Assessment, Teaching, and Learning

The Gordon Commission on the Future of Assessment in Education

Assessment, Teaching, and Learning (ATL) is Chairman Edmund W. Gordon's vehicle for the promotion of conversations concerning possible relationships of educational measurement and assessment to educational policy and practice. This instrument of communication reflects a bias that Gordon brings to the leadership of the Gordon Commission on the Future of Education — i.e., assessment and measurement in education should inform and improve the processes and the outcomes of teaching and learning, primarily, and may serve other purposes, such as accountability. In the title of this bulletin, the publisher assumes that assessment, teaching, and learning are the three legs upon which pedagogy stands, and that these are symbiotic and dialectically related processes that are as interdependent as they are reciprocally essential. Teaching cannot proceed — nor can it be fully understood — without assessment. It is questionable that teaching has occurred if learning does not result. Improved learning should be the primary purpose behind assessment and teaching. Assessment should be informed by the needs, ends, and processes of teaching and learning.

These perspectives have been anticipated by recent developments in the field of educational assessment and measurement. These developments are excellently reprised in *Knowing What Students Know (KWSK),* published by the National Research Council in 2001, a review of which provides the content for the current issue of *ATL.* The authors of *KWSK* came to several prescient conclusions. Among these are such commonsense notions as:

- The purpose of an assessment determines priorities, and the context of use imposes constraints on the design. Thus, it is essential to recognize that one type of assessment does not fit all.

- Often a single assessment is used for multiple purposes; in general, however, the more purposes a single assessment aims to serve, the more each purpose will be compromised.

- Assessment is always a process of reasoning from evidence, but assessment is, by its very nature, imprecise to some degree. Assessment results are estimates of what a person knows or can do.

- A model of cognition and learning should serve as the cornerstone of the assessment design process. This model should be based on the best available understanding of how students represent knowledge and develop competence.

- Widely used assessments of academic achievement are based on highly restrictive beliefs about learning and competence not fully in keeping with current knowledge about human cognition and learning.

In this issue of *ATL,* Dr. Ezekiel Dixon-Román of the University of Pennsylvania provides a commentary/review based on this important contribution. This provides a starting point for the work of the Gordon Commission and for what assessment in education will need to become.

Assessment, Teaching, and Learning is a bi-monthly bulletin that is the primary instrument of communication from the Chairman of the Gordon Commission to a broad audience of readers who are concerned with the relationships between psychometrics and education. The intent is to use this bulletin to stimulate conversation and debate concerning the multiple purposes of assessment in education; the possibilities for the improvement of teaching and learning processes and outcomes through the more creative use of measurement in education; visions of future change in the nature and practice of education; and the need for change in the capacity of the educational measurement enterprise necessary to the needs implicit in those visions. *Assessment, Teaching, and Learning* is available, without cost to the reader, electronically and in print. Subsequent issues will address: Assessment, Teaching, and Learning: The Pedagogical Troika; Systems of Educational Examinations; Curriculum Embedded and Proximal Assessments; Extant Varieties of Educational Measurement; and the Relational Analysis and Management of Educational Tests Data.

Assessment to Inform Teaching and Learning

By Ezekiel Dixon-Román

In a 2001 issue of *Pedagogical Inquiry and Praxis,* Chairman Gordon asserted that the affirmative development of academic ability should include "diagnostic, customized, and targeted assessment, instructional and remedial interventions" (p. 3). Later that year, the National Research Council (NRC) published a report titled *Knowing What Students Know,* which moved in promising directions toward his vision. This comprehensive report discussed the considerable advancements in integrating cognitive and measurement sciences and conceptualized assessment into the *formative* and *summative* domains.

Formative assessments, often instructionally embedded measurement methods, provide feedback to both the teacher and student on the student's learning and developmental progress. They inform instructional strategies, and can be used to assist in the development of a student's metacognitive and metacomponential competence. Summative assessments are large-scale assessments that focus on the most critical and central aspects of learning in a domain as defined by curriculum standards and informed by cognitive research. They are primarily used for post hoc decision making such as promotion, accountability, evaluation, and policy. In addition, summative assessments provide information for parents on their child's progress, and for teachers on students' instructional needs.

The contention here is that an aligned and developmental design of assessment, curriculum, and pedagogy will enable an equitable system of learning and development. The advancements and merging of the cognitive and measurement sciences, formative assessment, summative assessment, and an appropriate model for the development and design of a system of assessment are discussed.

Dixon-Román, an Assistant Professor of Social Policy at the University of Pennsylvania's School of Social Policy and Practice, is a member of the Gordon Commission. Dixon-Román does work on the intersections of the sociology of education, cultural studies, measurement, and policy. In 2005, he was a Harold Gulliksen Psychometric Research Fellow of Educational Testing Service and completed a two-year Institute of Education Sciences postdoctoral fellowship in the Institute for Policy Research at Northwestern University in 2008. Dixon-Román can be contacted at **ezekield@sp2.upenn.edu**.

Advances in the Cognitive and Measurement Sciences

As documented in the NRC report *How People Learn* (1999) and discussed in an earlier issue of *Pedagogical Inquiry and Praxis* (Bridglall, 2001), research evidence on learning and development from various disciplines has advanced the field of the cognitive sciences. The field has moved beyond traditional differential and behaviorist perspectives to more nuanced paradigms of learning such as cognitive and situative perspectives. The cognitive framework focuses on the development of structures of knowledge and procedures for reasoning and solving problems. The situative (or sociocultural) perspective goes beyond the cognitive focus on individual response to a task, describing behavior in relation to activity and context. The perspectives are not mutually exclusive, and need to be used in interaction with one another in assessment and teaching.

These framework developments refer to what is known about the acquisition of knowledge and the production of understanding. Advancements have been made in the distinction between working and long-term memory and how to move information from one to the other. We also know the importance of cognitive schemes and the organization of information, metacognition, the role of prior knowledge, and the importance of practice and feedback. Knowledge transfer and the role of the social context are additional, significant factors in pedagogical experiences. All of this research-grounded information has become integral in learning and development, which in turn require the design of assessment and teaching to be in that same spirit.

The field of psychometrics has made many advancements in its more than 100 years of existence,

particularly in response and alignment to the research and practice demands of psychology and education. Progress in measurement has developed beyond traditional classical test theory methods and factor analytic models to generalizability theory, item response theory, latent class models, unidimensional and multidimensional methods, and growth curve models or change analysis for continuous and categorical outcomes via structural equation or hierarchical linear modeling techniques. The development and increasing methodological acceptance of non-parametric and Bayesian statistical methods (Manly, 1997; Gelman, Carlin, Stern, and Rubin, 2003) in measurement has enabled greater flexibility and improved methods of estimation.

The enhanced methods of psychometrics in conjunction with the advancements in cognitive sciences has contributed to advancements in educational assessment (Mislevy, 2006). As Mislevy (2008) describes:

Assessment is structuring situations that evoke evidence about students' thinking and acting in terms of these patterns. It is an exercise of meaning-making, too: narratives about what students know and can do, in what kinds of situations; narratives cast in some conception of the nature of knowledge, its use, and its acquisition.

The cognitive theoretically grounded understanding of assessments has produced more learning-oriented models such as developmentally ordered progress maps as methods of assessment design (i.e., developmental assessments) and cognitive diagnostic methods. Developmental assessments are grounded in theories of the development of knowledge, skill, and understanding, more commonly referred to as progress maps. Interest in cognitive diagnostic information has spurred the formation of Tatsuoka's rule-space model (1983); Dibello, Stout, and Roussos's unified model (1995); Bayes Nets, which is based on the underlying theory of Bayesian statistics; conditional probabilities (Almond, 1995); and, relatedly, Mislevy's evidence-centered assessment design (Mislevy, 1994; Mislevy, Steinberg, & Almond, 1999). The evolution and merging of the cognitive and measurement sciences has enabled the refinement of formative and summative educational assessments.

Formative Assessment

The NRC Committee on the Foundations of Assessment argues that although summative assessments, historically and currently, receive the most research and policy attention, it is formative assessments that are most important for learning. Formative assessment

Successful Examples of Formative Assessment Designs:

Dynamic Pedagogy (DP), a multifaceted approach to teaching and learning that integrates the critical classroom processes of Assessment, Curriculum, and Instruction. Teachers constantly adapt to meet students' changing needs and strengths on a particular task (Armour-Thomas, Gordon and colleagues, 2005). Based on the work of Vygotsky (1978), Feuerstein (1980), and Sternberg (1988).

Intelligent Tutoring Systems, a computer-based tutoring system that incorporates well developed and highly specific descriptions of thinking about specific school subjects (Anderson and colleagues, 1990). The intelligent tutor is able to diagnose errors and provide specific instruction in order to mediate the student through the learning task.

Cognitively Guided Instruction (CGI) uses continuous classroom assessment strategies to modify instructional practices as needed (Carpenter and colleagues, 1996). Underlying CGI is a coarse-grained model of student thinking that is easily accessible to and employed by teachers.

Facets DIAGNOSER is an instructional program based on models of learning termed facets of student thinking. This instructional strategy is predicated on the cognitive principle that students come to instruction with initial ideas and preconceptions that the teacher should identify and build upon (Hunt & Minstrell, 1996).

Assessment via *epistemic games* (Shaffer, 2006) is designed to give learners the rich experience of professional practice within a discipline. They serve to develop domain-specific expertise based on principles of collaborative learning, distributed expertise, and complex problem-solving.

Stealth assessments (Shute, 2008) are unnoticed or unobtrusive to the learner and employ inferential practices that would be too hard for humans (e.g., estimating values of evidence-based competencies across a network of skills) by utilizing automated scoring and machine-based reasoning techniques. Internet-embedded assessments that are seamlessly woven into the fabric of the learning environments of virtual worlds is an example.

Successful Examples of Summative Assessment Designs:

The National Assessment of Educational Progress is the only assessment program that captures a nationally representative sample of students in order to evaluate state, regional, and national educational performance.

The Developmental Assessment Program of Australia (Masters & Forster, 1996) is designed to measure growth in competence on particular tasks by employing progress maps. Progress maps provide a description of skills, understandings, and knowledge in the sequence in which students typically learn and develop (NRC, 2001). The progress maps are also used to theoretically align both large-scale and classroom assessments so that they support one another.

The College Board® AP® Studio Art program is a curriculum-embedded design that requires students to compile a portfolio of work that may have been produced during or prior to the academic year for individual students' summative evaluations.

Facets DIAGNOSER may also be used as a curriculum-embedded assessment by student evaluations based on the most advanced module of student completion.

Formative and **summative** assessments are not independent of each other and, in fact, have a dynamic relationship. The dynamism lies in the feedback processes that each provide for teacher and student. In other words, formative assessments should be designed to optimize student performance on summative assessments and provide summative information for teacher and student. Summative assessments should be designed to provide general feedback to teachers of students' learning and development in a given area for instruction. In fact, independent of design, the distinction between formative and summative assessments is in the use of the produced information, whether for instructional feedback or decision making.

is in the developmental spirit of Vygotsky (1978) and incorporates cognitive and situative perspectives of learning (Greeno, 1998). It is formative assessment that provides feedback to teachers on the developed skills an individual student has in order to inform the pedagogical strategy and creation of necessary scaffolding. Formative assessments do not have the high stakes associated with decision-making summative assessments because they are embedded in instruction and performed unobtrusively throughout the instructional period. Formative assessments also do not have the high direct cost summative assessments carry, and have a greater effect on learning. In fact, in a study that Black and Wiliam (1998) conducted examining gains between pre-tests and post-tests, comparing programs that used formative assessment against matched control groups, effect sizes in the range of 0.4 and 0.7 were found. To put this in practical terms, an effect size of 0.7 would raise the U.S. ranking on the Third International Mathematics and Science Study from the middle of the 41 participating countries to one of the top five.

There are various types of formative assessments. Formative assessments may be conducted by way of classroom dialogue, questioning, seatwork and homework assignments, formal tests, less formal quizzes, projects, portfolios, etc. However, the NRC Committee stresses that "the development of good formative assessment requires radical changes in the ways students are encouraged to express their ideas and in the ways teachers give feedback to students so they can develop the ability to manage and guide their own learning" (p. 227). The most helpful type of feedback provides specific comments about errors and suggestions for improvement, and encourages students to thoughtfully engage the task rather than provide the right answer. The quality of this feedback can have direct implications for students' academic identity, the classroom culture, ideologies of ability, and affective development. Thus, underlying the effectiveness of formative assessment is the importance of the availability of high-quality professional development.

Summative Assessment

In addition to formative assessment, there are various summative assessments. The most widely recognized are the norm- and criterion-referenced large-scale assessments. Other types of summative assessments include grades, end-of-course exams, and curriculum-embedded assessments. Such curriculum-embedded assessments extract evidence of student performance from classroom work produced during the course of instruction in the academic year for student evaluation and promotion. This provides a more fair summative measure of individual performance that takes into account the students' personal educational experience.

The form of summative assessment utilized depends on the purpose. If it is for evaluation purposes at any level (beyond individual students) then a large-scale assessment such as the National Assessment of Educational Progress (NAEP) is more appropriate, but a measure fit for individual student performance would be a curriculum-embedded assessment. NAEP is a nationally representative sample of students designed to provide state-, regional-, and national-level evaluative information, whereas curriculum-embedded assessments are designed to provide summative information on individual students.

Development and Design of an Assessment System

When developing a formative or summative assessment, the NRC Committee emphasizes what they call "the assessment triangle." In order for an assessment to be effective, the following elements of the assessment triangle must be in synchrony: cognition, observation, and interpretation.

(1) Cognition refers to the theory or set of beliefs on how students represent knowledge and learn.

(2) Observation refers to the theory or set of beliefs about the kinds of tasks or situations that will prompt students to say, do, or create something that demonstrates important knowledge and skills.

(3) Interpretation is all the methods and tools used to reason from fallible observations and/or make sense of student's responses to various tasks.

A well-grounded and specified cognitive theory will provide the assessment designer with the sorts of tasks that will produce information on what a student knows in that domain. Likewise for interpretation, a well-specified cognitive theory also will indicate the limitations of the observations and what constitutes valid evidence.

The NRC Committee also stresses the need for a balanced assessment system. A balanced assessment system would be congruent in learning and developmental theory across assessments, curriculum, teaching, and learning, and would exhibit three properties: comprehensiveness, coherence, and continuity.

(1) Comprehensiveness refers to a range of measurement approaches that should be used to provide a variety of evidence to support educational decision making. In other words, high school graduation or grade promotion should not be based on one assessment, but rather should incorporate several assessments, such as a large-scale assessment, grade point average, a portfolio, or teacher recommendation. No one of these should be the determining factor; the point is to capture a definitive measure of a student's competence.

(2) Coherence implies that there needs to be alignment on the theory and goals of learning between large-scale assessments and classroom assessments as well as among curriculum, instruction, and assessment. In other words, the learning and developmental theory on which the curriculum was designed should also underlie the instruction and assessment. The large-scale assessment should be a broader conceptual version of the classroom assessments.

(3) Continuity suggests that, in order to provide real pictures of progress, there needs to be a continuous assessment system in which multiple sets of observations over time must be linked conceptually so that change can be observed and interpreted. Several measurements of an assessment over a period of time can provide rich information of growth in learning and development in a particular area.

Each one of these properties must be met in order to have an effectively functioning assessment system that is based on models of learning and development. In agreement, the NRC report on state science assessments (Wilson & Bertenthal, 2006) proposed an additional perspective to coherence. They posited that a successful standards-based assessment system would be *horizontally coherent* by having curriculum, instruction, and assessment theoretically in continuity; *vertically coherent* by having the classroom, school, school district, and state aligned to the same learning goals; and *developmentally coherent* by taking account of how students' understanding of specific subject areas develops over time. Thus, in a balanced assessment system, there must be coherence not just horizontally and vertically but also developmentally.

A balanced design to assessment is ever more important in the current standards-based system of

accountability. In response to this critical moment, the National Academy of Education working group on standards, assessments, and accountability (Shepard, Hannaway, & Baker, 2009) make the following recommendations:

(1) The federal government should encourage the redesign and clear connection of content and performance standards — and the curricula, teacher training, and high-quality assessments to go with them — with the goal of developing clearly articulated statements of the expected progression of learning.

(2) The federal government should support research on accountability system indicators to reflect both the status and growth of students. Performance standards should set ambitious but realistic targets for teaching and learning, and they should communicate to the public, parents, educators, and students themselves what is to be learned. Assessment results should be reported in ways that recognize progress all along the achievement continuum.

(3) The federal government should support the redesign and ongoing evaluation of accountability systems to ensure that they contribute to school improvement. Less-than-satisfactory school performance should trigger closer investigation of school operations before remedies or sanctions are applied, and stellar performances also should be verified. Different investigative approaches, including audit assessments, data-driven analyses, and expert-constituted inspectorates, should be considered.

(4) The federal government should support an intensive program of research and development to create the next generation of performance assessments explicitly linked to well-designed content standards and curricula.

These recommendations, along with the explicit goals of developing a comprehensive, coherent and continuous assessment system, will help to create a more equitable system of education where everyone potentially can achieve intellective competence.

Conclusions

It is contended in this NRC report that a comprehensive, coherent, and continuous design of assessment, curriculum, teaching, and learning will contribute to the development of a more equitable educational system — an effective educational system that will contribute to the development of competence in all students (Gordon, 1999).

As Gordon (2001) writes, intellective competence is a characteristic way of adapting, appreciating, knowing, and understanding the phenomena of human experience through the domains of cognitive, affective, and situative competence. The assumption is that most people have the potential to be intellectively competent in an educational system with equity and justice at its core.

In order for the field of psychometrics to move systems of education toward this goal, it is necessary to begin to bridge the various dimensions of human development with the contemporary theories of measurement. The various advancements in the merging of the cognitive and measurement sciences are certainly moving the field of psychometrics closer to addressing the construct of intellective competence. The model of formative and summative assessment set forth in the NRC report will help to further this goal of developing such competence in all children.

The NRC report came at an important time because of its emphasis on formative assessment despite the field's current reservations toward classroom assessment and emphasis on large-scale testing. If one were to take the report seriously, then one would recognize that current national policy is a departure from the research evidence on what we know works and what is needed for the United States' economic and intellectual development. In order for the United States to continue to move its educational enterprise into position to address the demands of the 21st century, it is imperative that the nation focus its attention on the universal development of human capital (Miller, 1995).

In an ever more diverse society, it is imperative that we implement balanced assessment systems that enable systems of education that better serve the critically important goals of equity and competence. That achievement will require that our system of education and our system of assessment use all of our emerging knowledge to inform what we do in educational assessment to inform and improve teaching and learning.

References

Almond, R. L. (1995). *Graphical belief modeling.* London: Chapman and Hall.

Anderson, J. R., Boyle, C. F., Corbett, A., and Lewis, M. W. (1990). Cognitive modeling and intelligent tutoring. *Artificial Intelligence, 42,* 7–49.

Armour-Thomas, E., Walker, E., Dixon-Román, E., Mejia, B., and Gordon, E. W. (2006). An Evaluation of Dynamic Pedagogy on 3rd Grade Mathematics Achievement of Ethnic Minority Children from Low Income Backgrounds. (Paper presented at the American Educational Research Association Annual Meeting, San Francisco, CA.)

Black, P., and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–73.

Bridglall, B. L. (2001, March). Research and practice on how people learn. *Pedagogical Inquiry and Praxis,* 1. New York: Columbia University, Teachers College, Institute for Urban and Minority Education.

Carpenter, T. P., Fennema, E., and Franke, M. L. (1996). Cognitively guided instruction: A knowledge base for reform in primary mathematics instruction. *The Elementary School Journal, 97*(1), 3–20.

Dibello, L. V., Stout, W. F., and Roussos, L. A. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In P.D. Nichols, S.F. Chipman, and R.L. Brennan (Eds.), *Cognitively diagnostic assessment.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Feuerstein, R. (1980). *Instrumental Enrichment: An intervention program for cognitive modifiability.* Glenview, IL: Scott, Foresman and Company.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). *Bayesian data analysis (2nd ed.).* Boca Raton, FL: Chapman & Hall/CRC.

Gordon, E. W. (2001, September). Affirmative development of academic ability. *Pedagogical Inquiry and Praxis,* 2. New York: Columbia University, Teachers College, Institute for Urban and Minority Education.

Gordon, E. W. (1999). *Education & Justice: A view from the back of the bus.* New York, NY: Teachers College Press.

Greeno, J. G. (1998). The situativity of knowing, learning, and research. *American Psychologist, 53*(1), 5–26.

Hunt, E. and Minstrell, J. (1996). Effective instruction in science and mathematics: Psychological principles and social constraints. *Issues in Education: Contributions from Education Psychology, 2*(2), 123–162.

Manly, B.F.J. (1997). *Randomization, bootstrap and Monte Carlo methods in biology (2nd ed.).* Boca Raton, FL: Chapman & Hall/CRC.

Masters, G. N. and Forster, M. (1996). *Developmental assessment: Assessment resource kit.* Hawthorn, Australia: Australian Council on Education Research Press.

Miller, L. S. (1995). *An American imperative: Accelerating minority educational advancement.* New Haven, CT: Yale University Press.

Mislevy, R. J. (1994). Evidence and inference in educational-assessment. *Psychometrika, 59*(4), 439–483.

Mislevy, R. J. (2006). Cognitive psychology and educational assessment. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.) (pp. 257–305). Portsmouth, NH: Greenwood Publishing Group.

Mislevy, R. J. (2008). "Some Implications of Expertise Research for Educational Assessment." Keynote address at the 34th International Association for Educational Assessment (IAEA) Conference, Cambridge University, September 8, 2008. Retrieved May 16, 2011, from http://www.iaea2008.cambridgeassessment.org.uk/ca/digitalAssets/180456_Mislevy.pdf

Mislevy, R. J., Steinber, L. S., & Almond, R. G. (1999). "Evidence-Centered Assessment Design." Retrieved May 16, 2011, from http://www.education.umd.edu/EDMS/mislevy/papers/ECD_overview.html

National Academy of Education. (2009). *Standards, Assessments, and Accountability.* National Academy of Education Working Group on Standards, Assessments, and Accountability. Shepard, L., Hannaway, J., and Baker, E. (Eds). Education Policy White Paper. Washington, DC: National Academy of Education.

National Research Council. (1999c). *How people learn: Bridging research and practice.* Committee on Learning Research and Educational Practice. M. S. Donovan, J. D. Bransford, and J. W. Pellegrino (Eds.). Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment.* Committee on the Foundations of Assessment. Pellegrino, J., Chudowsky, N., and Glaser, R. (Eds.). Board on Testing and Assessment, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Shaffer, D. W. (2006). *How computer games help children learn.* New York: Palgrave/Macmillan.

Shute, V. J., & Spector, J. M. (2008). "SCORM 2.0 White Paper: Stealth Assessment in Virtual Worlds." Retrieved May 16, 2011, from http://www.adlnet.gov/Technologies/Evaluation/Library/Additional Resources/LETSIWhite Papers/Shute–StealthAssessmentinVirtualWorlds.pdf

Sternberg, R. J. (1988). *The triarchic mind: A new theory of human intelligence.* New York: Viking.

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement, 20,* 345–354.

Vygotsky, L. S. (1978). *Mind in society.* Cambridge, MA: Harvard University Press.

Wilson, M. R., & Bertenthal, M. W. (Eds.). (2006). *Systems for state science assessment.* Washington, DC: The National Academies Press.

In line with Chairman Gordon's thinking, *ATL* is committed to pushing forward innovative and practical considerations from scholars that take seriously the advancement of human capital through the development of strong minds. Perspectives will be anchored in the desire and need to do better in the utilization of assessment, and will be supplemented in future issues with readings, resources and lists that help to frame the future of assessment in a way that is responsive to 21st-century learners. We look forward to public discourse and trust our readers also will make their perspectives known through contacting us.

Edmund W. Gordon, Publisher • David Wall Rice, Editor-in-Chief • Paola Heincke, Managing Editor

The Gordon Commission
on the Future of Assessment in Education

Contact us at:

contact@gordoncommission.org

Gordon Commission • P.O. Box 6005 • Princeton, NJ 08541