

***Beyond Testing:
Seven Assessments of Students and Schools
More Effective Than Standardized Tests***

Deborah Meier & Matthew Knoester

The title tells the reader much of what is in *Beyond Testing*. Meier and Knoester describe seven types of un-standardized “assessments” that they argue are better than standardized assessments. I put “assessments” in quotes because they use the term far more broadly than most in the field, seemingly to mean anything that might include discussion of student performance or school quality. The seven “assessments” they describe are student self-assessments, teacher observations of students and their work, descriptive reviews, reading and math interviews, portfolios and public defense of student work, school reviews by outside experts, and school board meetings and New England town meetings. The book also includes a brief discussion of the authors’ views of the role of education in a democracy and a short description of the New York Performance Standards Consortium, a group of 36 schools, mostly in New York City, inspired by and loosely modeled after Meier’s own Central Park East school. Woven throughout the book is a strong and one-sided criticism of both standardized testing and the current uses of testing, between which Meier and Koestner do not make a clear distinction.

While Meier and Koestner’s discussion of the limitations of standardized testing is weakened by one-sidedness and a lack of attention to research findings (as I will briefly explain below), it may nonetheless be helpful in an era in which many policymakers and educators stubbornly ignore these limitations. *Beyond Testing* offers a powerful argument that we need to broaden and refocus our notion of the goals of schooling, with a particularly timely emphasis on preparing students for lives as citizens in a democratic society. Meier and Koestner point to a number of potentially valuable ways to evaluate student learning. They also describe specific implementations of the seven “assessments,” which will help many readers understand what the authors’ suggestions entail for actual practice.

The diversity of Meier and Koestner’s set of “assessments” makes for interesting reading, but it also signals and contributes to one of the main weaknesses of the book. One of the core principles of assessment is that the best design depends on the intended use, but Meier and Koestner pay almost no attention to this. There are many points at which the use of assessments should be an issue in this book, but the most important for their core argument is perhaps the difference between assessments that require comparability and those that don’t. It has been amply documented that internal evaluations of student performance are often inconsistent, and one of the primary reasons for adding standardized measures to the mix of assessments is to address this problem. Originally, standardized assessments were designed in large part to provide classroom teachers with supplementary information that they could not obtain from their own evaluations, in part because of a lack of comparability, but we have increasingly relied on them to provide large-scale monitoring as well. For example, two of the most important aspects of trends in the achievement of American students are that while the Black-White and Hispanic-White gaps have been narrowing (albeit more slowly and erratically than we would like), the gap between rich and poor students has been widening. How do we know that? Standardized tests.

Although Meier and Koestner do not acknowledge this, some of their criticisms of standardized testing have long been described by proponents of standardized testing. For example, the limitations of the knowledge, skills, and dispositions that can be measured by standardized tests (and the concomitant principle that tests should be used as a supplement to and not as a replacement for teachers’ observations of their students’ work) have been discussed in detail in the measurement literature for well over half a century. That many policymakers and educators now routinely ignore this axiomatic principle is a criticism of the current misuses of testing, not of standardized tests. I always give my students a chapter published in 1951 in which E. F. Lindquist of the University of Iowa, one of

the most important developers and proponents of standardized testing in the history of the US, carefully described both the rationale for standardized testing and its many limitations. Those who are interested in Meier and Knoester's arguments about the relative value of standardized and un-standardized assessments would do well to start with Lindquist's more balanced and thorough chapter.

The one-sidedness of Meier and Koestner's argument is made apparent at the beginning of each chapter, where they list ways in which they consider the assessment in that chapter to be "more effective than standardized tests" (in general, not with respect to a given goal or use), without noting any limitations or any ways in which standardized measures may be superior. Examples of one-sidedness can be found throughout the book, often with a disregard for accumulated research evidence. One example is Meier and Koestner's treatment of measurement error. In several places, they excoriate standardized tests for having substantial measurement error. They are right to point out that the imprecision in the scores from even reliable standardized tests can be quite large. However, we have decades of research confirming what measurement theory predicts: scores from assessments comprising a small number of complex tasks are generally far less reliable than scores from standardized tests. A related issue: they criticize standardized tests for their limited sampling, but on many dimensions, complex performance tasks sample less per unit testing time. These weaknesses of performance assessment are among the important tradeoffs one faces in designing an assessment system, but Meier and Koestner do not acknowledge them.

Another example is Meier and Koestner's treatment of portfolio assessments. Research documenting the limitations of portfolios, including some conducted by me and my colleagues, extends back a quarter of a century. These limitations include issues of both reliability and validity, such as inconsistent and sometimes biased scoring, inconsistencies in the difficulty and novelty of tasks (which is confounded with differences in student performance), and confounding with assistance from peers and parents. These weaknesses do not imply that portfolios should never be used, but they impose constraints on their appropriate use, and educators planning to use them need to be aware of these tradeoffs. Meier and Koestner make no mention of them, instead focusing entirely on what they see as the benefits of a particular form of portfolio assessment.

The careful reader will find a number of unsupported claims and internal inconsistencies. For example, on page 23, in a section entitled "Local communities should decide what their students study in school," Meier and Koestner argue that with the exception of some basic skills, there isn't content that everyone should learn. They even argue that "there are competent adults who cannot name all 26 letters of the alphabet." Employers and postsecondary admissions officers might find this argument hard to swallow, and the authors themselves seem to contradict this notion a single page earlier, where they argue that "statistics should be a central part of the math curriculum at a much earlier age than it typically is."

Some of the core arguments and specific suggestions in *Beyond Testing* are timely as well as valuable given that the enactment of ESSA (the Every Student Succeeds Act) and growing parental dissatisfaction with current testing policies suggest a potential willingness to rethink how we assess both students and schools. To be clear, I agree with some of Meier and Koestner's arguments, and in fact, I suggest some of the same approaches, such as using expert external evaluators and placing more weight on a school's internal measures of achievement, in my own most recent book. Unfortunately, the book's lack of balance and disregard for research evidence lessen its potential to help move the nation toward a more reasonable and productive set of assessment methods.

Reference

Lindquist, E. F. (1951). Preliminary considerations in objective test construction. In E. F. Lindquist (Ed.), *Educational measurement*, 119–158. Washington, DC: American Council on Education.

Reviewed by Daniel Koretz in *Teacher College Record* January 08, 2018