

Education Review

Reseñas Educativas



Resenhas Educativas

October 1, 2025

ISSN 1094-5296

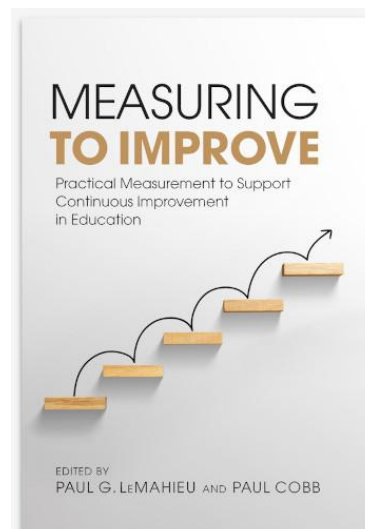
LeMahieu, P. G. & Cobb, P. (Eds.) (2025). *Measuring to improve: Practical measurement to support continuous improvement in education*. Harvard Education Press.

280 pp.

ISBN: 9781682539675

Reviewed by Brad McMillen
Wake County (NC) Public School System
United States

Editors Paul LeMahieu and Paul Cobb are well situated to offer advice on the relationship between educational measurement and school improvement. LeMahieu is a senior advisor to the president of the Carnegie Foundation for the Advancement of Teaching, on the graduate faculty in education at the University of Hawai'i at Mānoa, and former superintendent of education for the State of Hawai'i. Cobb is professor emeritus at Vanderbilt University. His work focused on improving the quality of mathematics teaching and student learning on a large scale.



The opening chapter, written by LeMahieu, Angel Yee-lam Li, and Cobb, provides a good framing for the book overall and defines “practical measurement”, a concept that aligns well with the burgeoning field of research-practice partnerships in education. In some ways, the research-practice partnership is contrary to the traditional expert-practitioner paradigm. The authors suggest that practical measurements are:

- tied to a clear theory of improvement;
- accepted as good measures by practitioners;
- actionable, in that the results lead to clear implications for improving practice;
- minimally intrusive and minimally burdensome to collect; and
- timely with respect to the turnaround of results.

The authors also articulate some of the myths that exist in the field regarding practical measurement, including a good reminder of Campbell's Law¹ (or Goodhart's Law from an economics perspective). The next five chapters (2–6) share case studies from various K-12 contexts in which practical measurement has been

¹ The more a quantitative social indicator is used for social decision-making, the more it is subject to corruption pressures and the more it will distort and corrupt the social processes it is intended to monitor.

developed and deployed effectively in support of plan-do-study-act cycles of improvement. Some of the cases represent classroom-focused instructional improvement work, while others operate at a more systemic and organizational level.

Sola Takahashi and Jon Norman provide the first of the organizational/structural studies in the book in Chapter 2, which serves as a good illustration of each of the five tenets of practical measurement from Chapter 1. The focus of the study was to improve Free Application for Student Aid completion rates at a California high school serving a high proportion of students from underserved populations who were not pursuing post-secondary education at a high rate. The focus on data visualization and power dynamics/inclusive sensemaking processes is particularly strong. This chapter also illustrates the importance of having explicit protocols and structures for data interpretation. Hearing more about the specific challenges experienced in the work and how they were overcome would have been a welcome addition; however, nonetheless, the case provides an informative, illustrative example.

The case study from Kara Jackson et al. in Chapter 3 is derived from a long-term research agenda focused on improving mathematics instruction across a network of school systems. The practical measurement strategies described include student surveys in addition to rubrics to evaluate the rigor of instructional tasks. This case was noteworthy in its use of multiple measures to triangulate data and for highlighting the use of iterative rounds of practitioner and student feedback to fine-tune their measures, illustrating the acceptability tenet of practical measurement put forth in Chapter 1. District staff helped the research team create survey items, collected pilot data, and engaged in cognitive interviews with students to refine their measures, using as many as five feedback cycles before conducting their primary data collection. Even the professional development the team provided to teachers was co-developed with district staff. This case is a great illustration of how practical measurement requires input from the individuals who are the focus of the study to avoid irrelevant variance in the resulting data. Moreover, it also illustrates the amount of time and energy sometimes needed to get it right. While both researchers and practitioners have significant demands on their time, it felt like the investment on the front end paid dividends later which is a good lesson for readers to take. This chapter also raises the issue, though not discussed explicitly, about how bespoke measures for specific interventions and contexts can increase measurement sensitivity to change. Although this may compromise generalizations to other contexts, with practical measurement/improvement work that is generally not the primary goal. The validity section is welcome, but it could have used some mention of the importance of reliability (esp. for observational measures) as a necessary precondition for validity.

The Chapter 4 case study by Linda Friedrich and Rachel Bear focuses on the implementation of the National Writing Project within a networked improvement community and provides a good illustration of the Chapter 1 tenet concerning a clear theory of improvement. The authors discuss the work necessary on the front end to calibrate teachers' thinking about what constitutes "argumentation" in writing, which resulted in a bit of a reset of the project in terms of data collection and professional development. They also describe the process of aligning teachers around the use of a

writing rubric by using common anchor responses to get clarity concerning scale points. Much like in Chapter 3, this case was instructionally focused and illustrated how practical measurement can be used to inform teachers about high-impact instructional strategies and key elements of student work. The authors also discuss the validity of the developed rubric in the context of perceptions of the perceived quality of students' writing, as well as through a correlational study with a separate outcome measure to demonstrate concurrent validity.

In Chapter 5, Elaine Allensworth describes work done at the Chicago Consortium to develop a freshman on-track indicator system for high schools, which was designed to facilitate early warning and intervention for students at risk for not graduating from high school. As another example of organizational-level practical measurement, this case illustrates the tremendous value of transactional and administrative data that schools already collect to drive improvement. This case in particular highlights the use of measurement to dispel often-held myths about cause-and-effect relationships. Early analysis during the project shows that ninth-grade student data regarding attendance and course failures were more accurate predictors of graduation than data related to student background characteristics and other internal factors, which challenged some status-quo thinking to pave the way for a common understanding of the theory of improvement.

The final case in Chapter 6 from Adrian Larbi-Cherif et al. illustrates the use of "exit tickets" to measure student perceptions of literacy instruction based on a comprehension task. The authors put forward four design principles that they feel can guide practical measurement based on their experiences: incorporating student voice; routine-embedded data collection (e.g., exit tickets were a quick and well-understood tool for teachers to use); designing measurements intentionally for equity (e.g., Spanish translation for English learners); and timely feedback (i.e., teachers received summary results the next day). These principles, similar to those promoted in Chapters 1 and 2, were then linked to key validity concepts.

In Chapter 7, Thomas Smith takes up the issue of validity in practical measurement across the collection of cases, delineating "validity-for-use" as a more classical understanding of validity argumentation as well as "validity-in-use," in line with more recent conceptualizations of consequential validity found in the research and professional standards literature. It contains a nice discussion of the tensions between practical measurement and classical validity strategies (e.g., domain coverage, abstraction to other contexts, and number of items) and outlines some key threats to validity-in-use for practical measures.



Paul G. LeMahieu

Chapter 8 is a wrap-up from Cobb and LeMahieu in which each case study is viewed in relation to the five principles from Chapter 1 and 2. Three key conditions are needed to make practical measurements useful in a continuous improvement cycle. The authors suggest that professional development, particularly in service of the first two principles of clear theory of improvement and acceptability, seems to be

key. They also highlight the need for expert coaches within the improvement cycle to help teachers with the sensemaking process and the “now what” instructional modifications that follow. Lastly, they highlight the key role that data systems and data representations can play. This last point is especially important, as it has implications for the actionable, minimally intrusive, and timeliness principles that underlie good practical measurement.

Overall this collection of case studies and the bookend chapters that provide the key through-lines is an excellent overview of the challenges of measurement for improvement at the ground level of K-12 education. The intense focus on high-stakes accountability measurement in education in recent decades has resulted in muscle atrophy in our field for practical measurement. Yet this book reminds us that those measurements are more proximal to learning, are often more sensitive to interventions, and can be operationalized in ways that accountability metrics cannot to support rapid-cycle improvements in schools. At the same time, the authors point out that much of what underlies practical measurement development is often just as rigorous and challenging as larger-scale accountability work. The only minor critiques I can offer are 1) I would have liked to have seen more discussion about addressing reliability concerns with the kinds of rubrics and surveys deployed in several of these cases, as it is difficult to think about validity without reliability; and 2) there was little discussion of the resources and staff time needed to support the development of data displays and dashboards to facilitate teachers’ timely and meaningful engagement with the data. Customized practical measures will require customized reporting technology which is often beyond the purview and budget of schools, so that is a potential impediment to practical measurement implementation that could prove difficult for others to pull off. Despite these two points, this book is a fantastic addition to the library for anyone engaged (or aspiring to engage) in applied school-based improvement work.



Paul Cobb

About the Reviewer

Brad McMillen is the Assistant Superintendent for Data, Research and Accountability with the Wake County (NC) Public School System. He has worked in educational research and assessment at the state and local school district level for more than 25 years and has published research on topics such as school size, year-round schools, and equity issues in course placements and related student outcomes. Brad has served in leadership roles for multiple organizations related to educational research and assessment including the North Carolina Association for Research in Education, Directors of Research and Evaluation, the American Educational Research Association, and the National Forum on Education Statistics. He also currently serves on the Board of Directors for the National Council for



Brad McMillen

Measurement in Education. He holds a master's and PhD in education from the University of North Carolina, Chapel Hill.



Education Review/ Reseñas Educativas/ Resenhas Educativas is supported by the Scholarly Communications Group at the Mary Lou Fulton College for Teaching and Learning Innovation, Arizona State University. Copyright is retained by the first or sole author, who grants right of first publication to the *Education Review*. Readers are free to copy, display, distribute, and adapt this article, as long as the work is attributed to the author(s) and *Education Review*, the changes are identified, and the same license applies to the derivative work. More details of this Creative Commons license are available at <https://creativecommons.org/licenses/by-sa/4.0/>.



Disclaimer: The views or opinions presented in book reviews are solely those of the author(s) and do not necessarily represent those of *Education Review*.