

# Assessment, Teaching, and Learning

The Gordon Commission on the Future of Assessment in Education

Volume 2, Issue No. 2 ■ April 2012

## Shifting Paradigms: Beyond the Abstract

With each issue of *Assessment, Teaching, and Learning*, there is a push toward futuring assessment. There is the coupling of extant knowledge involving education in a transdisciplinary way in an effort to position new paradigms as logical extensions of what we know to work in the development of whole persons and their intellectual competence. With this issue of *ATL*, we have provided an editorialized summary of An American Education Model<sup>1</sup>. This work provides logical alternatives in how it is that we consider, and might better position, assessment in the 21st century. The detail and nuance of this important piece is not captured in the summary provided, so we invite you to read the entire published piece here: <http://tiny.cc/zte1aw>. Still, the broad strokes of An American Education Model herein are at the core of what the Gordon Commission is intent on providing to practitioners and education scholars: tangible alternatives to present assessment systems that are able to engage and produce more responsive learners.

### Purpose and General Design

Advances in education research, statistics, technology, design, and policy have prepared the American education system for breakthroughs in standards, curriculum, assessment, and the relationships among them. This time, in ways that were not previously possible, we have the knowledge and the tools to keep ambitious teaching and learning at the center of the system, even as we sustain our commitment

to accountability and equity. We also have new tools that will enable us to use the data that emerge from the system to continuously improve it and tune it to the needs of each child.

We propose a new vision for an aligned system of standards, assessment, and curriculum.

*Assessment, Teaching, and Learning* is a bi-monthly bulletin that is the primary instrument of communication from the Chairperson of the Gordon Commission on the Future of Assessment in Education 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.

<sup>1</sup> Resnick, Lauren B., and Larry Berger. An American Examination System. National Conference on Next-Generation K–12 Assessment Systems. March 2010. Accessed at <http://www.k12center.org/rsc/pdf/ResnickBergerSystemModel.pdf> on March 8, 2012.

Our vision comprises three main areas of innovation working in concert:

1. ***Distributed Accountability Exams (DAEs)***, given periodically during the school year rather than once near the end, with items designed to be not only valid and reliable but also *educative*. These would be administered after a unit of curriculum that is expressly designed to prepare students for the exam.
2. ***A formative assessment system*** that is designed from the ground up around how teachers will make use of assessment in the classroom. Because all of the assessment data, both summative and formative, feeds a comprehensive learning profile of each student, the technology of *adaptive mass personalization* can be applied to shrink the testing burden and target assessment to each child's current place on relevant learning trajectories.
3. ***A technology platform*** that makes it easy for schools and teachers to manage the assessment process and that puts at teachers' fingertips the insights and actions that should follow from assessment data. The platform facilitates scoring many item types instantly and, when human scoring is required, streamlines the workflow to provide feedback as swiftly as possible.

## The Problem

Over the past two decades, our country has been trying to build a standards-based accountability system as a foundation for a more equitable and higher-achieving education system. In practice, however, we have created a *test-based* accountability system that does not reflect the standards we aimed for at the beginning of the 1990s, much less today's *fewer, clearer, higher* Common Core Standards.

Several studies, using several different methodologies, have shown that the state tests

do not measure the higher-order thinking, problem solving, and creativity needed for students to succeed in the 21st century. These tests, with only a few exceptions, systematically overrepresent basic skills and knowledge and omit the complex knowledge and reasoning we are seeking for college and career readiness.

The misrepresentation of standards by most current accountability tests has had negative effects on teaching and learning, especially for poor and minority students. The tests carry consequences, and many educators serving poor students aim to raise test scores in the most direct — in some cases, the only — way they know: They provide practice on exercises that substantially match the format and content of their state's end-of-year accountability tests. These exercises often depart substantially from best instructional practice. Some studies have documented a systematic *decline* from fall to spring in the quality of instruction. In reading, for example, the complexity of texts that students engage with is *lower* — in the same classrooms, with the same children — in March than in October. And there is *less* discussion of text and word meaning as teachers direct children through workbook exercises that mimic state test items (Anagnostopoulos, 2003; Koretz & Hamilton, 2006; McNeill, 2002). Principals and district administrators encourage this practice. They introduce interim assessments that largely mirror the end-of-year tests rather than model the kinds of performance intended by the standards. They do this because the tests count, and they are afraid that without practice, students will not do well enough to meet adequate yearly progress (AYP) requirements.

Calls now abound for even more frequent testing and for focusing teachers' attention early and often on which items their students are having difficulty answering on the interim assessments. But unless the process is guided by a fundamental understanding of *what kind of teaching* helps

children acquire robust competence, we should not be surprised when the most frequent response to weak early test scores is to practice the test. Though no one intended to do so, we have created a testing bind that, as it tightens, drives attention away from the intended standards. The effects are greatest in the poorest schools. The nation's current approach to raising achievement and increasing equity in the education system is having an effect opposite from the intended one. It is trapping poor children in a basic skills teaching program that gives them little chance to acquire the deeper knowledge and abilities we seek for everyone. And it may be lowering the learning opportunities even for many more privileged children as schools turn their energies to the test-based basic skills program.

## A Solution

Testing and accountability should remain at the heart of national education policy. Equity and national prosperity depend on a system that will stretch educators, the education system, and communities to work toward high achievement, and that will enable clear accountability when achievement goals are missed. But there should be new forms of assessment, functioning in new ways within the education system, to meet the needs. As early as 1992, scholars showed how, in many countries of the world, tightly linked examination and curriculum systems kept aspirations high, guided teachers in their work, and — sometimes — created pathways for young people who did not come from privileged families (Resnick & Resnick, 1992). The secret lay in charging teachers to *prepare their students for exams* and making sure that the exams were worth studying for. For the system to work, teachers and students needed to have a rough idea of the *kinds* of questions that would be posed on the exams, although not the specific questions that would appear. The systems also required trust that exam grades would be fair — that is, students would likely receive the same grade no matter who scored their written work.

(Written essays predominated over short-answer and multiple-choice items because the countries valued the kinds of thinking that were displayed in such essays.) Systems for checking on grade fairness (and allowing challenges in a few cases) varied among the countries studied, but all found ways of maintaining public trust in the system.

We outline an American Examination System, one that reflects key aspects of the substantive, cognitively demanding European systems while maintaining standards of psychometric rigor necessary to support America's accountability, comparability, and equity agendas.

The American Examination System we outline:

- ***models the kinds of instruction that are valued*** so that preparing students for assessment works for — rather than against — high cognitive demand instruction;
- ***situates exams within the stream of ongoing instruction*** so that assessments support teaching rather than distract from it;
- ***ensures content and instructional validity of all assessments*** so that the alignment problems that have plagued state testing systems can be resolved;
- ***provides reliable and valid accountability measures*** for student, school, and educator performance;
- ***includes diagnostic tools for instruction*** to meet individual student needs; and
- ***leverages advanced data collection and computational resources*** to mass personalize the formative assessments, improving their precision and usefulness.

The American Examination System we outline would be *educative* for those who use it. It would not just tell us how well students, teachers, and schools are performing, but also teach *teachers* how to teach, teach *students* how to learn, and

# On the Loss of Our Friend Bob Glaser: Reflections from James W. Pellegrino and Edmund W. Gordon

*James W. Pellegrino is a member of the Gordon Commission and is Distinguished Professor of Psychology and Education at the University of Illinois-Chicago.*

There are many words that could be written about the career of Bob Glaser, his many significant accomplishments in the fields of psychology and education, his many awards, and the many lives he touched across the globe. At a recent memorial service held on the University of Pittsburgh campus, many of his colleagues, former students, and friends — including members of the Learning Research and Development Center — came together to offer some words of remembrance, all of us acknowledging the deep gratitude we had for having experienced both his mentorship and his friendship. Everyone had a favorite story to tell about Bob's uncanny ability to help improve their papers by detailed editing comments and the constant question in the margin: "and so?" He gave of his time and intellect without question because he so loved the processes of learning, understanding, and communicating.

As all who knew him remarked, he always asked, "What have you discovered today?" ... and it was not just a throwaway remark of casual conversation. He was serious and expected an answer. He challenged each of us to think about our science and the problems we chose to study because he genuinely believed that the scientific study of learning could make a major impact on the education of all students. More than once, as a young assistant professor, after telling him about the elegant design for the next experiment I planned to do, he would then say to me: "Now tell me why that's the best way for you to spend your time." That was probably the most intimidating

question anyone has ever asked me, but it was also the most useful in shaping my career.

I had the great pleasure of collaborating with Bob early in my career on research on the analysis of aptitude, and we wrote a number of papers together — always a wonderful learning experience for a junior scholar. He introduced me to the world of educational testing and to the idea that we could use our emerging knowledge from cognitive psychology to construct better, more valid, and more useful educational tests. That fervor never changed for Bob, and he continued to pursue those ideas in productive collaborations with many scholars at LRDC. It was my distinct pleasure to collaborate with Bob again after a two-decade hiatus when we co-chaired the National Research Council Committee on the Foundations of Assessment. The *Knowing What Students Know* report from that committee reflects many of the concerns and ideas that Bob articulated years earlier in seminal papers on assessment, learning, and instruction. He was a true visionary whose ideas predated by many years the field's capacity for understanding and implementation. But he was never frustrated or impatient. Rather, he was committed to the cumulative nature of knowledge building in the cognitive sciences and the continual exploration of ways to bring scientific knowledge to bear on problems of practice. He saw education as a "design challenge," and he helped frame the idea of a field known as Instructional Psychology. He construed it as a design science, bridging the worlds of psychological theory and classroom practice. In seeking a vision for the future of assessment, the work of the Gordon Commission is very much in

keeping with the vision and commitments of Bob Glaser. While we will all miss him, his impact on many of us in how we frame the problems we seek to solve and the areas of scholarship from which we draw to seek solutions will continue forward unabated.

### Professor Edmund W. Gordon

I came into my professional maturity under the intellectual influence of the late Benjamin Bloom and the late Robert Glaser. Ben was my elder. Bob was my peer. But I came into intellectual personhood admiring their wonderfully analytic minds, their abilities for the synthesis of information, and their remarkable capacities for the creation of synergy in even disparate ideas. I also observed these developed abilities in my friend and mentor W.E.B. Du Bois. But in both Ben and Bob I saw these uniquely human capacities decline and disappear as the cognitive dysfunction we call Alzheimer's disease destroyed memory, precluded the capacity for recognition and adjudication of relationships, and impaired the capacity for judgment. While we curse the disease that is so destructive of human functions that we cherish, and while we mourn the passing of another intellectual giant, we may learn from the functional characteristics of these persons something of the phenomenon that they spent their lives studying — human intellect. To be an intellectually competent person is to have honed the abilities for analysis, synthesis, and synergesis in a mind that is compassionate, caring, and humane. I saw living examples of these characteristics in my friend Robert Glaser. I learned so much from him. May he rest in peace.

teach education *organizations* how to develop teaching expertise. It would meet this educative goal through a system that combines *distributed accountability exams* linked to specific topics for instruction with *diagnostic, formative assessments* designed for teacher use *during* instruction.

An online platform will make it possible to deploy and manage all of these elements at scale in a cost-effective way while minimizing additional burdens for teachers, students, and administrators. This online platform would be much more than a system for administering, scoring, and reporting on assessments. It can surround the *what* of assessment outcomes with useful representations of *so what?* (professional development) and *now what?* (more targeted instructional resources), so that everyone focuses on the consequential and instructional validity of assessment results and not just the accountability pressure.

## References

- Anagnostopoulos, D. (2003). Testing and student engagement with literature in urban classrooms: A multi-layered perspective. *Research in the teaching of English*, 38(2), 177–212.
- David, J. L., & Greene, D. (2008). *Improving mathematics instruction in Los Angeles High Schools: Follow up to the evaluation of the PRISMA Pilot Program*. Palo Alto, CA: Bay Area Research Group.
- Engel, S. (2010, February 1). Learning to play. *The New York Times*. Retrieved from <http://www.nytimes.com/2010/02/02/opinion/02engel.html>
- Koretz, D., & Hamilton, L. (2006). Testing for accountability in K–12. In R. L. Brennan (Ed.), *Educational measurement*, 4th ed., pp. 531–578. Westport, CT: American Council on Education/Praeger.
- Mariano, L. T., & Junker, B. W. (2007). Covariates of the rating process in hierarchical models for multiple ratings of test items. *Journal of Educational and Behavioral Statistics*, 32, 287–314.
- McConachie, S. M., & Petrosky, A. R. (Eds.). (2009). *Content matters: A disciplinary literacy approach to improving student learning*. San Francisco, CA: Jossey Bass.
- McNeill, L. (2002). *Contradictions of school reform: Educational costs of educational testing*. New York, NY: Routledge.
- Patz, R. J., Junker, B. W., Johnson, M. S., & Mariano, L. T. (2002). The hierarchical rater model for rated test items and its application to large-scale educational assessment data. *Journal of Educational and Behavioral Statistics*, 27, 341–384.

Phillips, V., & Wong, C. (2010, February). Tying together the common core of standards, instruction, and assessments. *Phi Delta Kappan*, 91(5), 37–42.

Rayn, K. E., & Shepard L. A. (Eds.) (2008). *The future of test based educational accountability*. Mahwah, NJ: Lawrence Erlbaum Associates.

Resnick, L. B. (in press). Nested learning systems for the thinking curriculum. *Educational Researcher*.

Resnick, D. P., & Resnick, L. B. (1977). The nature of literacy: An historical exploration. *Harvard Educational Review*, 47(3), 370–385.

Resnick, D. P., & Resnick, L. B. (1980). The nature of literacy: An historical exploration. In M. Wolf, M. K. McQuillan, & E. Radwin (Eds.), *Thought and language/language and reading* (pp. 396–411). Cambridge, MA: Harvard University Press.

Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement, and instruction* (pp. 37–75). Boston: Kluwer.

Resnick, L. B., Stein, M. K., & Coon, S. (2008). Standards based reform: A powerful idea unmoored. In R. D. Kahlenberg (Ed.), *Improving on No Child Left Behind: Getting education reform back on track*. New York, NY: The Century Foundation Press.

Rowan, B., Correnti, R., Miller, R., & Camburn, E. (2009). School improvement by design: Lessons from a study of comprehensive school reform programs. In B. Schneider, D. Sykes, & D. N. Plank (Eds.), *AERA handbook on education policy research*. New York, NY: Routledge.

Talbert, J. E., & David, J. L., with Lin, W. (2008, September). *Evaluation of the disciplinary literacy-professional learning community (DL-PLC) initiative in Austin Independent School District. Final report*. Palo Alto, CA: Stanford University, Center for Research on the Context of Teaching.

U.S. Department of Education. (2010). *State and local implementation of the No Child Left Behind (NCLB) Act: Volume IX—Accountability under NCLB. Final report*. (2010). Washington, DC: Author.

## Outlines of a Commission Paper

In this new section of *ATL*, we will provide a glimpse into the work of the Gordon Commission with themes that are being developed within a collection of more than two dozen papers in progress. This first installment is from *Changing*

*Paradigms for Education* by Edmund W. Gordon, E. Wyatt Gordon, Larry Aber and David Berliner.

## Changing Paradigms

One of the three stated missions of the Gordon Commission on the Future of Assessment in Education is to consider our best estimates of what education will become in the 21st century and what will be required of the educational assessment enterprise by the middle of this century. In the pursuit of addressing that component of our mission, Commissioners and Consultants to the Commission are considering a variety of anticipated and emerging changes in the paradigms by which the goals and processes of education are changing. Among these paradigms are such ideas as the examples that follow.

**A.** Led by such writers as William Butler Yeats, we see a shift from thinking about education as concerned with “filling buckets to lighting fires.” Increasingly, the goals of education reflect the growing concern with encouraging and enabling students to learn how to learn and to learn to continue learning; to become enquiring persons who not only use knowledge but who produce and interpret knowledge. The pedagogical challenge will be less concerned with imparting factual knowledge and more concerned with turning learners on to learning and the uses of their mental abilities.

**B.** “Reading, wRiting, and aRithmetic” will continue to be essential skills, but thought leaders in education — Sir Kenneth Robinson among them — increasingly point to varying combinations of “three Cs” as essential processes in education: Creativity, Conceptualization, Collaboration, Communication, and Computation. The “three Cs” are replacing the “three Rs” as the modern ends toward which education is directed. Learning how to think, reason, interpret, access, and create knowledge will be more and more privileged in the 21st century. The new century places high value

on communication as reading and speaking, but also as listening, collaborating, and processing information from multiple perspectives. The capacity to recognize and even create relationships between novel and disparate inputs of information will be rewarded in this new century. The illiterate members of 21st-century societies will be those who cannot navigate the world of digital technology, and computer literacy will require far more than word processing, social networking, and playing electronic games. Digitalization will change modern societies even more rapidly and radically than did industrialization and, in the process, education and its assessment will change.

**C.** The 18th, 19th, and 20th centuries privileged de-contextualization in the pursuit of precision in measurement and control in experimentation. When we turned to multivariate analysis, it was with a view to the sequential teasing-out of the contribution made by each component variable, even while we were beginning to understand the notion of dialectical interaction. The isolation of variables or components for the purpose of study may continue while the intent of such study is to know; however, as our purpose turns to understanding of the phenomena of the world and the relationships among these phenomena, experimenting, observing, and measuring things out of the contexts in which they have developed and function will become more and more dysfunctional. Education and its assessment will have to become capable of capturing aspects of context, perspective, and the attributions that come to be assigned to these conditional phenomena. The exactness and precision that have been gained by de-contextualization in the past will be challenged by the situativity required when contextualism and perspectivism are required for understanding.

**D.** In the interest of scientific validity, traditionally we have privileged “objective” knowledge over “subjective” information. We have been taught to try to control for or contain variance that is

associated with affect and social/psychological situation. We have tended to examine cognitive functions independent of their contamination or being influenced by human biases and feelings. Yet modern social and psychological sciences are pressing us to examine or assess human performance with greater respect for affective, emotional, situative, and social processes. Evidence mounts in support of the fact that these processes influence the character and the quality of human performance, yet “objective” documented human performance is the source of the data of assessment in education. Assessment in the future will have to be more sensitive to subjective phenomena — to affect, attribution, emotion, identity, situation, etc. — as will the teaching and learning transactions in which learners are engaged.

**E.** Assessment of the outcomes of learning in the interest of accountability will be with us for a while, but the future is likely to bring increased concern for assessment for the purpose of informing and improving learning and the teaching processes that enable learning. Political pressure continues to support a preoccupation with the possibly inappropriate use of educational assessment data for accountability purposes, even though such practices are not supported by the empirical evidence, and some of us feel that such practices are actually counterproductive for the intended purposes. We have known for more than a century that what we do in education is imprecise, that one model does not fit all, and that much of our intervention is underanalyzed trial and error. We believe that assessment in education can and should inform and improve teaching and learning processes and outcomes, without ignoring the importance of accountability. Whether the two purposes can be served concurrently and by the same assessment instruments and systems is one of the questions to be answered.

## Remembering our Friend Michael E. Martinez

It is with a deep sense of personal and professional loss that I inform you of the death of Professor Michael E. Martinez on Thursday, April 5, 2012. Michael suffered a seizure two days after returning from our meeting in San Juan. It was determined that a malignant growth for which he was under treatment for the past eight years had reached his brain and spine. Very aggressive treatment was unable to stabilize the process, but he was able to spend several days with his family before the end came for him. Michael was a full professor at the University of California at Irvine. He was a highly respected member of the Gordon Commission as well as a cherished personal friend. His book *Education as the Cultivation of Intelligence* occupies a special place in my library. His friendship and now his spirit occupy a special place in my heart.

Edmund W. Gordon

Michael Edward Martinez was Professor of Education at the University of California, Irvine (UCI). He taught courses on learning and cognition, and on intelligence, at the undergraduate and graduate levels.

A former high school science teacher, Dr. Martinez received his Ph.D. in educational psychology from Stanford University in 1987. He subsequently joined the

research staff of Educational Testing Service (ETS) in Princeton, New Jersey, where he developed new forms of computer-based testing for assessment in science, architecture, and engineering. This work resulted in two U.S. patents.

At UCI, Dr. Martinez conducted research on science and mathematics learning, as well as on the nature and modifiability of intelligence. He has published in such journals as *Educational Psychologist*, *The Journal of Educational Measurement*, and *The Journal of the American Society for Information Science*. His first book, *Education as the Cultivation of Intelligence*, was published in 2000. A second book, *Learning and Cognition: The Design of the Mind*, was published in 2010. A third book, *The Future Bright: A Transformative View of Human Intelligence*, is under contract with Oxford University Press.

While a faculty member at UCI, Dr. Martinez was awarded a Fulbright Scholarship at the University of the South Pacific in the Fiji Islands (1994–1995). He served as Program Director for the National Science Foundation (2001–2002), where he managed the NSF's role in the Interagency Educational Research Initiative (IERI). His honors include appointments as College Board Visiting Scholar (2002–2003), Mellon Visiting Scholar at Columbia University (2003–2004), Cambridge University Visiting Scholar (2009), and the APA Presidential Commendation for Contributions to Psychology (2003).

In line with Chairman Gordon's thinking, *ATL* is committed to pushing forward innovative and practical considerations from scholars who 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](mailto:contact@gordoncommission.org)**

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

The Gordon Commission was established by ETS to investigate and advise on  
the nature and use of educational testing in the 21st century. 19458