## Whether National Exam Scores are Student or Program Learning Outcomes

***January 12, 2016***

David Dirlam

Teaching to the national exam does not improve school performance. “No child left behind” has been abandoned. Higher Education assessment has been accused of not making progress in the last two decades.

In a fascinating posting to the [ASSESS Listserv](http://lsv.uky.edu/archives/assess.html) (Wed, 9 Dec 2015), Mary Herrington-Perry of Indiana State University posed a query that gets at the heart of the problem. **She asked whether to “earn a passing score on the national exam” is really a student learning outcome as opposed to a program outcome?**

Not all of the 14 replies responded to the important distinction between learning outcome and program outcome, but the consensus was that passing a national exam is a reasonable index of program performance that is independent of the local institution, but it needs to be supplemented. In an off-the-listserv response Catherine Wehlberg, AALHE President Elect and author of [*Promoting Integrated and Transformative Assessment*](http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470261358.html), added the distinction that its usefulness depends on whether the results lead to improvements in programs.

To capture the full import of the dialog, it is necessary to keep the focus on the tension between scales in education. Good teachers focus on the personal scales of students, lessons, and courses. Accreditation specialists and administrators focus on the aggregated scales of programs, institutions, and professions. Of course, there are interactions, since as Kelly McMichael of Mercer University and Keston Fulcher of James Madison University suggested, professions (on the aggregated scale) need to certify practitioners (persons). Assessment practitioners are caught in the middle, needing to help teachers improve programs so that institutions succeed. We aggregate the data, but if it does not get used at the personal scale of teacher-student interaction, then the effort is wasted.

Kelly McMichael, of Mercer University added the most detailed description of what kind of data should supplement national exam performance:

also combine other artifacts/sources-projects using faculty-devised rubrics or peer and faculty-reviewed group/lab projects-to substantiate learning. This triangulated approach reveals more of what is known as “authentic” assessment-where at certain points, the student is required to evaluate and articulate new information and problem-solving skills. Or as Mary Huba [and Jann Freed] [say], “…making connections between abilities and skills they have developed…or acquired in the major” (p. 41, [*Learner-Centered Assessment on College Campuses*](http://www.pearsonhighered.com/educator/product/LearnerCentered-Assessment-on-College-Campuses-Shifting-the-Focus-from-Teaching-to-Learning/9780205287383.page), 2000).

Linda Suskie, author of the well-known [*Assessing Student Learning, A Common Sense Guide*](http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470289643.html), provided the following analysis of Herrington-Perry’s listserv question based on her 7 traits of well-stated learning outcomes and identified whether national tests fit.

1. They’re outcomes (check).
2. They’re clear (check).
3. They use observable action words (check).
4. They focus on skills as well as conceptual understanding (not sure–depends on the exam).
5. They’re important (check).
6. They’re rigorous yet realistic (check).
7. They’re neither too broad nor too specific (too broad).

National exams can involve such a high level of aggregation that as Dennis Roberts of Pennsylvania State University pointed out, “What does it mean “to pass”? When it comes down to it … it’s arbitrary. So, there is nothing magical about a pass score” (ASSESS-L, Wed, 9 Dec 2015). I often compare such scores to evaluating a supermarket produce department by counting all the items in it: 6 watermelons plus 1,222 peas plus 85 oranges makes 1,313 items. There’s certainly “nothing magical” about that score. Furthermore, national exams are backwards looking, based on what is known, not what might be known. There is no room for the sort of authentic creativity Kelly McMichael alluded to that produces new information.

My interest in the educational usefulness of national exams was awakened over four decades ago by a group of campus school teachers at Plattsburgh State (NY), who were both exceptionally knowledgeable and much ahead of their time at using developmental research while interacting both one-on-one and whole-group with their students. As a faculty, they were adamantly opposed to standardized tests and paid little attention to the results. Nevertheless, the performance of their students on such tests was little short of astonishing, taking students who had often been rejected by public schools to some of the highest scoring class averages in New York State on the Regents’ Exams for reading, writing, and mathematics.

There were two factors that accounted for this success. One was weekly meetings aimed at identifying knowledge development. The teachers brought in work from their pre-K to 8th-grade students. We used the literature about development to make initial descriptions of developmental distinctions. The descriptions changed gradually based on improving agreement between independent raters with the question: “Did some raters miss something or do we need to improve the description?” These descriptions led to the first developmental rubrics – those used by raters of the New York State Regents Exam in Writing in the late 1970s.

The second factor for the campus school teachers’ success was their use of developmental indicators while interacting with students, planning lessons, and developing the curriculum. They held almost entire responsibility for lessons and curriculum, but were supplemented by Houghton-Mifflin’s Interaction Language Arts Program—a brilliant and engaging curriculum created under the direction of renowned expert in language arts development, James Moffett. Dorothy Hammond of the New York State Education Department’s Bureau of English spent over an hour reading a paper from each student in a fifth grade classroom and concluded that they could all pass the state high school exam in writing.

The bottom line is that the campus school faculty did not use the aggregated scores to improve their programs, rather they used the development of knowledge to make the improvements and the aggregated scores provided a simple communication of their results. But the students’ work itself provided a much more compelling answer than any single number per student could have.

Four decades later, well-known accreditation consultant, Peggy Maki, is working on a book on real-time assessment. Her goal, like that of Jim Moffett and the campus school teachers, is to get assessment back into the everyday minds of teachers and learners. If faculty members do assessment to comply with accreditation demands, they will abandon it as soon as those demands are realized. If they do it to improve their program, it will be an occasional (usually annual), practical activity. If they do it while developing courses, planning lessons, and interacting with their students, it will have astonishing and inspiring results.

The problem then becomes, how do we communicate such astonishing and inspiring results? I propose a revolutionary development in our paradigm and challenge the testing industry to transform its approach. Collectively, we must throw out the crutch of national exams with their sums of correct answers to questions of grossly uneven importance and ephemeral interests. In their place, we need to ask the industry, as well as administrators, accreditors, and education policy organizations to foster research on knowledge development that addresses each field represented in higher education and that helps educators like Kelly McMichael use their results to foster assessments that transform students. Learning outcomes that distinguish knowledge rather than count correct answers should be the focus of education, no matter the scale of our perspectives. These outcomes should not only reflect authentic work, but also be developmental—i.e., different at the beginning, middle, late, and years after programs. Programs should group and generalize these outcomes in the same way that we group and generalize the species of life, so that lesson and course outcomes connect to but are more specific than program outcomes. Curriculum maps should connect the lesson and course outcomes to those of the program. From an evaluator perspective, well-described outcomes should enable administrators, accreditation, and policy experts to randomly sample from student work to determine whether students have developed knowledge as a result of their experience in programs.

Leaving no learner behind does not happen with aggregated scores. Ultimately, education occurs between learners and teachers. Society benefits when knowledge develops, but such development always begins at the personal level.