



## When and What Should We Measure? Evidence for Educational Decisions

Francois Lombard

Université de Genève, Switzerland

**Abstract:** Educational guidance requires evidence, but different educational decisions call for different kinds of evidence. What may count as good evidence for choosing a curriculum or pedagogy at system or departmental level may be poorly suited to decisions about teaching, responding to student difficulties, refining a learning design, or understanding how specific design features shape learning in practice. Because the value of assessment depends on the decision it is meant to inform, this editorial argues that evidence-based education should not be tied to a single methodological model. The focus is on pedagogical guidance, design refinement, and the interpretation of learning in context. For such purposes, evidence gathered only before and after instruction is often insufficient. Learning may be heterogeneous, non-linear, and unevenly responsive to specific design features; what matters for teachers, designers, and researchers is therefore not only whether an intervention works on average, but how students' understanding develops during the learning process itself. The editorial highlights the value of timely, fine-grained evidence drawn from classroom activity, student productions, observation, and digital traces. Such evidence can make learning visible while there is still time to interpret it and respond to it. Broadening the evidentiary repertoire of educational research and practice may therefore strengthen the ambitions of evidence-based education.

**Keywords:** Design-based research, Educational decision-making, Evidence-based education, Formative assessment, Learning in progress, Learning traces

**DOI:** <https://doi.org/10.31757/euer.921>

### Introduction

Few would dispute that learners often benefit from well-structured guidance, even if its most appropriate forms remain debated (Tobias & Duffy, 2009). Educational actors, whether in curriculum or pedagogy design, assessment, lesson preparation, or classroom interaction, regularly make decisions intended to support learning. Such decisions require some basis for judgment: data, evidence, or at least interpretable signs of how learning is proceeding. Yet not all educational decisions call for the same kinds of evidence. As Mottier Lopez (2015) reminds us, the value of any assessment depends on the decision it is meant to inform. Accordingly, methods in educational research need to be judged in relation to the questions they address and the phenomena they seek to illuminate (National Research Council, 2002; Taber, 2019). The issue, then, is not whether one method should replace another, but what kind of evidence is most relevant to a given educational purpose.

This editorial focuses on educational decisions that concern, on the one hand, the design and improvement of learning environments and, on the other, the guidance of students while learning is underway. Such decisions arise in everyday teaching, local or larger-scale instructional improvement, action research, and design-based research. They are typically taken in naturalistic settings, where learning is embedded in complex contexts and where many relevant variables cannot realistically be isolated or controlled. In these contexts, the central question is often not simply how one whole intervention (direct instruction, inquiry, problem-based learning, and so on) compares on

average with another, but how particular features of a design matter in practice and what this implies for subsequent guidance or redesign. For such purposes, educational research and practice need evidence that is timely and fine-grained enough to inform interpretation, adjustment, and support. As Duschl et al. (2021) note, observations and measurements do not directly constitute evidence. They become data, and data become evidence through selection and interpretation, guided by a framework and oriented toward a particular question or decision.

### **Why learning in progress needs to be documented**

For teachers, if the educational decisions at stake concern guidance, then evidence is needed before the end of a sequence. Learning does not necessarily proceed in a linear way, nor do all students follow the same path at the same pace. It may develop unevenly, with advances, plateaus, misunderstandings, or reorientations that remain invisible in endpoint measures alone (Lombard et al., 2026a). For educational designers and researchers, likewise, some design features may fail to produce the expected effects, may generate unexpected positive effects, or may do so only after some delay; others may matter only for particular learners or in particular contexts. Under such conditions, it becomes especially important to make visible the finer variations that matter for understanding how a design is functioning and how guidance or redesign might be adjusted. This is why educational research and practice benefit from drawing on a wider range of methods capable of capturing learning in progress rather than only its final outcomes. Such evidence is especially valuable when the aim is to support learning while there is still time to do so, and may then be the most useful evidence available.

A further reason concerns the kind of variation educational inquiry must attend to. Large-scale comparison studies can provide powerful evidence when the aim is to estimate average differences between interventions. Yet an exclusive focus on “what works” can obscure the fact that education is not a matter of simple implementation, but of continuous judgment and adjustment in context (Biesta, 2007). As Taber (2018) notes, variation within conditions may sometimes exceed the average difference between them. What works well on average may therefore work unevenly across teachers, classes, and settings, as well as across learners who differ in prior knowledge, interests, dispositions, or ways of engaging with a task. The same applies within a learning design: some elements may be especially productive, others less so, and their effects may be blurred when only aggregate outcomes are considered. This does not reduce the value of such studies for the purposes they serve; rather, it underlines the importance of additional forms of evidence when the aim is to understand learning in context, refine designs, or guide ongoing activity. Educational inquiry often involves trade-offs between validity and reliability and, as Taber (2018) notes, in education what can be measured most precisely is not always what is most important for informing the next decision. For decisions about guidance and design refinement, fitness for purpose may matter more than maximizing repeatability across contexts or occasions. In such cases, what often matters most is how students are understanding, revising, explaining, or stalling while learning is still underway.

### **Where data on learning in progress can be found**

Data on learning in progress may come from several sources, including classroom observation, student productions, and traces generated in digital learning environments. When tasks, exchanges, or evolving written responses are produced on a platform, time-stamped and student-identified traces may offer insight into how learners progress through assignments, respond to resources, and interact with peers, often with less intrusion than many traditional data-collection methods. They may also reveal individual trajectories of understanding that would remain hidden in endpoint scores or averages, and do so at a stage when such evidence can still inform pedagogical intervention or design adjustment within the limits of classroom time and teacher availability. For example, across a sequence of short written explanations in a biology inquiry, one student may move from “Pollen causes allergy” to “When our body detects pollen, it releases histamine, which causes sneezing and a runny nose,” and only later to an explanation linking recognition of pollen allergens, mast-cell activation, histamine release, and the resulting symptoms. Another student may use the right terms from the outset yet never establish the relations that make them explanatory. Such traces can become evidence when interpreted through analytic frameworks capable of identifying meaningful units, relations, and patterns in students’ activity and productions (Lombard et al., 2026b).

Fine-grained data, and the evidence that can be built from it, are especially important in the heterogeneous classrooms typical of most teaching contexts, where learners do not engage with the same tasks in the same way, at the same pace, or with the same conceptual resources and difficulties. What may appear as a minor difference in wording can therefore signal a substantial difference in understanding, and one that matters for both pedagogical guidance and design refinement. Fine-grained evidence can thus help reveal heterogeneity not as noise around an average, but as a normal condition of learning. For teachers, this matters because broad categories often do little to indicate what to do next, whereas timely and fine-grained evidence can make visible how particular students or groups are engaging, where understanding is diverging, and when support may be most useful.

### **Toward a broader evidentiary repertoire**

The issue, then, is not whether education should be evidence-based, but whether the evidence we privilege is fit for the decisions we hope it will inform. Educational research has understandably valued designs that can support comparison, generalization, and claims about average effects. But educational practice, design work, and many forms of research, including action research and design-based research, also depend on a different family of judgments: judgments about how learning is unfolding, how particular features of a design are being taken up, and where support or revision may still matter.

Although the examples discussed here mainly concern conceptual understanding, the same perspective may also inform the study and support of other educational aims, including critical thinking, scientific methods literacy, and some aspects of students’ understanding of science itself (Lombard et al., 2020). At this finer grain, it becomes possible to examine not only conceptual relations, but also the ways students justify claims, interpret data and

evidence, mobilize epistemic criteria, or bring prior beliefs and values to bear on what they make of scientific knowledge and its production.

Seen in this light, learning in progress is not a marginal concern. It is one of the places where data can be transformed into evidence that is educationally actionable. To attend to it is not to reject outcome measures, but to recognize that some of the most important decisions in education depend less on what can be shown after the fact than on what can be interpreted in time. A broader evidentiary repertoire is therefore not a methodological indulgence. It is part of taking seriously the actual work of teaching, design, and educational improvement. As AI creates new possibilities for analyzing finer-grained traces of learning, this issue may become even more pressing: the challenge will not only be technical, but educational, requiring that the analytic frameworks guiding such interpretations remain aligned with the aims and values of education rather than primarily with external commercial interests.

### Acknowledgments

I am grateful to Patrice Potvin for his careful reading of an earlier draft and for comments and suggestions that helped sharpen the argument.

### References

- Biesta, G. (2007). Why “what works” won’t work: Evidence-based practice and the democratic deficit in educational research. *Educational Theory*, 57(1), 1–22. <https://doi.org/10.1111/j.1741-5446.2006.00241.x>
- Duschl, R., Avraamidou, L., & Azevedo, N. H. (2021). Data-texts in the sciences. *Science & Education*, 30(5), 1159-1181. <https://doi.org/10.1007/s11191-021-00225-y>
- Lombard, F., Schneider, D. K., Merminod, M., & Weiss, L. (2020). Balancing emotion and reason to develop critical thinking about popularized neurosciences. *Science & Education*, 29(5), 1139-1176. <https://doi.org/10.1007/s11191-020-00154-2>
- Lombard, F., Merminod, M., & Weiss, L. (2026a). Is learning always linear, uniform, and parallel to teaching? Rethinking common assumptions about learning. In O. Morin, C. Bruguière, & M. Hammann (Eds.), *Challenges of a changing world in biology education* (pp. 289-306). Springer Nature Switzerland. [https://doi.org/10.1007/978-3-032-05346-6\\_19](https://doi.org/10.1007/978-3-032-05346-6_19)
- Lombard, F., Sudriès, M., Larpin, C., & Weiss, L. (2026b). Not wrong, but not a good answer: Assessing discipline-specific quality of student explanations. *Science & Education*. <https://doi.org/10.1007/s11191-026-00722-y>

Mottier Lopez, L. (2015). *Evaluations formative et certificative des apprentissages. Enjeux pour l'enseignement*. De Boeck.

National Research Council. (2002). *Scientific research in education*. National Academies Press.

<https://doi.org/10.17226/10236>

Taber, K. S. (2018). *Masterclass in science education: Transforming teaching and learning*. Bloomsbury Academic.

<https://doi.org/10.5040/9781474289450>

Taber, K. S. (2019). Experimental research into teaching innovations: Responding to methodological and ethical challenges. *Studies in Science Education*, 55(1), 69–119. <https://doi.org/10.1080/03057267.2019.1658058>

Tobias, S., & Duffy, T. M. (Eds.). (2009). *Constructivist instruction: Success or failure?* Routledge.

**Please Cite:** Lombard, F. (2026). When and What Should We Measure? Evidence for Educational Decisions. *The European Educational Researcher*, 9(2), 1–4. DOI: <https://doi.org/10.31757/euer.921>