

UC Berkeley

UC Berkeley Previously Published Works

Title

Classroom Assessment: Continuing the Discussion

Permalink

<https://escholarship.org/uc/item/4z59n9pg>

Journal

Educational Measurement Issues and Practice, 37(1)

ISSN

0731-1745

Author

Wilson, Mark

Publication Date

2018-03-01

DOI

10.1111/emip.12194

Peer reviewed

Classroom Assessment: Continuing the Discussion

Mark Wilson
UC Berkeley

Introduction

First, I would like to thank the Commentary writers for their thoughts and perspectives. Each of the Commentaries is an important piece in its own right, and I am both humbled and delighted to have four so thoughtful and experienced authors reacting to and enlarging upon the focus article. In this brief response, I will first give some brief background, then discuss some clarifications of points raised by the commentators. (Note, at many points, the Commentary writers make common remarks about both Focus articles, but I will, of course, only respond regarding my own.) After that I move on to a discussion of the commentator's observations regarding the core themes of the paper, then follow that up with some notes on matters that were not commented upon, and finish with a brief conclusion.

The commentators have (very appropriately) taken a broad focus for their views—addressing the larger perspectives inherent in the papers. However, in the case of my own paper, it was written with the intention of speaking in a focused way to NCME members rather than to the broader range of educators and educational researchers (as a direct consequence of it's being a NCME presidential address). And, as such, I have taken as its aim an exploration of the roles of assessment experts in education, thinking about what that role has been historically, and speculating on ways that it could be expanded in the future, particularly with the aim to put us in the position of improving education, rather than being cast, as we sometimes are, as being facilitators of the “bad guys.”

One interesting commonality across three of the four commentaries was reference to the National Research Council's "Knowing what students know" Committee report (NRC, 2001). This is surely no coincidence, as two of the four authors of the Focus articles were members of that Committee (Jim Pellegrino, who was co-Chair, and myself). More importantly, I see this as being a tribute to the pioneering work done by that Committee and its report. But, I will also mention one anecdote from the Committee's deliberations that I see as being relevant here. At an early meeting, the three self-confessed psychometricians on the committee roster, Bob Mislevy, Kadriye Ericikan and myself, decided that we wanted to "take the pulse" of the Committee membership regarding the main purpose of the Committee's work. So we asked the group's indulgence for an exercise where each member gave his or her idea about what was the source of the problem that the Committee should seek to solve. The results were, overall, quite plain—the problem was one that was created by the unseemly influence on the education system of assessment experts and (especially) psychometricians. Now, there may have been some pay-back on us for being so forward, but it was certainly true that the role of educational measurement in general (and hence, NCME and its members, in particular) was seen as being mostly negative. This was, for me, a quite daunting experience, and, although I was happy with our contributions to the eventual Committee report (which I think are reflected in the citations it received in the commentaries, and the many others besides), I believe I was still responding to it, so many years later in that presidential address.

Some clarifications

In this section I will consider a few points that I thought needed clarification, doubtless due to my own lack of success in clearly stating my aims and scope in the paper. One point that was noted was a distinction between the two Focus papers—whereas the paper by Shepard, Penuel and Pellegrino (2018) “emphasizes the importance of curriculum specificity to provide an interpretative framework for both instruction and assessment” (Marion, 2018), my account of the concept of learning progressions places them at a deeper level, perhaps more like a backbone for a curriculum (or indeed, a common backbone for multiple curriculums), and hence, the concern was that assessments developed according to a learning progression may lack the specificity of a particular curriculum. But I think this point needs to be expanded somewhat, as a learning progression could be the basis for *both* assessments proximal to a curriculum, and assessments that were designed to be distal from that curriculum. Indeed, there are contexts where different choices here would be appropriate, and, for example, one design tactic would be better in a formative context for the assessments and the other would likely be better in a summative context (see later for a discussion of these two terms). As Margaret Heritage (this volume) notes: “if curriculum is designed to support individualization by generally defining the order of learning experiences, with room for variations in the pace of learning, supported by assessment, it can honestly be said to represent the same expectations for all students, ... (Mosher & Heritage, 2017).”

A second point of clarification arises concerning definitions of classroom assessment and related topics. Susan Brookhart (2018) rightfully points up the distinctions among formative assessment, summative assessment and classroom

assessment as being important and worthy of attention and discussion. In my paper, I was reticent to give too strong a definition of classroom assessment, and hence used the views of others to indicate particular positions on this. Generally, I would like “classroom assessment” to remain a broad and inclusive concept—as assessment that is carried out under the aegis of the classroom and for classroom purposes. I do agree with the Black & William (1998) position that specific assessment events can end up in either the formative or summative camp (or even both): This is because they (i.e., Black and William) define the formative/summative difference as being related to the *usage* of the information from the assessment, rather than by some other (more superficial) feature. This would make classroom grading a summative type of classroom assessment (in most usages), while the assessments I was describing in my paper would most likely be used formatively, and hence be formative types of classroom assessment. Similar distinctions can be made about large-scale assessments, although, as I point out in the paper, I have grave concerns about some attempts to make formative use of large-scale assessments.

The core themes

All four commentaries were very clear in their identification of core themes in my paper: as Brookhart (2018) states “the primary connection between classroom and large-scale assessment isn’t measurement, but rather learning.” She also points out the crucial role of teachers in the process. I absolutely agree—the most important work of the assessment expert and the learning theory specialist is to help teachers in this crucial role—they must develop curriculum and assessment ideas and prepare materials that help teachers not only to implement these ideas, but also to adapt them and develop them

within their own classrooms. But it is what the teachers actually *do* in the classroom that is ultimately responsible for “learning” in our schools. According to Heritage (2018) these materials “should not lead teachers to discrete assessment events, but rather to ways of shedding light on students’ emergent, partial or even fragmentary understanding as they grapple with important ideas and analytic practices within the disciplines.” This is a wonderfully encapsulated account of how I would see a construct map helping a teacher in the classroom, beyond specific items etc. But, exemplary items and instructional descriptions associated with the construct map are also essential in this by giving teachers concrete links to what they might do in their classrooms.

The commentators make several points about these themes, and help bring important perspectives to the challenge for assessment. Wiliam (2018), for example, points out that “reverse engineering” items from existing achievement tests to create better assessments and instruction is not a sound road to success. This is a observation that I heartily agree with—and, in fact is one of the reasons for proposing the Bear Assessment System (BAS—as described in my paper) as a “from scratch” method of doing the opposite, engineering large-scale assessment starting from sound classroom assessment and instruction.

Several of the commentaries focus on the challenges of implementation in regular school settings. Brookhart (2018) notes that “both papers call for development work that is beyond the scope of a classroom teacher, or even a group of teachers, and beyond the scope of most commercial publishers of print or digital instructional materials.” Marion (2018) notes approvingly the following position from the Shepard et al., (2018) paper: “Implementing learning progressions through the creation of high-quality curricular

materials should be a district initiative.” In my own view, I’m not so sure that school districts are up to the basic development that is needed—I do think that they could be centers of adaptation of materials and dissemination of ideas. But, there needs to be a new alliance among key players in this effort—teachers, curriculum developers, learning science specialists, and assessment experts, engaged in the development of associated curriculum and assessment materials. Beyond that, administrators and policy experts, who are committed to change, and who are savvy at designing implementation plans that work, are needed. Of course, this is asking a lot; Wiliam (2018) notes the two Focus articles as giving “... a highly attractive vision...[but, the] radical nature of the vision means that its realization will require a great deal ...”. And, Marion (2018) ponders that they are: “so aspirational in fact, I question how either of these visions can be implemented at scale. “ I agree that the ideas are innovative and tremendously challenging, but without some vision of where we want to take our education system, and the assessments that are an integral part of it, we will never know whether we are moving in the right direction. My own experience is that teachers who have engaged deeply with the approach I described (such as those in the ADM project described in my paper) have found the experience both effective in their classrooms and professionally rewarding.

Some points not covered

There are some points in the paper (Wilson, 2018) that went somewhat un-noted in the commentaries, and I will say a few things about those too, as I think they are interesting and hopefully will not be overlooked by the reader. One such is the distinction between *information* and *signification* uses of assessments. An information use is a

usage that is derived from the results of the assessment, such as, say, a decision about the content of the next few lessons in a class, or whether a student seems to have succeeded in learning some particular concepts. These are what many would think of as the typical uses of assessments in education. But, indeed there are usages of assessments that do not depend directly on the results, and these I termed “signification” usages. For example, a teacher might give a brief assessment to the students before starting on the instructional aspects that the assessment is aimed at. Now, an informational use here would be to establish a baseline for the students in the class—how much do they know about that topic. But the teacher might also be directing the attention of the students on the items of the test to alert them to what is going to be important in the coming instruction—that is, if they don’t know how to respond to some items, then they should be particularly watchful when questions like those arise in the instruction¹. This is a situation where the assessment is being used to *signify* the importance of the erstwhile content of the assessment. In the paper, I used large-scale examples to make this point, but have used classroom examples here to show how the concepts are general to a broad range of contexts.

Even though information usages are probably the most commonly held idea about assessment, I made the comment in the paper that, in my view, signification usages tend to have a larger impact on educational systems. This seeming paradox is reflected in many journals in educational assessment and measurement, where the articles focus almost exclusively on information contexts. It may also be reflected in its absence from

¹ Of course, this tactic depends on a certain degree of sophistication among the students—what would happen, for instance, if the student did not realize that they couldn’t respond correctly to the item—this might lead to them being inattentive at a point where they need to pay a lot of attention!

the commentaries, even though the importance of signification usages is referred to several times in the commentaries (but without mentioning the distinction). I do believe that appreciating this distinction is particularly important if one wants educational measurement and assessment to (positively) influence education.

A second distinction that was submerged in the commentaries was that between a learning progression and a roadmap. I will not repeat the descriptions of each of these here, as they are given in some detail in the paper, but feel it is sufficient to note that a learning progression is a very complex concept that will involve “descriptions of states of student learning, instructional strategies to support this type of student learning, well-designed schemes of assessment that relate the states to student performances, and deep professional development that helps teachers to learn and master the new pedagogy of fostering student progress along the learning trajectory” (Wilson, 2018). But a roadmap is a much simpler projection of that complicated structure into a simpler structure that delineates just the assessment aspects of the learning progression (and is illustrated in the paper in in Figures 5 and 8). This perhaps will help explain several mentions in the Commentaries mentions of a lack of “curricula specificity” in the approach described in my paper: By focusing on the roadmap, I have avoided having to address these issues, but indeed this issue is still present, as these issues would have to be addressed in the accompanying learning progression. Now, one virtue of the idea of a roadmap is that it might be relatively robust to different curricula, and hence could provide a means of generalizing across curricula, but this is not something that would be guaranteed by design, but would need empirical investigation, by, say, examining DIF between the students involved in different curricula. It also raises the possibility of having a two-level

item-bank, one that is contextualized to specific curricula, and one that is contextualized in a more generic way.

Conclusion

The commentators have succeeded in expanding our perspectives and correcting misunderstandings, and they have enriched our thinking. My own aim in writing my presidential address was principally to foment discussion and evidence-based argument—and these commentaries are wonderful examples of that.

References

- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-147.
- Brookhart, S. M. (2018). Learning is the primary source of coherence in assessment. *Educational Measurement: Issues and Practice*, xx, y, n-k.
- Heritage, M. (2018). Making assessment work for teachers. *Educational Measurement: Issues and Practice*, xx, y, n-k.
- Marion, S. F. (2018). The opportunities and challenges of a systems approach to assessment. *Educational Measurement: Issues and Practice*, xx, y, n-k.
- Mosher, F., & Heritage, M. (2017). *A Hitchhiker's Guide to Thinking about Literacy, Learning Progressions, and Instruction*. CPRE Research Report #RR 20172. Philadelphia: Consortium for Policy Research in Education.
- National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Committee on the Foundations of Assessment. J. Pellegrino, N. Chudowsky, & R. Glaser (Eds). Washington, DC: National Academy Press.
- Shepard, L. A., Penuel, W. R., & Pellegrino, J. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement: Issues and Practice*, xx, y, n-k.
- Wiliam , D. (2018).How can assessment support learning? A response to Wilson, and Shepard, Penuel, and Pellegrino. *Educational Measurement: Issues and Practice*, xx, y, n-k.

Wilson, M. (2018). Making measurement important for education: The crucial role of classroom assessment. *Educational Measurement: Issues and Practice*, xx, y, n-k.