Assessment, Teaching, and Learning

The Gordon Commission on the Future of Assessment in Education

Volume 2, Issue No. 5 ▪ December 2012

Paradigm Shifting with Epistemology

An important dimension of the work of the Gordon Commission on the Future of Assessment in Education has to do with the experience of paradigm changes in education and shifts in the epistemologies by which educational policies and practices are informed. A great deal of attention has been given to the exploration of changes in the ways that we teach and learn, changes in the processes of teaching and learning, and to shifts in the ways in which we think about the nature of the knowledge and techniques that we teach. Of perhaps even greater consequence is the concern with shifts in conceptions of the nature of what it means to know — that is, we, along with most who take education seriously, note the need to change how it is that we are educated and how it is that we are educating within the United States.

The troika of assessment, teaching and learning gives reasonable entry to change with constructs that are relatively accessible, albeit not necessarily as straight ahead as one might have thought at first blanch. Epistemologies, of course, are frequently discussed in inquiries concerning learning about learning, and in figuring how best to situate the pedagogical troika. But too often underlying notions concerning epistemology, the study of the origins and meanings of knowledge, are isolated as philosophical jargon some distance removed from the context in which they should be applied. With this awareness, the Commission has engaged the task of looking to epistemology in meaningful, practical ways to help define challenges of futuristic projections concerning education and its assessment. This issue of *Assessment, Teaching, and Learning (ATL)* offers commentary from the Commission Chair on the intersection of epistemology, education and assessment. That commentary is followed by abstracts of three papers that were commissioned or selected for the Gordon Commission to inform its consideration of related issues.

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.

A Word from the Chair…

In varying degrees, the members of the Gordon Commission have become involved in discussions and in the generation of papers concerning the challenges posed for education and assessment in education by the fact that our conceptions of knowledge and what it means to know continue to change. In our considerations of the future of assessment in education, we anticipate that shifts in the epistemologies that inform human thought will continue to occur, and that we are likely to experience these shifts as occurring more rapidly and in greater conflicting interaction than has been true in all of human history. These shifting epistemological perspectives are especially important in the work of the Gordon Commission because of the phenomena of focus for the Commission: education and its assessment. Both are so firmly grounded in conceptions of human behavior, traditions in observation and measurement, and even the nature of reality that were the consensus positions of the 19th and 20th centuries. Many of these notions are being challenged or at least re-conceptualized in contemporary thought and educational practice. Educational and psychological measurement theory has not stood still in the face of these changes. However, testing practice has remained relatively stable. Scholars of the measurement sciences are deeply involved in exploration of and reflection upon emerging challenges and contradictions. Some have ventured the consideration of alternatives. However, charged with the tasks of inquiry into the possible character of education in mid-21st century and the demands likely to be made on assessment in education, the members of the Gordon Commission have been forced to examine the relationships between the ways in which we think about human behavior and its assessment and the ways in which we practice assessment. In this context, anticipated and observed changes in the ways in which we think about relevant issues have forced serious examination of the theoretical assumptions that underlie what we do in assessment as well as those assumptions and ideas that will form the conceptual context for what we can expect of assessment in future years.

I wish it were possible to report light at the end of the tunnel in which we are digging. There is, perhaps, no aspect of the work of the Gordon Commission about which there is less ambiguity. Most of us who are engaged with the Gordon Commission were born into and cut our teeth on modernist empiricist scientific thought. Our minds are programmed to privilege positivist thought. But our examination of human history and the emerging epistemologies are convincing of the tentativeness of what we know and the limitations of the conceptual products of modernity. Despite the enormous technological progress that modern ways of thinking have enabled, and despite the power of the logical reasoning that has been framed by these established ways of science, we are being forced to consider that models reflect particular ways of knowing and that these ways of knowing are socially determined, essentially subjective and subject to error.

The largest single group of scholars on the Gordon Commission identify themselves with educational assessment and measurement. Our field of specialization rests on the empirical sciences and positivist thought, borrowed from the "natural sciences" and applied to the behavioral and social sciences. **However, our examination of the scholarship leads us to conclude that what we believe and know may not be a sufficiently accurate reflection of reality to be used as the sole basis for thinking about the future of assessment in education.** Some of us believe that we may not be able to get where we need to go with education and its assessment using the

extant knowledge base and the conceptual frames upon which that knowledge and related techniques rests. The Gordon Commission has included in its work three syntheses of relevant knowledge and thought concerning the shifting epistemologies that inform our work. These papers are abstracted in this issue of *Assessment, Teaching, and Learning*. The positions advanced do not reflect a consensus position endorsed by the members of the Commission. The papers do reflect a perspective to which the Chairperson of the Gordon Commission is sympathetic. The fully developed papers can be found on the Commission's website, **www.GordonCommission.org**. Readers are encouraged to seek out these provocative discussions of limitations and potentials for assessment in education that are reflected in the epistemologies that inform us.

Outlines of a Commission Paper

This section of *ATL* provides a glimpse into Gordon Commission work in real time, with themes that are being developed across a collection of more than two dozen Gordon Commission papers in progress. The present Outline provides overviews from three epistemology papers. One is part of a previously published piece that helps in providing a complete view of Commission thinking on epistemology, and two others make significant strides in advancing how epistemology and assessment presently intersect. Together, the three papers portend what we would do well to consider in the future of assessment and epistemology.

Postmodern Test Theory[1]

Robert J. Mislevy, Educational Testing Service

Introduction

This paper attempts to reconcile work applying statistics and psychology to improve educational assessment with work being labeled as "neopragmatic postmodernist test theory." That conception has the potential to create more educationally useful assessments, with the downside that those assessments won't be easily implemented from above. Throughout the paper there is comparison of the effect of postmodernist theory on the practice of law to how it might affect assessment.

Modernism and Postmodernism

Modernism proposes that there are real, correct or right interpretations and answers to the questions posed in a field. For example, modernism in physics is exemplified by the pre-20th-century thinking that objects interact according to fixed laws within universal notions of time and space. Modernism in law proposes that "justice" is a real system that can be reflected in law. In education, modernism informs the thinking that assessment can characterize people through variables that represent objective underlying characteristics. **Modernist approaches define a context in which to observe students, specify how to observe their behavior, and delineate how to draw inferences about students using those observations.**

Postmodernism is based in the idea of discourse, where the use of language constructs rather than reflects the meaning of

[1] 1997. Postmodern Test Theory. In A. Lesgold, M. J. Feuer, & A. M. Black (Eds.), *Transitions in Work and Learning: Implications for Assessment* (pp. 180–199). Washington, DC: National Academy Press.

events. In physics, the discovery of relativity and quantum theory — the realizations that there is no universal scale of time to reference and that objects don't exist in single points in space — undercut the earlier modern perspective. In the law, economics and critical feminist and race theory accomplished the same goal. In assessment, cognitive psychology made the observation that students' perspectives and social settings have profound impacts on learning and measurement.

***Neopragmatic* postmodernism accepts the postmodernism critique of modern theory, but attempts to build theories and tools that have empirical and practical value despite their fundamental shortcomings.** It is less concerned with pointing out the flaws of modernist thought and more concerned with devising solutions for problems. Neopragmatic postmodern test theory attempts to use the modernist tools of measurement to "to support learning in ways conceived in a postmodern perspective."

Progenitors of Change

The postmodern critique of assessment became necessary not least because improved testing methods made more apparent the difficulty of assessing underlying traits in an objective test. Even as item response theory and other advances made it possible to adapt tests to individual examinees and describe classroom and school effects on test scores, they revealed an inability to describe changes in test scores over time or to adequately explain why different students can find the same problem easy or hard. It became clear that the modernist methodology was insufficient to meaningfully describe learning processes.

More recent research into mental models, schema, the nature of expertise and situated learning has improved our understanding of how learning occurs. While there is disagreement about models within cognitive psychology, neopragmatic postmodernist test theory draws on all lines of research as potentially useful tools for solving practical problems. Alongside traditional tests, good teachers also use evidence from projects, classwork, peer interaction and other "informal" sources to make judgments. The challenge facing the neopragmatist is to do the same within assessment.

Some Implications of a Postmodern Perspective

Traditional assessments have been shaped by constraints on how data could be gathered, transmitted and analyzed. Economic and logistical pressures limited the use of essays and interviews that require human interpretation or the analysis of ongoing performances. Ubiquitous electronic communication, mass data storage and the proliferation of cheap computing power is opening new possibilities, some of which are outlined:

- *New kinds of tasks and scoring:* Computers can present tasks that are interactive, dynamic, constructive and more open-ended, while automating some scoring.
- *Distributed testing and scoring:* Student responses can now be recorded and transmitted electronically, allowing assessment of a rich task record in remote places and at different times.
- *"Replayability":* Capturing performances instead of only responses frees us from the limitation of evaluations by only those raters present at the original performance. This enables broader scoring and the use of exemplars to establish shared standards of evaluation.

However, **despite advances in technology and statistical models, the more examinees differ on contextual and experiential factors, the more likely it is that complex tasks will fail to provide evidence on how students would fare on other**

tasks. The same performance that proves valuable evidence for context-grounded instruction can be worthless in a broadly cast one-size-fits-all survey of students.

What is clear is that the limitations of traditional formal assessments prevent them from telling us all we would like to know about student learning. The challenge facing the neopragmatist is retaining the contextual relevance of informal assessment while simultaneously providing the communicative and credibility-based functions of traditional formal assessment. The example of the AP® Studio Art portfolio assessment suggests that it is possible using constructive and individuated data that is built over time. A common framework for interpretation, enabled by technology, can provide greater cross-contextual credibility to the judgments a good teacher has always had to make about students.

An instructive example is found in adult literacy assessments. Adult education programs reflect a multidimensional definition of literacy, and as such vary considerably in the nature and level of skills of the students with whom they work and that they emphasize. Skepticism about our ability to describe the relative effectiveness of widely disparate programs is justified. Statistical procedures can neither create nor determine the relationships among test results. They can be used to examine the relationships that do exist and harness that information for various purposes.

No single score can give a full picture of the range of skills important to different students in different adult literacy programs, and no statistical method can compare any two arbitrarily selected adult literacy tests. What is possible are three less-ambitious goals requiring the understanding that no one test can capture a full range of evidence about student proficiencies:

- *Comparing across literacy programs levels of performance in terms of consensually based tasks in standard conditions:* Some aspects of competence will be considered useful by a wide range of programs, and assessment can gather information about them in the same way for all.
- *Estimating levels of performance within clusters of literacy programs with similar objectives with shared assessments focused on those objectives:* This would necessarily be in different ways in different clusters, and at levels of accuracy demanded by the particular purposes defining each cluster. Components of programs' assessments might gather evidence for different purposes, types of students or levels of proficiency.
- *Making projections about how students from one program might perform on the assessment of another program:* Such studies are restricted by time, place, program and population. The more assessments differ in form, content and context, the more uncertainty is associated with the projections.

Conclusion

The challenge for the next century will involve new ways of understanding how the practice of assessment can strengthen the validity of its use as part of an educational system in light of reflection on and reevaluation of the legacy of the inadequacies of the modernist measurement regime. There is an understandable impulse to reject any use of assessment models and information-gathering tools because they cannot objectively measure the underlying constructs they were designed to reveal. However, doing so would unnecessarily discard decades of knowledge about some ways to structure and communicate observations about students' learning. Models and theoretical frameworks are useful tools that can enrich our understanding of student learning and program effectiveness by incorporating evidence from multiple perspectives and diverse sources.

Epistemology and Measurement: Paradigms and Practices

I. A Critical Perspective on the Sciences of Measurement

Ezekiel J. Dixon-Román, University of Pennsylvania
Kenneth J. Gergen, Swarthmore College

The political and economic interests of Western industrialist nations have centered on scientific measurement as the dominant orientation to effective policy decision making. The federal turn to evidence-based policy and decision making represents a recent instantiation. This orientation has especially materialized in education policy, more generally, and its reliance on testing, in particular. The past 10 years of federal education policy have relied on standardized measures of student achievement as means to evaluate school performance. Policymakers have assumed that the instruments of measurement are sound, robust and valid. At the same time, these assumptions rest on an early philosophy of science of measurement, a philosophy that has until recently been given little critical attention (Boorsbom, 2008; Michell, 1999; Mislevy, 1997). In the two yoked contributions that follow, we shall offer a critical analysis of the epistemological grounds of traditional measurement and provide an alternative epistemology that, in our view, holds far more promise for educational practices and policies of assessment. It is in the context of the emerging transformations in world conditions that we view this shift from a positivist/empiricist to a social constructionist epistemology as paramount in significance.

The first of these two papers lays out the main tenets and assumptions of positivist epistemology, and discusses how the positivist paradigm became a philosophical movement ultimately employed in the legitimation of the social sciences. We will pay particular attention to the influence of positivism on measurement and how measurement became a hallmark modernist science. We then shift our focus to the history and philosophy of the science of measurement, tracing each of its epistemological shifts and assumptions. Given that the major shifts in psychological measurement occurred with the development of measurement theory, this section will be organized around each of the major theoretical paradigms of measurement — i.e., classical test theory, latent variable modeling and representational models. We also discuss the more recent shifts and emergent approaches to measurement that take more of a socio-cultural and situative perspective.

In the final section, we raise and explore critical questions of the positivist paradigm of measurement that should be addressed as we move into the future. These questions include: Why should measurement sciences sustain the assumptions of universality at the cost of the multiplicity in social and cultural particularities? In what ways can the seeming objectivity represented in instruments of measurement be removed, so as to reveal the ideological investments that inhabit the production and reading of the instruments? And, is it still theoretically tenable to assume that the manifest behavioral responses to items provide any sort of privileged access to the inner workings of the human mind? Through this line of critical questioning, we set the stage for the considerations of alternative

epistemological possibilities in the second paper, and their implications for practice in the broader domain of assessment to which measurement is a limited special case.

As a précis to the analyses that follow, it will first be useful to take brief account of what many see as seismic changes taking place in world conditions. Such changes amplify the weaknesses in the traditional paradigm of measurement and invite exploration into the implications of a paradigm more adequately attuned to the emerging social conditions.

Global Change: The Fluid and the Frozen

There is now voluminous commentary on what are commonly sensed as major changes taking place in cultural — and indeed, global — life. Although a vast simplification, for present purposes it is useful to see these changes as comprising two interdependent movements: one toward disorganization and upheaval, and the other toward systematization and control — or, metaphorically, movements in the opposing directions of the *fluid* and the *frozen*. On the side of massive and disruptive change, for example, Marshall Berman (1988) focuses on the mounting ambiguities and complexities thrust upon the culture in the past century. Hardison (1995) describes the general drift away from the vision of nature as solid and tangible, and the accompanying acceleration in social change. Ritzer (2009) ascribes major transformations in patterns of contemporary life to newly emerging channels of consumption. Eitzen and Zinn (2011) stress the effects of globalization on contemporary life. Rodgers (2011) describes the contemporary condition as an "age of fracture," focusing on the loss of collective purpose; Bauman (2011) views our condition as one of increasing "liquidity," with early cultural traditions replaced by the continuing demands for the new.

Yet none of these works touches sufficiently on mammoth changes in communication resulting from a raft of newly emerging technologies. In the time it takes to read this sentence out loud, for example, over 80 million email messages will have been launched into the world. In the last year alone it is estimated that eight trillion text messages were sent via cell phones. And this is to say nothing of the increasing reach of television, or the 1.5 billion people in the world who surf the web, or the estimated more than 177 million tweets per day. This enormous expansion in communication not only facilitates rapid change, but it also multiplies the possibilities for organizing and stabilizing behavior.

As people communicate, they create meaning; they generate together a sense of what is taking place, what is valuable to do, and appropriate logics of action. The existing technologies of communication enable people to locate like-minded others from across their societies and, indeed, from around the world. Thus we find, for example, an enormous expansion in the number of religious sects, NGOs, online interest groups and politically active enclaves, and the potential for their participants to remain in communication 24/7. In effect, there is a splintering of societies into multiple groupings, or as Maffesoli (1996) has put it, we live in a "time of tribes." At the same time, at every institutional level — in business, government and religion, for example — this same potential for increased organization takes place. Rules, plans, requirements, surveillance and the like are all used to stabilize order … and the many and increasing forces for disruption bring about increased demands for control.

By and large, traditional practices of measurement in education have been employed in the service of freezing the social order. Typically, they are used to establish and sustain standards, and to facilitate policies that ensure effective performance in these terms. As one might say, national testing in education is used to order the society in terms of the

values and rationalities of those occupying positions of institutional power. To be sure, there is much to be said for realizing a vision of society in which all members enjoy the benefits of education. However, in light of the emerging conditions of change, this vision becomes increasingly perilous. In particular, it does not take into account:

- The ability of multiple groups — religious, ethnic, political, and so on — to organize themselves around alternative visions of education, to develop agendas, and to pose a critical challenge to what appear as educational impositions from elsewhere.
- The continuously shifting character of what counts as useful knowledge. With rapid transformations in world conditions, new topics and requirements are constantly developing, rendering older curricula irrelevant. Demands for courses related to environmental sustainability, digital capability, the media and the Middle East are illustrative. Differing groups in different parts of the nation also will view these educational needs in different ways.
- The continuously shifting character of pedagogy. Traditional top-down, preformed pedagogical practices are becoming outmoded by virtue of the communication skills, proclivities and needs of contemporary students. Teachers are less capable of creating and sustaining a "dominant reality" or set of values within their classes, as communication technologies enable "counter-realities" to rapidly form within student populations. In effect, what is taught and how it is taught become increasingly dependent on innovation and improvisation.

It is within this context that we turn to the justifying logics of contemporary testing.

Below are central issues that the Gordon Commission continues to address as problematic in considering the future of assessment in education.

New and Multidimensional Characterizations of Ability

- *Transfer* is the ability to recognize and use skills and knowledge gained in one context in another context or for a purpose other than that for which it was learned.
- Some of the models of knowledge that we use in traditional assessment may not bear out when data made available by new technologies are analyzed.
- Context sensitivity is important, but we have to get smarter about understanding relationships between situations and performances.
- We need better indicators of the status of intellective competence and deeper understanding of the processes by which those indicators develop and are expressed.
- There is a tension between the concept of intelligence as a function of an individual's effort and as a function of social group interaction.

- We must understand diversity of accomplishment as variation in the context of what accomplishment means in a variety of contexts.

Balancing the Different Purposes of Assessment

- We need to broaden our understanding of the attributions that are assigned to assessment data.
- More attention must be paid to the different kinds of evidence demanded by different education situations and stakeholders. It's not a simple question of whether to standardize assessment, but a question about its role in a system of judgments being made.
- Teaching to a consequential summative test can be positive if the structure of the test focuses on the right kinds of learning and performances.

Ubiquitous Assessment

- Incorporating information as it comes in to change a model of competence in real time is a promising model for bridging the divide between formative and summative assessment.
- Whatever the problems in accumulating and working with data, we can certainly still improve on the inferences made from a single decontextualized testing event.
- How is information collected for teaching used, by whom, and for what purposes?

Control and Ownership of Assessment Instruments and Processes

- Proprietary control of assessment will weaken as we introduce more diverse perspectives and practices.
- The mechanisms in place to prevent misuse of assessments are imperfect. We have to anticipate the threat of inappropriate use and establish mechanisms and procedures to control and prevent misuse.

Changing How People Think About Assessment

- We want people to think of assessment as a system that collects valid evidence in order to guide intervention decisions.
- Rather than saying we have new data and don't need traditional measurement models, say that we need new frameworks for new data, some of which may resemble the previous models.
- Note that there is a difference between the idea of assessing a student's capability as a status or as a process.

Epistemology and Measurement: Paradigms and Practices

II. Social Epistemology and the Pragmatics of Assessment

Kenneth J. Gergen, Swarthmore College
Ezekiel J. Dixon-Román, University of Pennsylvania

As developed in Part I of our offering, there are significant conceptual flaws inherent in the empiricist orientation to measurement. In light of the rapidly emerging challenges to educational practices and policies, these flaws are especially significant. With increasing challenges to the adequacy and relevance of "one-size-fits-all measurement," there are no foundational logics to support existing measurement practices. The aim of Part II is first to sketch the contours of the most promising alternative to the empiricist paradigm — namely, a socially based epistemology. With this orientation in hand, we are positioned to consider assessment as a socio-cultural practice, and to inquire further into its potentials and shortcomings. After scanning several alternatives to traditional practices, we offer recommendations for the future of assessment.

On the Future of Assessment

The nations of the world offer a wide panoply of approaches to educational testing. One might contrast, for example, the stringent and high-stakes practices of testing in Japan and Korea with the low levels of standardized testing in Finland. By the same token, one may ask about the ways in which these differing practices function in society. Interestingly, while the strong testing approaches taken in Japan and Korea are associated with high rates of adolescent suicide (along with laws to prevent excessive study), Finnish minimalism is co-terminal with what is regarded as the most effective educational system in the world. Further exploration of such differences would be useful in expanding deliberation on alternatives to present practices.

Yet, as we proposed at the outset of our initial paper, technology-based shifts in global conditions appear to militate against current testing practices in the United States. The attempt to impose a single set of standards on an enormously varied domain of practices fails to recognize the rapidly shifting and endlessly complex landscape of educational desiderata. What is needed in the way of education — the knowledge and the skills required for productive and meaningful participation — is ever changing. To lock in a highly circumscribed set of ideals and to shape the entire educational system by these specifications is to reduce the potential for productive participation in the future. Further, these same technologies of communication also foster multiple affinity groups, enabling them to generate rationales, agendas and plans of action. In effect, there is an expansion in the array of voices that demand to be heard, voices that are set against otherwise totalizing agendas. In our view, the way must be prepared for more democratic, inclusive

and dialogic contributions to the ways in which local systems of education function.

Let us briefly summarize what we see as the major outcomes of our deliberations. In our view, we should move toward modes of evaluation that will support and enable the participation of all of the nation's people in building effective educational systems and the flourishing of human capabilities more generally. Assessment should in no way drive these efforts, but should serve ancillary or supportive functions. Our views support a human development approach to achieving national goals, over and against the reliance on the neoliberal gaze and the economic marketing of education. In what follows, we first attend to what we believe assessment should not be and then shift the focus to what, for us, assessment should become. In planning for the future, we should move away from:

- employing assessment as an instrument for administrative sorting, selecting or predicting.
- assessing internal mental structures such as cognitive or affective knowledge, reasoning or other mental abilities.
- using assessment results to establish policies that differentially reward and/or penalize students, teachers or school administrators.
- forming policies demanding that all educational systems meet the same standards in terms of student performance.

To be clear, we are *not* arguing against active and continuing deliberations on the performance adequacy of students, teachers, administrators or school systems in general. Such reflections can make vital contributions to educational success in rapidly changing and highly differentiated world conditions. Assessment instruments or tests can contribute to such deliberations in significant ways.

Thus, we propose that such tests be employed in evaluation practices that:

- are dialogic in nature and situated primarily in local knowledges and practices.
- include multiple stakeholders and, importantly, those whose performance is under evaluation.
- include multiple criteria, reflecting the needs and values of multiple stakeholders.
- center on processes of continuous improvement of whole systems, including students, teachers and surrounding communities.
- provide for the education of teachers, school administrators and other relevant stakeholders in dialogic-centered practices of evaluation.

What could this mean for the future of testing and measurement instruments more generally? In our view, the above recommendations would not mean a diminishment, but rather, an enrichment and expansion of such services. With respect to testing, there is good reason to move toward the following:

- *Making standardized tests available (as opposed to mandatory) to all educational institutions.* Whether and how local school systems or districts employ test scores in their deliberations should be locally determined.
- *Radically expanding the kinds of tests available to schools for evaluating students.* For example, depending on locale, schools might wish to have tests that would enable them to benchmark students in terms of computer literacy, career fluency, civic and political participation, bilingual capacity, dialogic skills, environmental knowledge, musical aptitude, physical competence, health, and so on.

Continued on page 12

Continued from page 11

- *Expanding the range of tests available to schools and school districts for evaluating their own development.* For example, schools might varyingly wish to benchmark themselves in terms of parental participation, excellence as a learning community, internal collaboration, civic contribution, relationships with business and government, and the like.
- *Offering educational services enabling local schools to generate effective practices of participatory evaluation.*

In conclusion, if we take into account the increasing development of communication technologies and the resulting shifts in demands and opportunities, it is imperative to explore new ways of practicing evaluation. Along with Nussbaum (2011), we argue here for evaluation in the service of creating capabilities as opposed to judging them.

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

The Gordon Commission was established by ETS to investigate and advise on the nature and use of educational testing in the 21st century. 21463