The Gordon Commission
on the Future of Assessment in K–12 Education

Gordon Commission

Fellow Synthesis Paper

Towards Assessment Vérité

Sherice N. Clarke

Learning Research and Development Center
University of Pittsburgh

Acknowledgments: I gratefully acknowledge the contributions of Professor Edmund W. Gordon, Professor Ernest Worrell and the Gordon Commission Fellows (Keena Arbuthnot, Amanda Johnson, Juliette Lyons-Thomas, Jordon Morris and Catherine Voulgarides) in helping to develop my thinking for this paper.

Introduction

This paper is organized into two parts. The first part summarizes the commission papers included in this synthesis paper. These papers are ordered thematically from the philosophical and theoretical examinations of assessment (Bereiter and Scardamalia, 2012; Cauce and Gordon, 2012; Gorin, 2012) to the implications of technology on assessment (Behrens and DiCerbo, 2012; Chung, 2012) and, finally, to the use of assessment for teaching and learning (Amour-Thomas and Gordon, 2012; Hill, 2012). The second part looks across these papers and synthesizes the substantive themes.

In my commentary, I introduce the term *assessment vérité*, borrowed from the mid-century filmmaking movement, *cinema vérité*, which pushed toward greater realism in technique and subject matter in film as a means to undercover aspects of the human experience. I found the idea of striving for greater realism in technique and subject particularly useful for characterizing the major themes of the papers included in this synthesis. Thus, I will use the term *assessment vérité* to refer to the use of ethnographic methods to measure learning in naturalistic conditions in order to understand students' knowledge and skills in development with greater precision. My commentary will consider how these papers work toward the idea of *assessment vérité* and the implications of this concept on assessment.

Bereiter, Carl and Scardamalia, Marlene (2012). *What Will It Mean To Be An Educated Person in Mid-21st Century?*

Overview

A little more than a decade into the 21st century, Bereiter and Scardamalia's discussion imagines *what it will mean to be an educated person in the mid 21st century*, considering the implications on the processes and outcomes of education for decades to come.

Examining the effect of recent technological advances on ways of being and doing in this part of the 21st century suggests that the future will bring yet more new and unforeseen challenges and inevitably more change. Consequently, the question the authors pose to policy makers and educators is: what model of education should be used to support learners in changing and uncertain times? What competencies and "habits of mind" should educators work to foster in the next decades?

Author Recommendations

The authors posit that a mid-21st century person should possess a range of competencies, including possessing a wide range of knowledge and understanding problems at various levels of depth and complexity.

Education for changing times would necessitate preparing learners to engage in lifelong learning and enabling learners gain new competencies while adapting to the "accelerating pace of change." They suggest that education foster breadth, depth, and the ability to navigate diverse ideas, peoples, and cultures. In addition, education should continue to develop students' capacity to engage in complex reasoning and rationality and seek to develop learners into thinking persons.

In particular, the authors highlight the following areas of need that schools do not yet adequately address, toward developing knowledgeable persons in the mid-21st century:

1. understanding nature of knowledge and capacity to create new knowledge
2. working flexibly with the concrete and abstract
3. understanding complexity scientifically
4. developing collective cognitive responsibility

Implications

Bereiter and Scardamalia pose the questions: who do we (society) want to be in terms of knowledge and competencies in the mid-21st century, and what can education do in the interim to

actualize this vision? An important point to bring forward from their argument is that as our learning, cognitive, and social theories develop, so should our instructional purposes and approaches. I would add that assessment also should adapt with theory and instruction. That is, as we develop more complex models of cognition (e.g., social factors of learning), assessments should adapt to capture our emerging models of cognition. For example, as we adopt competencies like knowledge creation as desired outcomes of education, instruction should adapt so that learners are given the space to create new knowledge. An implication for assessment might mean more emphasis being placed on learners' ability to create, manipulate, and transfer new knowledge. Likewise, as our theories of collaborative learning develop, learners should be given instructional space to collaborate, and assessment should adapt so that we might be able to measure and evaluate individual and collaborative contributions to solving problems.

Cauce, Ana Marie and Gordon, Edmund W. (2012). *Toward the Measurement of Human Agency and the Disposition to Express It.*

Overview

If we try to unpack educational achievement, considering the conditions that help to foster educational outcomes and the requisite capacities and abilities, it becomes clear that these capacities alone cannot lead to educational achievement. Cauce and Gordon suggest that human agency, and its expression, is a central mechanism, albeit under-theorized, that helps to drive educational outcomes. Drawing on theoretical and methodological insights from social psychology, sociology, and economics, they present a multidimensional conceptualization of human agency, considering its relationship to cognitive and educational outcomes, and how it might be developed as a critical aspect of the educational process.

Cauce and Gordon define human agency as "the capacity and disposition to recognize and act in one's own best interest and that of chosen others." They suggest that knowledge, skill, and

social context are pre-conditions for agency, where social context can enable or preclude its expression. One of the critical insights from economic and social theory on human agency is that it cannot be expressed if the social context is one that prohibits choice. They conceptualize human agency as the mechanism that mobilizes our human resources toward achieving accomplishments.

Author Recommendations

Cauce and Gordon argue for the development of human agency to be an end that education seeks to foster, as it not only creates the conditions that support true democracy and participation, but also may serve to support yet greater agency in future generations. They suggest that researchers and educators develop methods for promoting students' agency. In addition, they urge for the development of holistic measures of agency, beyond survey methods, to provide more robust data for understanding learners' agency, its emergence, and its expressions. They also suggest that research and measures could be developed to examine how educators and administrations could help to support or preclude the development of student agency.

Implications

Cauce and Gordon's multidimensional model of human agency presents the opportunity to think about multidimensional approaches for developing human agency with and through education. Their model bridges sociological and psychological definitions of human agency, where the sociological emphasizes the role of context in enabling or precluding agency and the psychological emphasizes the role of the individual in mobilizing their competencies toward accomplishments. The former has implications for what education, writ large, can do to foster agency (e.g., enabling opportunities and choice), whereas the latter has implications for what

educators might do to help foster agency and its expressions (e.g., help learners recognize their and mobilize their competencies towards accomplishments). Future directions for research might examine how these levels of a multidimensional model of human agency work in concert toward the development of agency.

Gorin, Joanna (2012). *Assessment as Evidential Reasoning*

Overview

Assessment as evidential reasoning advances Mislevy's (2006) notion of assessment as an evidentiary argument, which brings forward two critical aspects of the assessment process: 1) the purpose of assessment (i.e., to make claims about what students know and can do) and 2) the nature of the evidence (i.e., the quality and quantity of evidence used to advance claims about students' knowledge and ability). In the current assessment paradigm, testing enables the observation and collection of evidence of students' knowledge and ability. The question that Gorin poses is whether the evidence that we have been using to make inferences about student ability and, subsequently, claims about knowledge and ability, is the most valid and reliable evidence. Conceptualizing assessment as argument, drawing on Toulmin's (1958) framework of argumentation, places the robustness of evidence-based claims at the center of assessment design and development. Thus, the challenge for assessment design becomes:

1. What constitutes the best evidence with which the most persuasive arguments and inferences about student knowledge and ability can be made?
2. Under what conditions can we produce robust evidence of student knowledge and ability in order to make these claims?

Gorin suggests that if the future of educational assessment is to take on these questions, it would require a fundamental shift from current views of assessment purpose and its design.

Author Recommendations

Gorin proposes that assessment design as evidentiary argument build on Mislevy's (Mislevy and Haertel, 2006) Evidence Centered Design (ECD) approach, which is a framework for developing assessments that considers 1) the most relevant claims we want to make as a result of assessments, 2) the most robust evidence to support those claims, and 3) the assessment environment that is needed for students to display the behaviors we wish to observe.

There are several implications with respect to the assessment paradigm shift that Gorin proposes. First, it will necessitate a reexamination of precisely what we wish to know about students' knowledge and ability and the quality of the evidence produced by current assessments. Second, as new educational goals, objectives, and theories of learning are adopted and practiced, new constructs and forms of evidence may be needed to make claims about student learning; e.g., collaborative learning, human agency, and self regulation. Finally, the technological revolution enables the collection of new forms of evidence on student learning that can be embedded in the learning process, and collected continuously. Technology creates new possibilities in terms of assessment items and tasks which can build on contemporary theories of learning to facilitate the observation of complex cognitive processes, thereby providing multifaceted evidence that can aid in the analysis of these data.

Implications

There are two aspects of Gorin's model of assessment as evidentiary argument that I feel are important to bring forward. First is the role of evidence for making claims about student knowledge. Second is the importance of the conditions under which the best evidence might be generated.

With respect to evidence, conceptualizing assessment as the production of evidence from which claims about knowledge can be made brings the question of validity of evidence into the discourse on assessment. Gorin's model conceptualizes the validity of evidence as a driver of assessment design. For example, if we consider a problem of practice, such as the assessment of English Language Learners' (ELL) listening comprehension, we know that multiple tasks might be used to elicit evidence of listening comprehension (e.g., dictation, comprehension questions, note-taking). The question of validity necessitates that the assessment designer examine the evidence in relation to a learning claim. In this sense, would dictation provide the best evidence of understanding or lexical recognition? To make claims about comprehension, additional forms of evidence might be needed.

With respect to context, Gorin's model demands the same attention be given to the conditions under which assessment evidence can be produced. Using the same example case, ELLs' listening comprehension, a dictation task would require a student to recognize lexical items in sentential context and write them. These conditions would require them to make connections between phonological and orthographic knowledge of the target language, but might do so in a form that is abstract from why language is used. Thus, a more contextualized task for the skills that dictation elicits might be more appropriate. Gorin's model helps to bring into focus the behaviors that certain conditions produce.

Behrens, John T. and Dicerbo, Kristen E. (2012). *Technological Implications for Assessment Ecosystems: Opportunities for Digital Technology to Advance Assessment.*

Overview

The digital revolution has given way to a host of new technologies that extend human capabilities. New methods of examining the natural and social world are now possible, including

understanding learning, and thus provide a unique opportunity to completely re-imagine educational assessment. In this essay, Behrens and DiCerbo provide a road map that examines how digital technologies can be collected, analyzed, and used to model activity, and thus can be harnessed for advancing educational assessment.

Behrens and DiCerbo identify three core aspects of technological developments that can be exploited for educational assessment. First, computers enhance humans' capabilities in their ability to store, process, and mine large sources of fine-grain data. What this means for assessment is that new insights can be gleaned of student learning using computational methods of storing, analyzing, and modeling student data. Second, the digital age and our increasing use of digital technologies makes it possible to gather new forms of data based on human interaction in digital environments. Data capture, such as player interaction in a computer game, or clicks and scans on a mobile device, can therefore be both ubiquitous and unobtrusive ways of collecting data, making it possible to gather rich, fine-grain data of human performance with which to make inferences about student knowledge, ability, and learning. Finally, digital technologies can visualize fine-grain data so that observations and inferences can be made of these data; e.g., visualizations of eye-tracking data of students' visual attention in a cognitive tutor to understand which features learners attend to in order to solve problems (e.g., Letourneau, 2012).

Author Recommendations

Behrens and Dicerbo highlight several implications of the aforementioned technological developments on assessment that could be transformative in nature, including:

1. shifting from an item paradigm to an activity paradigm
2. shifting from an individual paradigm to a social paradigm
3. shifting from "assessment situations" to assessment ecosystems

Moving toward an activity paradigm reframes assessment practices from identifying correctness of items to capturing a constellation of learning transactions using digital technologies to make inferences about student cognition and learning. Shifting toward a social paradigm would bring assessment closer to the conditions where existing theory and empirical research suggest learning occurs through collaboration. Finally, Behrens and Dicerbo argue that developing an assessment ecosystem would help to counter the disruptive nature of the present assessment paradigm by building on the digital technology to extract data unobtrusively that could provide rich data on student learning.

Implications

While technology presents incredible opportunities for the assessment system, I would like to highlight some of the implications of Behrens and DiCerbo's proposed model. First, more research is needed on how process data collected unobtrusively correlates to cognitive ability. In other words, do these new methods of data collection constitute more accurate assessments of cognitive ability and/or cognitive processes or are they simply measuring new phenomena; e.g., student engagement? Chung (2012) provides some initial studies to address this question, but I would argue that more research is needed if these forms of ubiquitous assessment are to provide meaningful data on student learning and development. Second, the collection of fine-grain data as the authors propose could constitute large corpora and essentially a record of individual behavior and activity. What are the ethical implications of the collection of these data in terms of ownership and use? For example, could transactional data be used to perpetuate the achievement gap with respect to groups already underserved by education? Finally, as Chung (2012) suggests, one advantage of technological advances is they can make possible the visualization of fine-grain

data on students' emergent learning in the manner of assessment as dynamic pedagogy, for teachers intervene in the service of student learning.

Chung, Greg (2012). *Toward the Relational Management of Educational Measurement Data.*

Overview

We now have the capacity to collect large corpora of fine-grain data on student behavior in digital environments. This data can be used to help understand what students know. Less is known, however, about precisely what insights these fine-grain data can provide on cognitive and affective processes, and how these process data can be used for adaptive instruction to enhance learning. Chung reviews the evidence on the use of data on learners for instructional interventions and considers their implications for teaching and learning.

Examining the relationship between interaction data in digital environments and think-aloud data of those same interactions, Chung found a statistically significant correlation suggesting that when a task required reasoning, learners' online behavior was problem-solving oriented. Testing the extent to which students' online behavior was problem-solving oriented, Chung and colleagues adapted a user interface to trace the features of the interface design to which students attended when solving problems. They found that high achieving learners attended to the most essential information to solve problems and were faster at solving problems than low achieving students. Chung suggests that this provides evidence that learners' online behavior is deliberate and systematic, and correlated with cognitive processes. In a third study, student behavior in a designed gaming environment was analyzed in order to understand the extent to which students' behavior in this digital environment displayed students' understanding and misconceptions. Like the aforementioned studies, they argue that learners' behavior in a gaming environment was both systematic and purposeful.

Author Recommendations

All three examples provide evidence for how fine-grain data helps to model cognition in development. Chung argues, however, that more empirical research is needed to fully understand the relationship between fine-grain data, learning processes, and educational outcomes.

Chung suggests developing adaptive educational systems that use fine-grain data on student learning to adapt instruction to individual needs in order to enhance learning processes and outcomes. In addition, he suggests that by putting the data, and its representations, in the hands of educators, they can be better equipped to make decisions about how to adapt instruction to support student learning.

Implications

Technology enables new ways of collecting data on student behavior and, as a result, new inferences can be made of student learning processes. However, as Chung highlights, more research is needed to verify the extent to which these fine-grain data can be correlated with cognitive processes. For example, does keystroke or mouse-click data provide robust measures of cognitive processes and student knowledge? Educational neuroscience opens up a wealth of new possibilities for relating behavioral data with cognitive activity and cognitive processes. For example, like Chung, CREATE Lab is examining student attention in a digital learning environment to problem solving using eye-tracking data (Letoureau, 2012). These data are used to better design the learning environment, but it also might be possible to use these kinds of measures to better understand cognitive processes in digital environments.

Armour-Thomas, Eleanor and Gordon, Edmund W. (2012). *Towards an Understanding of Assessment as a Dynamic Component of Pedagogy.*

Overview

The current assessment paradigm, which has given primacy to year-end standardized achievement tests, yields information on student proficiency and performance that makes possible the stratification of students into classes of proficiency. These kinds of assessments, however, are limited in their ability to provide information about students that can be used to support student learning. If the charge of assessment were conceptualized as one where supporting learning is at its very core, then *pedagogy* needs to be re-conceptualized as a dynamic process in which assessment, curriculum, and instruction work together to support student learning.

Armour-Thomas and Gordon propose a model of assessment as a dynamic component of pedagogy, in which assessment provides essential feedback that can be used to adapt curriculum and instruction to optimize learning. This model of formative assessment would bring assessments into the classroom and serve to monitor students and provide feedback for teachers on students' learning processes, including prior and emergent learning, transfer, and metacognitive processes. These data can be leveraged to make appropriate selections of texts, tasks, and processes that promote learning of subject matter; in short, adapting the curriculum to student needs.

Author Recommendations

The authors provide a framework that serves as a guide for examining the interdependency of assessment, curriculum, and instruction in the service of learning. They suggest specifying learning goals and objectives and examining how current assessments are congruent with those goals and objectives. They argue that these forms of assessment can be integrated into all phases

of instruction; e.g., assessments of prior knowledge, emergent learning, and consolidation. Their framework suggests examining the extent to which assessments provide information for teachers to adapt instruction and curriculum, and learners adapt learning.

Given the current assessment climate, adoption of pedagogically driven assessments, like the ones the authors propose, would require changes at all levels of the assessment system for these kinds of assessments to become practice and count as a measure of student learning. Toward this end, the authors propose the following:

- make learning-centered assessments count in the evaluation of teaching and learning
- use computer technologies to develop learning-centered assessments
- ensure the validity and fairness of learning-centered assessments
- train teachers to use assessment as a dynamic component of pedagogy

Implications

Amour-Thomas and Gordon explicate the architecture of assessments as a dynamic component of pedagogy. A critical component of these forms of assessments, I believe, is that they provide a unique opportunity to steer assessments toward both productive purposes and productive uses of assessment data (Ho, 2012). The post-NCLB era finds instruction as distinctly test-driven, however, with an eye toward multi-level accountability. In this sense, the assessment system shapes instructional topics, instructional approaches, and the time spent on test preparation. We can characterize the purposes of these practices as instrumental in nature, targeted toward student achievement on assessments, and the uses of assessment data as inherently summative, focused on accountability. If the purpose of assessment becomes supporting learning and adapting instruction and curricula, then the uses of these assessment data become important tools to support teachers in making informed, student-relevant decisions about instruction and curricula in the service of student learning. Formative uses of assessment

outcomes might help to support teachers to integrate these forms of assessments into instructional practice.

Hill, Clifford (2012). *Assessment in the Service of Teaching and Learning.*

Overview

Tracing the genesis of assessment from Imperial China and its evolution over the last three centuries as the use of assessment moved West, Hill provides a historical and comparative perspective on assessment that beckons a reexamination of contemporary purposes and methods of assessment.

The purposes of assessment over time have shifted from functioning as a shibboleth to a tool used to advance meritocratic aims in order to widen access to education to a tool used to certify cultural competence and a tool used to integrate learners from schooling into the professional world. As assessment became a practice in the United States, we saw traces of the earlier models of assessments, such as an emphasis on meritocratic and egalitarian ideals, however, as Hill notes, the invention of machine-scoring multiple-choice assessments drove test design and content. Test construction driven by scoring methodology has introduced a host of constraining features in these assessments that are antithetical to current theories of reading and text comprehension.

Hill offers two alternative models of assessment to bridge the gap between the purposes of assessment and the methodological tools we use to assess. The Digital Testing Model evaluates students' reading and media literacy through a series of theoretically grounded activities that scaffold reading comprehension. The second alternative model, the Digital Project Model, evaluates students' media literacy through a series of activities designed to scaffold research, analysis, and presentation of findings using a suite of digital tools. What both models of reading

assessment have in common is that they present a suite of activities modeled on theories of text comprehension to serve the dual purpose of supporting student learning and providing measures of student learning.

Author Recommendations

Hill argues that the integration of a digital assessment model could produce a rich archive on student development and learning. Integration of digital assessments into regular classroom practice could begin to diminish the lines between testing situations, instruction, and learning. In addition, teacher involvement in the evaluation of student learning using scoring rubrics in a digital assessment model could help to heighten their awareness of the learning goals and objectives, which also might help to infuse these standards into their daily instructional practices.

Hill recommends that an infrastructure be developed to help to legitimate these forms of assessments so that they might penetrate the assessment system and become widespread practice. Among his recommendations are:

- generate digital archives of student performance and achievement
- certify digital archive data so that institutions might use this data in place of standardized tests; e.g., SAT

Implications

Hill provides a practical model of assessment as a dynamic component of pedagogy (Amour-Thomas and Gordon, 2012). There are three implications of Hill's digital assessment model that I would like to bring forward. First, the emphasis on assessing authentic activities, with authentic materials, in authentic digital environments, akin to what Varenne (2012) refers to as "ethno-assessment," "ethno-curriculum," and "ethno-pedagogy"; i.e., assessments, curriculum, and pedagogy that emerge from everyday activities in the service of learning and development. The ethno of ethno-assessments in the model that Hill proposes serves to bring

assessment practices closer to real-world activities and, thus, real-world competencies. Second, a retrospective view of assessment is helpful to be reflexive about contemporary purposes of assessment, as well as envisioning other models. It raises some fundamental questions, such as whether our assessments should be shaped by our scoring methodology or our theories of learning, to which I would argue for the latter. Finally, there is a need to develop alternative scoring methods for formative assessments like the Digital Testing Model. Hill's recommendations for scoring digital assessments are expensive and, as a result, may limit widespread use. Other computational methods are needed to assess linguistic data if we are to move away from multiple-choice scoring. As technology develops, there may be new possibilities for measures and scoring complex assessment data.

Commentary: Toward *Assessment Vérité*

We find ourselves at a tipping point with respect to educational assessment. A primary goal of the Gordon Commission on the Future of Assessment in Education is to take stock of assessment means and ends in order to imagine future possibilities. In doing so, the authors have revealed aspects of the current assessment system that are particularly problematic (Bereiter and Scardamalia, 2012; Hill, 2012; Gorin, 2012) and have considered emerging possibilities in light of enabling technology, which could fundamentally change the way we do assessment (Behrens and DiCerbo, 2012; Chung, 2012). What is common across these papers is their emphasis on naturalistic observation in naturalistic conditions to make inferences about student knowledge and ability, or what I will refer to as *assessment vérité*.

Technology has enabled new ways of collecting evidence from which to make inferences about student learning. Behrens and DiCerbo, Chung, and Hill all highlight how it is now possible to collect data on a wide range of student behavior in computer environments, which can constitute evidence on student knowledge and ability; e.g., keystrokes, eye-tracking, and timestamp data. These forms of data present a fine-grain picture of cognition-in-interaction, from which it might be possible to gain greater insight into learning processes. Chung's paper reviews some of the current evidence that suggests online behavior is correlated with cognitive processes. More of this kind of research is needed, to which I would add, using different methodologies. However, what these studies begin to show are possible directions in understanding process data and its relationship to cognition. It also might be the case that these forms of data generate evidence of cognitive and social processes that have not previously been measured in formal assessments; e.g., attention and agency. In addition to technology's capacity to collect new forms

of assessment data, these tools make it possible to collect a greater quantity of data on students, from which modeling student development longitudinally is possible.

The above technological possibilities are all concerned with naturalistic observation of student ability. That is, moving away from what Behrens and DiCerbo refer to as an item paradigm to an activity paradigm in assessment by capturing student processes in interaction. One of the fundamental questions that the commission authors propose is whether the current item paradigm produces adequate evidence of student ability (Gorin, 2012; Chung, 2012; Behrens and DiCerbo, 2012; Hill, 2012). Limitations of assessment models like year-end standardized tests with multiple-choice items make it difficult to produce and observe the targeted behaviors assessments are designed to elicit. Examining learning processes naturalistically, as the authors propose, could mean generating educative tasks that draw on theories of learning and contextualize the skills and abilities we hope to elicit. Data that is gathered in these generative assessment environments could then be used to understand student learning, as well as be used to help to support it (Hill, 2012; Chung, 2012; Behrens and DiCerbo, 2012).

In other words, naturalistic observation could present an opportunity to converge learning and assessment situations, such as Amour-Thomas and Hill propose, as it necessitates the careful design of an environment (or conditions) under which those observations could be made. For example, if the goal of instruction is to increase students' dialogic reasoning (O'Connor and Michaels, 2007), assessment designers would have to create the conditions within which dialogic reasoning might be observed. Generating those conditions would, in effect, create a situation in which reasoning could be both displayed and developed. This might make increasingly obsolete

the concept of a separate occasion in which assessments data are recorded, as technology enables the collection of student activity in situ.

One of the advantages of moving toward a model of *assessment vérité* is that it might help to make assessment more of an ubiquitous part of education, rather than a disruptive in the form of year-end high-stakes tests. That is, aligning in-school assessments to the ways in which assessment operates in out-of-school environments; e.g., in scientific research, on the basketball court, in museums, and in child rearing (Osborne, 2010; Varenne, 2012). Another advantage is the idea of assessment in situ; that is, assessment occurring in the places where the desired behavior occurs. Drawing the power of technology, this might be in spaces like gaming environments, class discussions, collaborative reading activities, or problem solving. These are spaces that contemporary learning theory suggests are productive for learning, but are not yet reflected on high-stakes tests.

However, a disadvantage of a model like *assessment vérité* might be that more methodological and technological development is needed to understand complex cognitive processes in dynamic learning environments like classrooms. Using the example of classroom discussion from my own line of research, we are just beginning to model growth in dialogue over time to understand the affordances of participating in class discussions (Clarke et al., 2012). More work is needed to better understand dialogic learning in complex and dynamic instructional environments like classrooms.

References

Amour-Thomas, E. & Gordon, E. W. (2012). *Toward an Understanding of Assessment as a Dynamic Component of Pedagogy.* Report of the Gordon Commission on the Future of Educational Assessment. Princeton, NJ: Educational Testing Service.

Behrens, J. T. & DiCerbo, K. E. (2012). *Technological Implications for Assessment Ecosystems: Opportunities for Digital Technology to Advance Assesment.* Report of the Gordon

Commission on the Future of Educational Assessment. Princeton, NJ: Educational Testing Service.

Bereiter, C. & Scardamalia, M. (2012). *What will it mean to be an educated person in the mid-21st century?* Report of the Gordon Commission on the Future of Educational Assessment. Princeton, NJ: Educational Testing Service.

Cauce, A. M. & Gordon, E.W. (2012). *Toward the Measurement of Human Agency and the Disposition to Express It.* Report of the Gordon Commission on the Future of Educational Assessment. Princeton, NJ: Educational Testing Service.

Chung, G. K. W. K. (2012). *Toward the Relational Management of Educational Measurement Data.* Report of the Gordon Commission on the Future of Educational Assessment. Princeton, NJ: Educational Testing Service.

Clarke, S. N., Resnick, L. B., Rose, C. P., Chen, G., Stainton, C., Katz, S., . . . Granger, R. (2012). *Torwards Discursive Instruction: From I-R-E to Accountable Talk.* Paper presented at the LearnLab's Annual Learning Science Workshop on Use of Technology Toward Enhancing Achievement and Equity in the 21st Century, Pittsburgh, PA.

Gorin, J. S. (2012). *Assessment as Evidential Reasoning.* Report of the Gordon Commission on the Future of Educational Assessment. Princeton, NJ: Educational Testing Service.

Hill, C. (2012). *Assessment in the Service of Teaching and Learning. Report of the Gordon Commission on the Future of Educational Assessment.* Princeton, NJ: Educational Testing Service.

Ho, A. (2012). *Variety and Drift in the Functions and Purposes of Assessment in Education.* Report of the Gordon Commission on the Future of Educational Assessment. Princeton, NJ: Educational Testing Service.

Letourneau, S. M. (2012). *Real-time interactions between attention and behavior in multimedia learning environments*. Paper presented at the Annual LearnLab Learning Science Workshop on the Use of Technology Toward Enhancing Achievement and Equity in the 21st Century, Pittsburgh, PA.

Mislevy, R. J. & Haertel, G. D. (2006). *Implications of Evidence-Centered Design for Educational Testing. Educational Measurement: Issues and Practice, 25*(4), 6-20.

O'Connor, C. & Michaels, S. (2007). *When Is Dialogue "Dialogic"? Human Development, 50*(5), 275-285.

Osborne, J. (2010). *Arguing to Learn in Science: The Role of Collaborative, Critical Discourse. Science, 328*(5977), 463-466. doi: 10.1126/science.1183944

Toulmin, S. (1958). *The Uses of Argument*. Cambridge: Cambridge University Press.

Varenne, H. (2012). *Education: Constraints and Possibilities in Imagining New Ways to Assess Rights, Duties and Priviledges.* Report of the Gordon Commission on the Future of Educational Assessment. Princeton, NJ: Educational Testing Service.