/ISEE03

Julien Mercier [1]
Iris Bourgault Bouthillier [2]

To more reliably achieve educational goals based on values and policies (Brighouse et al., 2018), quantitative and qualitative traditions should complement each other to strengthen the quality and impact of empirical research under the broad banner of evidence-based education (EBE). And since human learning and development are the cornerstones of educational goals, the domains contributing to educational research rest on a vast number of areas, including in the following fields: learning sciences and cognitive science (psychology and neuroscience [behaviour and brain processes]); computer science (computer-based learning systems, learning analytics); economics; and social sciences (the learner and their broader context).

INTRODUCTION: "WHAT WORKS" IS NOT ENOUGH

This research brief identifies three overarching problems in educational practice which a different approach to EBE can help solve.

Problem 1: There is a need to apply a high minimum standard for what counts as evidence of improved learning. From the perspective of EBE, decisions about which practices to use in a given learning context should ideally be based on evidence (Slavin, 2020). Evidence begins with a demonstration of the effect of a treatment on a defined outcome (Connolly et al., 2018) and, more broadly, empirical support that a policy works generally or in a specific context (Joyce & Cartwright, 2020). However, it is not enough to simply know whether a given intervention works or not. As important as that is, we should expect a higher standard: that is, to know whether an intervention will work better than what we are already doing, compared to a control group, and after eliminating as many sources of bias as possible.

[1] Director, NeuroLab, University of Quebec in Montreal, Montreal, Canada
[2] Ph.D. student, University of Quebec in Montreal, Montreal, Canada

* The first author is Co-Chair of the ISEE Assessment Data and Evidence Group. The second author is a lead author on a full chapter on the present topic, which is to be part of the ISEE Assessment. The authors thank Stephen Fraser, Anantha Duraiappah, Yoko Mochizuki and Clara Cardona for their reviews, comments and inputs on earlier drafts.

Problem 2: Before the need for additional evidence in the form of new interventions, we argue that there is a need for what we call "relative evidence." Relative evidence arises from the combined results of multiple studies, meta-analysis, and thorough comparisons of multiple extant interventions. Such comparisons enable an assessment of all pertinent interventions with respect to their effectiveness against specific outcomes, allowing practitioners to answer the question: *Given all the possible interventions available to me, which is most likely to yield success in my specific context?* The consistency or variability of effect sizes across studies of similar interventions is critical to support assertions regarding their *general effectiveness.* However, there is a lack of relative evidence in extant literature regarding a majority of educational issues: new interventions are being tested against a control group (business as usual) and well-documented interventions are rarely compared and aggregated through proper meta-analytic approaches.

Problem 3: There is a need for strong assertions about how the local context in which the evidence is to be applied ought to affect our expectations of impact. An effectiveness prediction (Joyce & Cartwright, 2020) is the prediction that a given intervention abstracted across causal ascriptions and general effectiveness claims will work within the specific constellation of variables of a given context of application. Effectiveness predictions are generally either absent from implementation efforts, or tackled through biased, non-scientific reasoning, such as beliefs, peer pressure, marketing, etc. In education, it is possible to be a lot more efficient in implementing best practices by applying a rationale increasingly used in other fields.

In this research brief, we are concerned with the information and reasoning needed to address three questions that jointly determine the best course of action for obtaining the best educational outcomes: What works? What works best generally? Will it work "here," tomorrow, in my classroom? We provide an overview of the nature of scientific evidence in education and suggest a framework—the EBE3 framework—that: 1) encompasses current efforts related to the development of educational knowledge; and 2) posits the overall progress of educational research in a dialogue between theory-building and validation. Thus, the framework will be of practical use for educational researchers, policymakers, and practitioners.

THEORY-BUILDING AND THEORY-TESTING IN EDUCATIONAL RESEARCH: DIVIDE, COMPROMISE, OR SYNERGY?

In addition to the levels of scientific evidence that emphasize empirical demonstration, theory—the conceptual model used to explain phenomena—is another essential aspect of educational research (National Research Council, 2002). We place limited emphasis on the grand theories that generalize theoretical understanding and focus instead on mid-range theories that attempt to account for social aspects and particularities of situations. Mid-range

theories consist of representations or abstractions of aspects of reality that can be approximated by conceptual models, which can then be empirically tested. Such conceptual models drive the research question, the use of methods, and the interpretation of results (National Research Council, 2002).

Theory is indispensable in educational science, allowing researchers, decision-makers, and practitioners to consider both the application of interventions in specific contexts and to understand the underlying causal mechanisms. Research contributes to theory in two main ways: theory-building and theory-testing (validation), which are not mutually exclusive. In some fields, the more an article contributes to theory, whether it builds it, tests it, or both, the more impactful it tends to be for the scientific community.

A taxonomy by Colquitt and Zapata-Phelan (2007) captures many facets of the theoretical contributions of an article. The taxonomy is built on two axes, theory-building and theory-testing. According to its authors, the disadvantage of this taxonomy is that it only depicts what empirical studies are intended to do without capturing how well they actually do it. Many other important factors could be added to this taxonomy: how interesting a new construct is, how much a new relationship adds to a literature, how rigorously a theory is tested, and so on.

WHAT IS WORKING BEST GENERALLY?

We consider purely causal ascriptions to be limited in their capacity to inform the implementation of EBE. Consequently, we begin our discussion of the necessary ingredients of an empirical demonstration of effectiveness with the notion of general effectiveness claims. Insofar as applied research improves professional practices in education, and given the impact of these practices on learners, it is desirable to be able to judge the relative value of available research results relevant to practice, following a set of considerations pioneered by Cochrane (1972). For each aspect of the role of the teacher or professional, it must be possible to determine either an absence of research, the presence of poor-quality research, the presence of quality research, and possibly the accumulation of relevant and converging research.

From an interventionist perspective that follows a basic premise, namely that the best information for practice is of an applied and causal nature (Joyce, 2019), it is necessary to arrive at unambiguous inferences between an intervention and its effect on the learner. In this regard, the most consensual criteria on which these causal inferences can be established, and taken up across a majority of applied fields emanating from the human sciences, are brought together through levels of scientific evidence.

In light of the cumulative nature of empirical evidence, the levels of scientific evidence are operationalized domain by domain, by grading the internal and external validity of the available evidence. Also, employing a standard benchmark of effectiveness—the most

common being effect size—is essential when merging evidence about relative effectiveness across increasingly broader educational areas of intervention in order to prioritize intervention among them.

Levels of pseudoscientific and scientific evidence

Table 1 provides a classification of the pseudoscientific and scientific evidence applied to educational research. This proposal of criteria concerning the efficacy of a given intervention seeks to extend prevalent hierarchies of evidence to encompass the various types of evidence created and disseminated, including inadequate, pseudoscientific evidence (e.g. Burns et al., 2011; Evans, 2003) and allows us to distinguish 1) information of a pseudoscientific or non-scientific nature, 2) the results emanating from a scientific approach, and 3) probative evidence concerning the relative convergence and divergence of the integrality of available research results. The terms “probative,” “scientific,” and “pseudoscientific/non-scientific” are used for clarity in relation to the common language of education researchers, practitioners, and policymakers. They are used to provide clear benchmarks to classify sources of evidence, and should not be seen as exclusive or unrelated. Hansson (2009) defines a pseudoscientific assertion using three criteria: (1) it pertains to an issue within the domains of science (in the wide sense); (2) it is not epistemically warranted; and (3) it is part of a doctrine creating the impression that it is epistemically warranted. “Scientific,” in the context of applied educational research, is meant to provide limited empirical indications about the efficacy of a given intervention. “Probative” is understood as the ability of evidence to make an assertion more or less true, in this case the assertion pertaining to “effectiveness.”

Range	Level	Sources of evidence	Main limitations
Probative: provide effectiveness generalizations	1	Mega-analysis, meta-analysis, narrative literature review, evidence-based review	Abstracted, decontextualized recommendations
Scientific: provide causal ascriptions	2	Experimental studies	Do not provide effectiveness generalizations
	3	Quasi-experimental studies	Internal validity
	4	Correlational studies, quantitative case studies	Impossible to verify causality
	5	Expert committees, clinical experience from experts (teamwork reports)	Opinions subject to political or personal influences
	6	Qualitative research, single-case protocols	Lack of generalizability
Pseudoscientific and non-scientific: beliefs not related to solid observation or reasoning	7	Poor-quality research (qualitative or quantitative)	Improper methodology
	8	Absence of research, practice reports, trends	Lack of systematic empirical observations

Table 1 *Levels of evidence applied to educational research and practice*

In short, the levels of evidence are useful in guiding the decision-making process of policymakers and practitioners to improve learning outcomes. Furthermore, levels of scientific evidence, as an index of "readiness for application," are mostly useful for consumers of research to guide educational practice regarding the effectiveness of interventions; they are not an indication of innovation. From a research production perspective, to foster innovation, it is always necessary to have evidence moving across/ up the hierarchy for different topics/ research questions, and having evidence at every level of the hierarchy is essential to the ongoing development of knowledge in a given field. This distinction helps debunk the impression of superiority and inferiority of the different levels—each in its own right is essential for innovation in research. Furthermore, it is important that we continually seek to move innovative knowledge from one level to the next—towards sound application—thereby accumulating the necessary evidence for responsible practice.

Theory-building and theory-testing, and the need to move up and across levels of scientific evidence in educational research

Although the evidence-based trend is relatively widespread in education, its application has been the subject of numerous criticisms targeting the external validity of the studies

constituting the best evidence. Internal validity is the extent to which an empirical study establishes and clearly explains a relationship between an intervention and its outcome and external validity refers to the possibility of applying the conclusions of an empirical study outside the context of the study.

The main limitation of the hierarchy of scientific evidence is the external validity of the evidence (Joyce, 2019). Higher-level evidence aims to increase the internal validity of studies to better demonstrate the effect of an intervention, but the external validity of these studies remains limited (Orr, 2015). In psychosocial fields, including education, such similarities between entities are more difficult to demonstrate. Therefore, interventions are more likely to produce different results in groups, contexts, etc. Even meta-analyses are likely to introduce biases concerning the external validity of a body of research since they pool studies carried out in several contexts that are not necessarily comparable (Parkhurst & Abeysinghe, 2016).

Higher-level evidence can be very useful in determining the effects of an intervention at the practical level (Slavin, 2020), but many other levels of evidence are needed from a context-specific policymaking perspective. Particularly in an area like education, where practice is policy-based, aspects such as popular opinion on practices, social determinants of target groups, and other contextual variables are important to consider (Parkhurst & Abeysinghe, 2016). It is the accumulation of appropriate evidence at different levels that supports the use of an intervention; higher-level evidence is not always sufficient support due to lack of external validity. These criticisms can be addressed by stronger theory-building and theory-testing.

WILL IT WORK HERE AND NOW?

Answering the question “Will it work here and now?” amounts to demonstrating how the local context affords a causal pathway through which the intervention can make a positive contribution. While local effectiveness predictions will never be certain, incorporating this information in the reasoning supporting the implementation of evidence-based practices can improve them (Joyce & Cartwright, 2020). Ultimately, we do not just want to know if an intervention works, we want to know if it will work in the specific context for which it is intended to be used. This question implies a shift towards a context-focused approach to EBE (Joyce & Cartwright, 2020).

Proponents of EBE generally attribute the gap between research and practice to shortcomings in the way tasks are performed on one or both sides of the gap (Joyce & Cartwright, 2020). To the contrary, we argue that a major part of the necessary reasoning in EBE has been overlooked and not appropriately formalized. With this in mind, what appears to be lower-level evidence in the context of establishing what works generally becomes higher evidence in the context of establishing a fit with a local context. Such evidence, which includes ethnographic approaches or evidence from local surveys, is thus also needed to assemble a

body of evidence that supports the utility of an intervention in a specific context (Parkhurst & Abeysinghe, 2016).

A realist approach to the review and synthesis of evidence from the literature and to the evaluation of a given intervention's implementation seems particularly useful in answering the question "Will it work here?" A realist synthesis is a narrative summary focused on interpretive theory that applies a realistic philosophy to the synthesis of primary study results that affect a single research question. It uses an interpretive inter-case comparison to understand and explain how and why the observed results in the studies included in a literature review occurred (Wong et al., 2012). Realist evaluation provides a framework for understanding how the context and underlying mechanisms affect the outcomes of an intervention (Ericson et al., 2017). These methods were originally developed by Pawson and Tilley to evaluate complex intervention policies in health and social services (Pawson, 2006; Pawson et al., 2005; Pawson & Tilley, 1997). In a realist approach, data are collected and analyzed to determine context–mechanism–process effect configurations (Haynes et al., 2017). An explanation and understanding of the interaction between the context, the mechanism, and the impact of the intervention is then produced (Wong et al., 2012). This joint focus on context, mechanism, and process effect should overcome one crucial limitation of quantitative research: authors have argued that traditional study designs such as randomized controlled trials, and non-randomized and prospective cohort studies, although useful, overlook a key element, namely being able to identify contextual information that is useful when replicating the results in another context (Graham & McAleer, 2018).

In other words, the success of an intervention depends on how participants interact with it in local contexts (Haynes et al., 2017), and a realist approach should uncover these processes. The working hypothesis behind a realistic synthesis is that a particular intervention (or class of interventions) will trigger particular mechanisms somewhat differently in different contexts. In realism, it is the mechanisms that trigger change rather than the interventions themselves, and realistic reviews therefore focus on "families of mechanisms" rather than "families of interventions" (Pawson, 2002, as cited in Wong et al., 2012, p.94).

Levels of contextual fitting

Table 2 provides a classification of the contextual fitting of effective interventions based on scientific evidence. Akin to the previous levels of evidence, this proposal of criteria allows us to distinguish 1) information of a pseudoscientific/non-scientific nature, 2) the results emanating from a scientific approach, and 3) the probative level at which the relative convergence and divergence of results is uncovered based on a thorough literature review. The facts needed to improve the level of contextual fitting come from empirical research, observations, and credible theory.

Range	Level	Evidence required	Main limitations
Probative	1	Realist review	
Scientific	2	Qualitative research during implementation work	Correspondence between studied population/context established for the target population, but without taking into account all contextualized elements from the literature
	3	Qualitative research during experimental work	Correspondence between studied population/context established only from the population studied
	4	Exclusive reliance on general effectiveness claims	Correspondence between studied population/context unestablished
Pseudoscientific/non-scientific	5	Exclusive reliance on causal ascriptions	Based on arbitrary[1] choices among "what works"

Table 2 *Levels of contextual fitting applied to educational research*

[1] *"Arbitrary" includes, but is not restricted to, epistemological biases, personal preferences, emphasizing the latest research, and more generally acting without the required information.*

Theory-building and theory-testing, and the need to move up and across levels of contextual fitting in educational research

Effectiveness predictions are obtained through the identification of contextual influences (Joyce & Cartwright, 2020). We specify this as a process of disaggregation of contextual influences. We also suggest that this process cumulatively leads to an increase in what we call levels of contextual fitting, which hinges considerably on theory-building—making increasingly reliable predictions about *what* might work for their school, their district, and their students, and *how* it might do so (Joyce & Cartwright, 2020).

CONCLUSION: THE FRAMEWORK AND ITS IMPLICATIONS

The three cumulative questions underlying the EBE3 framework imply a repositioning of sources of evidence, some of them traditionally considered the best (i.e. experimental studies) but now considered as a first step towards more powerful evidence. On the other hand, others previously neglected in EBE (qualitative research at large) are taking a prominent role. Indeed, articulating the two additional main ingredients posited in this research brief—general effectiveness claims and effectiveness predictions—in an effort to go beyond "what works"

leads to a new articulation of applied empirical research within a given educational field, as demonstrated in Figure 1.

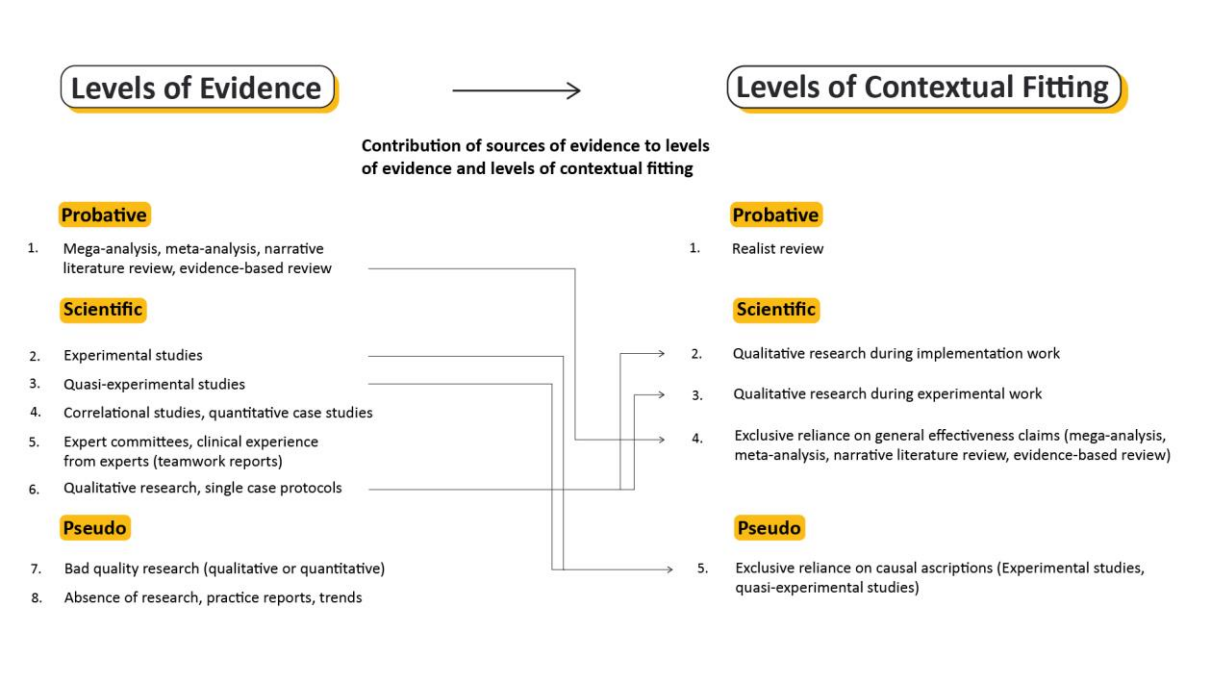

Figure 1 *Contribution of sources of evidence to levels of evidence and contextual fitting within the EBE3 framework*

A striking realization is that sources of evidence that leave something to be desired in terms of scientific evidence are among the best sources of evidence in terms of contextual fitting. Moreover, higher levels of evidence are largely insufficient in terms of contextual fitting. Finally, qualitative research, traditionally considered anecdotal in EBE, occupies a crucial role in improving the contextual fitting of the best evidence emanating from the highest levels of evidence. The proposed articulation of causal ascriptions, effectiveness generalizations, and effectiveness predictions generated by empirical research in education in the form of the EBE3 framework has implications for future research, policymaking, and improving educational practice.

With regard to the orientation of applied scientific research, the framework in Table 1 may shed light on the need for specific kinds of quantitative studies, meta-analyses, and synthetic work, as well as qualitative implementation work. Thus, it may help to bridge the perceived divide between quantitative and qualitative research in education by suggesting a sound integration of quantitative and qualitative methodologies around a common applied goal: to provide the necessary information to improve educational intervention. By reviewing and integrating the state of the art in EBE, it is clear that quantitative and qualitative research leverage each other to achieve the cumulative steps necessary for better intervention in a given domain.

In light of the importance of meta-analyses and systematic reviews in the need for effectiveness generalizations, it should be noted that the realist review process presented as a method for establishing effectiveness predictions can be reused to facilitate the automation of meta-analyses and enable living reviews of evidence. Indeed, realist reviews can be key in standardizing coding frameworks for studies, with common coding of cohorts, intervention delivery mechanisms, and core components. In addition, the framework presented in Table 2 helps to: focus research efforts in building local effectiveness predictions; outline various kinds of information that can improve predictions; and encourage the use of methods that are better equipped to acquire that information.

The framework presented in Tables 1 and 2 may contribute key ingredients to the mechanisms identified by Langer et al. (2016) that facilitate research use by policymakers. By insisting on a more complete scientific demonstration of efficacy, from causal ascriptions to effectiveness generalizations and effectiveness predictions, the framework may provide the materials for interventions that facilitate access to research evidence and build decision-makers' skills to access and make sense of evidence.

With regard to organizations and systems, the more complete scientific demonstration of efficacy outlined in Table 2 may help identify the right information for the right people, thus underlying the design of interventions that foster changes to decision-making structures and processes. Notably, an increased focus on core components, that is, mechanisms that represent active ingredients in interventions, can help policymakers avoid a-priori biases towards scientific disciplines that may seem compelling but do not provide the best explanations about how and why interventions work. The consequences of evidence-based reform operationally refined in this research brief could be profound. If educational policies begin to favour programs with clear evidence, publishers, software developers, university researchers, and entrepreneurs of all kinds will have an incentive to engage in serious development and evaluation. Governments, witnessing the cumulative impact of such research and development, might provide substantially greater funding for these activities in education.

Finally, practice should be greatly improved by a widened view of the necessary evidence in the implementation of so-called best practices, especially regarding effectiveness predictions. Effectiveness predictions help frame the practitioners' reasoning with regard to the match between general, abstracted evidence and their own specific and idiosyncratic context around a specific kind of inference that is amenable to analysis and testing in the context of day-to-day practice.

Evidence brokerage is also crucial to bridge the gap between research and practice (Langer et al., 2016). Because the EBE3 framework identifies the reasoning and supporting information for next-generation EBE, it could be used in information design to enhance the structure of evidence repositories and other resources. Langer et al. (2016) also concluded that interaction

among professionals can build a professional identity with common practices and standards of conduct, thereby fostering EBE. Using social influence and peer-to-peer interaction as catalysts, districts may be able to use support specialists (e.g. curriculum specialists, program specialists), and schools may be able to use onsite personnel, including literacy facilitators, or highly effective general or special education teachers (peers) as coaches. The focus could then be on those teachers who need follow-up support instead of providing the same support for all teachers across all professional development activities.

In sum, the EBE3 framework presented in this research brief may be one of the most integrative with respect to research traditions and the different roles (teachers, researchers, policymakers) involved in EBE. Future work should appraise the implications of such an integration—conceptually, operationally, and organisationally.

KEY MESSAGES

The results from a collection of high-quality studies comparing an experimental group given a target intervention with a control group (usually receiving business-as-usual teaching) has been the cornerstone of EBE for decades under the label "what works." It is the main, but not a sufficient, building block of EBE, and there is a need for a higher minimum standard for evidence of improved learning.

For a given educational issue, what is needed is a complete inventory of available interventions, rank-ordered in terms of relative efficacy to answer the question "What works best generally?"

An EBE initiative is not complete without solid indications that a specific context of application will enable the intervention that is working best overall to yield the expected benefits, which will answer the question "Will it work here?"

KEY RECOMMENDATIONS

The potential of the EBE3 framework to go beyond "what works" will be fully realized by:

- *emphasizing effectiveness generalizations through the production of meta-analytic work as soon as there are enough published experimental studies on a given issue;*
- *emphasizing effectiveness predictions by undertaking qualitative work regarding effectiveness predictions in given contexts as soon as meta-analytic results are available.*

The potential of the EBE3 framework to augment the cohesion of applied empirical work on a given issue will be fulfilled by:

- *focusing on theory-building and theory-testing in empirical studies, despite the applied nature of educational research;*
- *aligning the goals/research questions in quantitative and qualitative research with the maturity of a field, in order to optimize outcomes with respect to sound application in educational intervention.*

REFERENCES

Brighouse, H., Ladd, H. F., Loeb, S., & Swift, A. (2018). *Educational goods: Values, evidence, and decision making*. University of Chicago Press.

Burns, P. B., Rohrich, R. J., & Chung, K. C. (2011). The levels of evidence and their role in evidence-based medicine. *Plastic Reconstruction Surgery*, *128*(1), 305–310. https://doi:10.1097/PRS.0b013e318219c171

Cochrane, A. L. (1972). *Effectiveness and efficiency: Random reflections on health services*. Nuffield Trust.

Colquitt, J. A., & Zapata-Phelan, C. P. (2007). Trends in theory building and theory testing : A five-decade study of the Academy of Management Journal. *Academy of Management Journal*, *50*(6), 1281–1303. https://doi.org/10.5465/amj.2007.28165855

Connolly, P., Keenan, C. & Urbanska, K. (2018) The trials of evidence-based practice in education: A systematic review of randomised controlled trials in education research 1980–2016. *Educational Research, 60*(3), 276–291. https://doi:10.1080/00131881.2018.1493353

Ericson, A., Löfgren, S., Bolinder, G., Reeves, S., Kitto, S., & Masiello, I. (2017) Interprofessional education in a student-led emergency department: A realist evaluation. *Journal of Interprofessional Care, 31*(2), 199–206, https://doi:10.1080/13561820.2016.1250726

Evans, D. (2003). Hierarchy of evidence: A framework for ranking evidence evaluating healthcare interventions. *Journal of Clinical Nursing*, 12, 77–84.

Gage, N. A., Cook, B. G., & Reichow, B. (2017). Publication bias in special education meta-analyses. *Exceptional Children*, *83*(4), 428–445.

Graham, A. C., & McAleer, S. (2018). An overview of realist evaluation for simulation-based education. *Advances in Simulation, 3*(13). https://doi:10.1186/s41077-018-0073-6

Hansson, S. O. (2009) Cutting the Gordian Knot of demarcation. *International Studies in the Philosophy of Science, 23*(3), 237–243, https://doi:10.1080/02698590903196007

Haynes, A., Brennan, S., Redman, S., Williamson, A., Makkar, S. R., Gallego, G., & Butow, P. (2017). Policymakers' experience of a capacity-building intervention designed to increase their use of research: A realist process evaluation. *Health Research Policy and Systems, 15*(99). https://doi.10.1186/s12961-017-0234-4

Joyce, K. E. (2019). The key role of representativeness in evidence-based education. *Educational Research and Evaluation*, *25*(1–2), 43–62. https://doi:10.1080/13803611.2019.1617989

Joyce, K., & Cartwright, N. (2020). Bridging the gap between research and practice: Predicting what will work locally. *American Educational Research Journal, 57*(3), 1045–1082. https://doi:10.3102/0002831219866687

Langer, L., Tripney, J., & Gough, D. (2016*). The science of using science: Researching the use of research evidence in decision-making.* EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.

National Research Council. (2002). *Scientific research in education*. National Academies Press. https://doi.org/10.17226/10236.

Orr, L. L. (2015). 2014 Rossi Award lecture: Beyond internal validity. *Evaluation Review*, 39(2), 167–178. https://doi:10.1177/0193841X15573659

Parkhurst, J. O., & Abeysinghe, S. (2016). What constitutes "good" evidence for public health and social policy-making? From hierarchies to appropriateness. *Social Epistemology*, *30*(5–6), 665–679. https://doi.org/10.1080/02691728.2016.1172365

Pawson, R. (2006). Evidence-based policy: A realist perspective. Sage Publications.

Pawson R., & Tilley N. (1997). *Realistic evaluation*. Sage Publications.

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2005). Realist review: A new method of systematic review designed for complex policy interventions. *Journal of Health Services Research & Policy*, *10*, 21–34.

Slavin, R. E. (2020) How evidence-based reform will transform research and practice in education. *Educational Psychologist, 55*(1), 21–31, https://doi:10.1080/00461520.2019.1611432

Wong, G., Greenhalgh, T., Westhorp, G., & Pawson, R. (2012). Realist methods in medical education research: What are they and what can they contribute? *Medical Education, 46*, 89–96. https://doi.org/10.1111/j.1365-2923.2011.04045.x

Note: This text rests on many more academic shoulders than the references listed above. The research brief is a summary of Chapter 1 from the ISEE Assessment Data and Evidence Group. The ISEEA Report will be published in 2022, with more than 100 references.

International Science and Evidence-based Education Assessment

ISEE Assessment Advisory Board

Advisory Board Co-Chairs

Kevan Collins, *Chair, Youth Endowment Fund, UK*
Najat Vallaud-Belkacem, *France Director, ONE; Former Minister of Education, France*

Advisory Board Members

Anurag Behar, *CEO, Azim Premji Foundation, India*
Peje Emilsson, *Founder, Kunskapsskolan Education Sweden AB, Sweden*
Mary Helen Immordino-Yang, *Professor of Education, Psychology and Neuroscience, University of Southern California, USA*
Roza Otunbayeva, *Former President of Kyrgyzstan*
Rémi Quirion, *Chief Scientist of Quebec, Canada*

Ex-Officio Members

Anantha K. Duraiappah, *Director, UNESCO MGIEP*
Nienke van Atteveldt, *Professor, Vrije Universiteit Amsterdam, The Netherlands*

ISEE Assessment Panel

Assessment Co-Chairs

Anantha K. Duraiappah, *Director, UNESCO MGIEP*
Nienke van Atteveldt, *Professor, Vrije Universiteit Amsterdam, The Netherlands*

Working Group Co-Chairs

Working Group 1 on Human Flourishing

Nandini Chatterjee Singh, *Senior National Project Officer, UNESCO MGIEP*
Oren Ergas, *Senior Lecturer, Faculty of Education, Beit Berl College, Israel*
Tal Gilead, *Senior lecturer, Seymour Fox School of Education, Hebrew University of Jerusalem, Israel*

Working Group 2 on Context

Kenneth Pugh, *President and Director of Research, Haskins Laboratories, USA*
Edward Vickers, *Professor of Comparative Education, Kyushu University, Japan*
Latika Gupta, *Assistant Professor, Delhi University's Central Institute of Education, India*

Working Group 3 on Learning Experience

Stanley T. Asah, *Associate Professor, College of the Environment, University of Washington, Seattle, USA*
Gregoire Borst, *Professor of Developmental Psychology and Cognitive Neuroscience of Education, University Paris Descartes, France*
Stephanie Bugden, *Post-doctoral Fellow, Psychology Department, University of Pennsylvania, USA*

Data and Evidence Group

Stephen Fraser, *Deputy Chief Executive, Education Endowment Foundation (EEF), UK*
Julien Mercier, *Director, NeuroLab (FCI), University of Quebec, Montreal, Canada*

ISEEA Fellows

Joanne Marieke Buil *(PhD, Vrije Universiteit Amsterdam, The Netherlands)*
Anya Chakraborty *(PhD, University of Reading)*
Adriano Linzarini *(PhD, Paris-Descartes University)*
Thomas Macintyre *(PhD, Wageningen University)*
Rebecca Merkley *(PhD, University of Oxford)*
Moses Oladele Ogunniran *(PhD, Beijing Normal University)*
Mohsen Saadatmand *(PhD, University of Helsinki)*
Rongxiu Wu *(PhD, University of Kentucky)*

Review Editors

Drew H. Bailey, *Associate Professor, School of Education, University of California, Irvine, USA*
Iroise Dumontheil, *Reader in Cognitive Neuroscience; Director of MSc/MA Educational Neuroscience, Department of Psychological Sciences, Birkbeck, University of London, UK*
Roland Grabner, *Professor of Educational Neuroscience, Institute of Psychology, University of Graz, Austria*
Greg W. Misiaszek, *Associate Professor, Institute of Education Theories, Beijing Normal University, China*
Elaine Unterhalter, *Professor of Education and International Development, UCL Institute of Education, University College London*
Anne-Laura van Harmelen, *Professor of Brain, Security and Resilience, Institute of Education and Child Studies, Leiden University, The Netherlands*

ISEE Assessment Secretariat

Yoko Mochizuki, *Head of Policy, UNESCO MGIEP*
Shailly Gupta, *Assistant Project Coordinator, UNESCO MGIEP*

The ISEE Assessment

The International Science and Evidence-based Education (ISEE) Assessment is an initiative of the UNESCO Mahatma Gandhi Institute of Education for Peace and Sustainable Development (MGIEP), conceived as its contribution to the Futures of Education process launched by UNESCO Paris in September 2019. In order to contribute to re-envisioning the future of education with a scientific and evidence-based report, UNESCO MGIEP embarked on an ambitious project of the first-ever large-scale assessment of knowledge in education.

The overall goal of the ISEE Assessment is to pool multi-disciplinary expertise in educational systems and reforms from a range of stakeholders in an open and inclusive manner and undertake a scientifically robust and evidence-based assessment that can inform education policymaking at all levels and scales. It aims not to be policy prescriptive but to provide policy-relevant information and recommendations to improve education systems and the way we organize learning in formal and informal settings. It is also meant to identify information gaps and priorities for future research in the field of education.

In the education sector, the term "assessment" generally refers to activities used to measure student progress. Going beyond this narrow notion of education assessment, and drawing lessons from the IPCC Assessment Reports and other scientific environmental assessments (such as Millennium Ecosystem Assessment and IPBES), UNESCO MGIEP aspires to initiate a **scientifically credible, legitimate, relevant, and inclusive** process that will assess the state of education as a complex system, in addition to its role in achieving sustainable and peaceful societies.

The ISEE Assessment uses the 1996 Delors report's four pillars of education—*Learning to be*, *Learning to know*, *Learning to do*, and *Learning to live together*—as evaluative benchmarks and the lenses of "what," "where," "when," and "how" we learn and teach. The assessment is designed according to three working groups: (1) Human Flourishing, Education, and Learning; (2) Education, Learning, and Context; and (3) Learning Experience. In addition, there is a technical group on Data and Evidence. The ISEE Assessment Report is expected to be released in 2022.

Contact: ISEE Assessment Secretariat
issea@unesco.org