The Gordon Commission on the Future of Assessment in Education has been a truly remarkable undertaking. Over the past two years, the 30 Commission members, supported by some 50 consultants, have produced an extraordinary body of papers, policy briefs, and reports. The work is fresh, sound, and well reasoned, offering a new vision of why, what, and how educational testing should be carried out—a vision in which the most important assessments are deeply connected to teaching and learning, and designed to support those processes. This vision places highest priority on assessment for education, while still attending to assessment of education. It is developed in full recognition that learning, including out-of-school learning, will look very different in tomorrow’s digital future from the way it looks today.

The Commission’s Technical Report, Public Policy Statement, and of course the individual papers comprised in the Knowledge Synthesis Project point to a future where assessment is carried out in new ways for new purposes. It would have been folly for the Commissioners to attempt to set forth a single, coherent picture of educational assessment in the middle of the 21st century. There are far too many unknowns. However, the rich materials produced by the Commission can serve to inform, stimulate, and support the imagination of anyone who wishes to attempt to create such a picture. To that end, in my brief time this afternoon, I would like to try to convey my own picture of an assessment future consonant with the vision implicit in the Commission’s work. I’ll begin with large-scale testing as it mostly is today, and then suggest how that state of affairs might change as the promise of the Gordon Commission’s vision is fulfilled.

**Testing Today**

Educational testing today is a very strange business. We socialize students about how to engage in an activity called “taking a test,” which is quite unlike any of their other day-to-day activities, either in school or out of school. Even though assessment tasks are intended to elicit knowledge, skills, and abilities that are valued in their own right or that support future school learning, most of these assessment tasks are themselves highly artificial, contrived activities. People smile at the assertion that “Life is not a multiple-choice test,” but the concern I take from the Gordon Commission papers is, rather, that “Multiple-choice tests are not life.” Students taking a test are answering questions, solving problems, or writing out their thoughts for no reason except to demonstrate that they are capable of doing so. They are performing,
for the teacher or for some more distant audience. Various adults in children’s lives struggle to communicate the seriousness of this testing endeavor. Because testing is so divorced from any authentic purpose or pursuit, parents and teachers must wrestle with the dilemma of inducing student effort and motivation without inducing undue stress and anxiety.

Test performance really is a performance. Like a play acted out on a stage, it is a simulation, not the real thing. After rehearsals, referred to as test preparation, the event itself is staged. It unfolds with pencils and booklets and answer sheets as props, beginning with the examiner reading some scripted lines. Most importantly, special rules constrain the social context of the activity. Help seeking is tightly controlled and peer interaction is generally forbidden. To borrow the phrase used by Bransford and Schwartz (1999), examinees engage in “sequestered problem solving,” demonstrating what they are able to accomplish not in a “smart workplace” or using “smart tools,” not as members of collaborative teams, but instead using a meager set of permissible resources, perhaps a calculator or a translating dictionary, and working in isolation. Some do well at this odd activity, some poorly, and for better or worse, the individual consequences of their test performances can be substantial.

We could ask, “Why do we do this?”, but it might be more productive to ask, “How is it that educational testing has reached this point in its evolution?” The answers are complex, of course, but I believe some key factors are first, a focus on comparability and fairness; second, a demand for efficiency; third, an historic interest in teaching and testing of factual knowledge; and fourth, the privileging of objectivity in the scoring and interpretation of test results. Let me speak to these four points in turn.

Comparability and Fairness

Comparability is at the heart of assessment. In order to compare persons to one another, to generate scores that are stable and interpretable, we standardize the situation in which the test takers are observed and the rules whereby their performances are interpreted. This seems reasonable because the definition of what a test measures—the meaning of the test performance—is thought to inhere in the test itself, not in the educational histories or the unique constellations of identities, experiences, goals, and motivations that each individual person brings to the testing situation. Each examinee interacts with same the test materials and receives a score. It seems self-evident that if they have taken the same test, then comparing their scores is unproblematical. Since all have been asked to do the same thing following the same directions, with the same time limit, and so forth, the test has been fair. This appearance of fairness is crucial if tests are to serve the major purposes for which they are now used. Testing must appear to be objective and impartial, even if that implies that what is measured diverges from what really matters.

Efficiency

One unavoidable requirement for any assessment is that there be some record of each examinee’s performance. This scorable record might be a piece of writing in response to a
prompt, a video recording of an impromptu speech, a lab notebook, a cabinet produced in a woodshop, a proof of a mathematical theorem, or a sheet of paper with darkened circles in precisely determined locations. Large scale educational testing was helped enormously by IBM’s mark sense scoring machines and E. F. Lindquist’s later invention of optical mark recognition scoring. The speed and accuracy of optical scanning contributed greatly to the widespread adoption of multiple-choice tests. Of course, it also helped that multiple-choice tests were viewed as objective—Teachers reading essays might be influenced by their own biases and opinions about good writing, but no such subjectivity threatened to distort multiple-choice test scores.

Focus on Factual Knowledge

Today’s testing technology also comports well with deeply rooted, popular views of school learning. In education circles, it may seem painfully naïve to equate the goal of schooling with mere acquisition of information, but imparting factual knowledge is still a huge part of what schooling is about, and a huge part of what the public expects of schools. Children are supposed to learn their multiplication tables, important dates, and how to spell. The person in the street wants schools to teach “the basics,” and if pressed for examples of what that means, is very likely to come up with learning outcomes that are perfectly amenable to evaluation using multiple-choice questions.

Objectivity

Finally, bound up with expectations of comparability and fairness as well as a focus on factual knowledge, the way we do educational testing seems legible. The rules of the game seem clear and make sense to people. Testing is supposed to be scientific, and science is supposed to be precise and objective. Scoring errors on large-scale high-stakes tests make national news because they are rare, and because the public has high expectations for the scientific accuracy and defensibility of test scores—and, I might add, a rather tenuous grasp of the concept of measurement error.

A Future Vision Inspired by the Gordon Commission

Now, drawing on the rich and provocative papers and reports produced by the Gordon Commission, let me offer my own imagining of a very different assessment future.

First of all, much assessment in the classroom will be truly integrated with instruction, based on student pursuits that are educationally useful and intrinsically meaningful in the classroom context. Assessment inferences may be informed by observations of the processes as well as the products of student learning activities. Children will work alone or together on engaging tasks, and much of their work will be supported by various technology-mediated systems. Records of students’ actions will be captured and analyzed to support high-level inferences about their reasoning and their evolving expertise.
A New View of Comparability and Fairness

Giving assessment for education priority over assessment of education will invite a focus on criterion-referenced rather than norm-referenced score interpretations. Assessments will show directly what students know or are able to do, not just how they compare to one another. In this context, comparability and fairness will take on new meanings. Students have different backgrounds, different talents, and different passions, and assessments will reflect the reality of variation in what different children are coming to know and accomplish. Even if comparison is no longer as simple as a rank ordering of numerical test scores, comparability need not be sacrificed. Rigorous criteria may be framed at a level of abstraction that allows for individual learning choices. We know today that students given the freedom to choose from a range of topics for research reports or creative writing assignments nonetheless produce works that can be evaluated using a common rubric. As children pursue a broader range of complex, meaningful tasks in school, there will be greater scope for this kind of flexibility.

A New View of Efficiency

The Gordon Commission offers an inspiring and believable vision of the implications of digital technology for assessment. When computers are used as a medium of instruction and communication, the old idea of a “scorable record” like a bubble sheet is completely transformed. New media for representing and interacting with knowledge are already reshaping the world of work for many adults, and as digital knowledge representation technologies become more prevalent in schools, it will become possible, in the words of the Commission, “to gather new forms of data based on human interaction in digital environments” (Gordon Commission, 2013, p. 13). Digital technologies will support data processing as well as data collection, yielding much richer, more detailed portrayals of the processes, as well as the products, of students’ academic work. Data streams generated in the course of students’ ongoing classroom activities may serve as the basis for assessments that today would be prohibitively expensive or simply impossible.

A New View of Schooling Outcomes

Generations of curriculum developers, educators, and policy makers have tried to shake the public loose from the notion that school learning is primarily about facts and basic literacy skills. The 1983 Nation at Risk report (National Commission on Excellence in Education, 1983) sent a wake-up call. Education Secretary Terrell Bell’s infamous “wall charts” pushed a similar message. The National Assessment Governing Board tried to stimulate concern by introducing the NAEP Achievement Levels. Lauren and Daniel Resnick’s highly influential 1992 chapter on “Assessing the Thinking Curriculum” offered a surprisingly contemporary vision of expanded learning outcomes, and linked their attainment directly to new forms of assessment (Resnick & Resnick, 1992). Lorrie Shepard (2000) argued the point cogently in her 2000 AERA Presidential Address on “The Role of Assessment in a Learning Culture.” The National Research Council’s report last year on Education for Life and Work observed that terms like “deeper learning,”
“21st century skills,” “college and career readiness,” “student centered learning,” “next
generation learning,” “new basic skills,” and “higher order thinking” “[reflect] important
dimensions of human competence that have been valuable for many centuries, rather than
skills that are suddenly new, unique, and valuable today” (National Research Council, 2012, p. 3).

The Gordon Commission echoes and expands upon these calls for a broader view of
learning outcomes. I believe it goes beyond earlier appeals in also presenting a more
sophisticated vision of assessment and assessment use, linked to the transformative power of
new digital technologies and capitalizing on the affordances of those technologies to bring a
more expansive vision of learning outcomes closer to realization.

As the Commissioners observe, “Modern information technologies afford student access
to almost limitless quantities and varieties of information resources. Competence in accessing
and utilizing available resources could replace the more traditional privileging of memory store”
(Gordon Commission, 2013, p. 16). New tools for knowledge representation and use, from the
abacus to the iPhone, amplify human abilities. At the same time as students’ work with new
tools makes available new forms of data for assessment, it also enables new cognitive
attainments as children, along with adults, become more powerful creators and users of
knowledge.

A New View of Objectivity

Finally, when assessment is primarily for education rather than of education, objectivity
becomes less important. The times are changing. As Professor Gordon wisely concludes in his
introduction to the executive summary of the Commission’s Technical Report, we are beginning
to understand that we need a contextualist and relativist science as new developments in
science, technology, and scientific imagination carry us beyond the limits of our familiar,
positivist epistemology. I believe that the foundational principle of assessment as evidentiary
argument linking observations to inferences and actions will endure, but patterns of reasoning
will be more complex and more particular to the circumstances of individual students.

This is an exciting time indeed for assessment. The Gordon Commission has pointed the
way toward a bold new future.

Thank you.
References


