



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

Gordon Commission
Fellow Synthesis Paper

Juliette Lyons-Thomas
University of British Columbia

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

Findings

This paper proposes that a “dynamic pedagogy,” that is, the interaction between assessment, curriculum, and instruction, is instrumental to student learning. The authors make the case that each of the three strands is related to one another in the context of learning, showing similarities, and ways in which one can be used to inform the others. Given these relationships, the authors then provide a framework for examining assessment within the dynamic pedagogy model. Finally, the authors provide a number of suggestions if assessments are to be learning-centered. These suggestions include using learning-centered assessments when evaluating both teachers and students, including computer-based assessments, ensuring that validity and fairness is present in these assessments, and changing the way in which teachers are prepared for their service so that assessment is viewed as a practice that is embedded in daily practice.

Commentary

This paper offers a thoughtful approach to using assessment as one of three components to maximize student learning. What is especially advantageous about the paper is that it does not offer assessment as a stand-alone method, but rather acknowledges that other components cannot and should not be teased away from it.

The questions asked with respect to phases of learning (pg. 17-18) can be especially useful as purposes for validity if assessments are to be used for learning-centered assessments. Before assessment strategies are determined, these purposes should be solidified as intended uses (and of course, not changed once the assessment has taken place).

Finally, the four points addressed as future directions are each important and vital in creating useful assessments for dynamic pedagogy. If assessments are to be truly useful for

learners and educators, these points should each be given the attention and concern that they require. Though each future direction listed in this paper includes a paragraph that briefly explains how to promote learning-centered assessments, each topic certainly deserves its own paper that goes into depth to explain how this view of assessment can be realized.

Recommendations

In order to ensure that assessments are learning-centered, each of the four future directions provided in this paper should be further examined. For instance, one idea that can be expanded upon is the suggestion that a common form of assessment should be used for both learning-centered assessment and the evaluation of learning and teaching. As has been pointed out in other papers, using assessments for multiple purposes is common, though not always ideal. It would be interesting to see a discussion on the type of assessment that the authors have in mind, with details about the elements that would be included. Further discussion on this topic could ensure that the assessment is appropriate for its purposes.

With respect to including aligning assessment, curriculum, and instruction, a model which may be useful in communicating this idea is Wilson and Sloan's (2000) Berkeley Evaluation and Assessment Research (BEAR) Assessment System, which essentially embeds assessment into the teachers' instruction, and lists a match between instruction and assessment as one of its main goals (Ercikan, 2006).

Bereiter, Carl and Scardamalia, Marlene. (2012). *What Will it Mean to be an Educated Person in the Mid-21st Century?*

Findings

In their paper, the authors discuss the changing characteristics of an educated person in this century compared to those from past centuries. It is argued that while many former traits still hold true, changes in society have resulted in an additional set of qualities. As well, the role that technology has played in individual and group behaviors is discussed.

Throughout the article, the authors point to relatively recent ways in which a person must successfully navigate his or her world: a person must be cosmopolitan, able to understand and critique information from various forms of media, and have moral reasoning skills beyond what has been required in the past. The authors also specify four traits which they see as being “real 21st century competencies”: 1) knowledge creating, 2) working with abstractions, 3) systems thinking, and 4) collective cognitive responsibility. Finally, it is acknowledged that technology surely plays a role in the assessment of competencies, though it is difficult to determine how it will be used in the future.

Commentary

This article provides a comprehensive list of abilities that are needed now and in the next 40 years. These abilities represent a variety of domains; for instance, changes in moral reasoning based on globalization brings together two concepts that now affect one another.

This article serves as a starting point for developing assessments of the competencies listed here. Though the paper does not appear to provide many specific suggestions for how it can be done, even acknowledging that they are difficult to assess, some competencies that test developers may choose to target at least have been identified.

Recommendations

In the Implications for Measurement and Assessment section, a good deal of the discussion is aimed toward the ability to assess critical thinking skills. Something that is not mentioned here is the area of research that examines cognitive processes used for assessing critical thinking. For instance, Leighton (in press) describes some possible strategies for ensuring that students are actually using the higher order processes that assessments more frequently require. Including cognitive processes in test design may be a helpful approach to assessing critical thinking skills, especially if it is something that future assessments will increasingly entail.

Dixon-Roman, Ezekiel and Gergen, Kenneth (2012). *Epistemology and Measurement: Paradigms and Practices*.

Findings

This paper describes the history and philosophy of assessment, focusing on the positivist roots of measurement. The paper begins by describing the conflict between testing in education and the global changes that have taken place in large part due to technological advances, and then takes a look back to the origins of measurement, including classical test theory and modern test theory. The paper ends by briefly discussing emerging postmodern approaches to understanding measurement, pointing out that the aims of measurement are at odds with the social relational processes of education. It also sets up the second of the two papers, which focuses on social epistemology.

Commentary

The topics discussed in this paper are important to anyone in education, but especially those who study measurement and are interested in large scale assessment issues. Efforts to describe and understand the history of assessment, including the misuses of testing (e.g., *The Mismeasure of Man*), are essential for ensuring that past mistakes are not repeated by fledgling test developers.

One sentence in the first paragraph of the paper was especially notable, and properly sets the tone of the article. This sentence asserts that “policy makers have assumed that the instruments of measurement are sound, robust, and valid” (p.1). These assumptions are undoubtedly based in part on the technical foundations of measurement, and while incredibly important in many respects, should not be trusted blindly as they appear to be (by a broader array of people than just policy makers).

Recommendations

One suggestion stems from the way in which the two concepts (positivism and social constructionism) are polarized. While they are unarguably very different concepts in the area of measurement, they both offer ways in which to improve assessment. Rather than positioning them as competing epistemologies, it may be helpful to present them as complimentary to one another in the area of assessment.

Overall, the paper is very useful in clarifying the roots of measurement. It should be used to explain the origins of practices that are still used today, while pointing out that other epistemologies may be beneficial for framing assessment in the 21st century.

Gergen, Kenneth and Dixon-Roman, Ezekiel (2012). *Social Epistemology and the Pragmatics of Assessment*.

Findings

This paper serves as a follow up to Part I, which outlined fundamental problems with the empiricist tradition of measurement. In this paper, the authors present an alternative approach to measurement, moving away from the previous positivist/empiricist epistemology and instead regarding measurement from a social constructionist perspective. In particular, the authors point toward changes in the global use of technology as a key factor for why this shift should take place.

Commentary

This paper is useful in that it shows the great value of moving measurement away from the traditional epistemology of positivism in favor of a social constructionist approach, especially given the transformation that we, as a society, are currently undergoing and moving toward. Aside from that, a notable strength of this paper is that it provides concrete

recommendations and warnings with respect to addressing assessment in the 21st century. The last line of the paper perhaps best sums up the hopeful results of these recommendations for assessment; i.e., the authors are optimistic that that these methods will lead to “evaluation in the service of creating capabilities as opposed to judging them” (p.15).

Recommendations

One suggestion for the future is to take the warnings (i.e., the points that the authors suggest moving away from) and link them with existing practices that are used in American education.

Furthermore, the recommendations that the paper presents should, at some point, be linked to specific assessment strategies that could be followed by education systems. The final page of the paper provides some context of how schools can use the recommendations, but specific examples could help turn the suggestions into reality. For instance, one recommendation is to “radically [expand] the kinds of tests available to schools for evaluating” (p.14). A next step would be to provide examples of tests that would allow schools to assess computer literacy, or perhaps recommend a strategy for policy makers to follow so that schools are using similar tests across the country.

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

Findings

This paper describes the use of assessments an evidential argument. In particular, it draws upon Mislevy's (2006) concept of building an evidentiary argument about what a student knows or is able to do, rather than using assessments as tests. The article discusses how this can be achieved by using various forms of assessments, such as alternative assessments, medical assessments, and information technology, and also points out that in gathering evidence for 21st century skills, new types of data are likely needed to build a new type of argument.

Commentary

Though assessments as an evidential argument is covered in other papers (e.g., it is listed as one of four metaphors for understanding assessments in Mislevy's paper), the suggestions presented in this paper are motivating for making this a reality. Not only is re-conceptualizing the evidence (section III) described with respect to changes in the educational context, but item types, data sources, and advances in analyses are described as well.

Recommendations

It is clear that novel types of evidence are needed for an evidentiary argument of 21st century competencies, especially given the topics presented in other papers (notably, the Bereiter and Scardamalia paper). However, this paper is especially useful because it presents some ways of collecting that evidence and advocates for multiple sources of information.

The method of collecting response processes is very interesting and adds to the typical method of having students think out loud (although this method also could be used and improved upon). These methods are relatively new compared to other ways of collecting information about assessments, and should be encouraged as investigative methodologies in both academic settings and testing organizations.

Ho, Andrew (2012). *Variety and Drift in the Functions and Purposes of Assessment in Education*.

Findings

Ho's paper examines the role that validity plays in educational assessment and, in particular, discusses 1) Haertel's classification of purposes of assessment while contrasting it with other prominent frameworks and discussing and proposing multiple dimensions in which the framework might work, and 2) considers the factors that cause the purpose of assessments to change over time (appropriately referred to as "purpose drift") while suggesting some strategies that may be useful in ensuring that tests are used appropriately.

Commentary

Ho's examination of key validity players' frameworks for the purposes of assessment is comprehensive and very useful for categorizing the many uses of assessment.

In *Knowing What Students Know*, it is acknowledged that assessments are often used for multiple purposes. Notably, however, it also is stated that "the more purposes a single assessment aims to serve, the more each purpose will be compromised" (NRC, 2001, p.40-41). Ho's paper, and especially the section on purpose drift, does a great job of explaining why multiple purposes of assessments can be problematic, as well as providing real life examples for consideration.

The paper suggests "raising the standards of validation to proactive efforts from developers, policy makers, and analysts to anticipate and eventually stem the drift of assessment score purpose along easily anticipated vectors" (p. 14). This is an appropriate course of action for the widespread uses, and potential misuses, of assessments. However, additionally, the current state of understanding of assessment among stakeholders should be examined. For instance, Sireci (2012) claims that "there is great debate, and even ignorance, regarding the types of

information tests can provide.” Therefore, in addition to raising standards of validity and validation, the understanding of current standards should be verified.

Recommendations

Among those in the measurement field, it is widely accepted that validity “is the single most important criterion for evaluating achievement testing” (Koretz, 2008, p. 215). Presenting the concept of validity to assessment stakeholders in an accessible manner should be the foremost concern in ensuring that purpose drift does occur or, if it is inevitable, does not negatively affect the students partaking in educational assessments.

Linn, Robert (2012). *Test-Based Accountability*.

Findings

This paper provides a historical perspective of accountability in education in the United States, beginning with the ESEA of the 1960s and ending with the value-added (VA) movement that has seemingly continued to grow despite many shortcomings, which are highlighted by Linn. The paper goes on to describe the ranging effects of accountability systems, details on the adoption of CCSS and PARCC, as well as future directions of accountability, drawing upon both federal and international examples. Linn concludes by proposing that combining qualitative and quantitative approaches in a complimentary manner would be the most effective approach to realizing the goals of assessment-based accountability.

Commentary

Of particular interest is the discussion and recommendation of supplementing qualitative information to test-based accountability systems. Not only could this potentially provide contextual information to educators regarding the basis for quantitative findings, but it also would offer a degree of convergent evidence to the validity of test-based accountability. Often, a large amount of trust is given to quantitative findings, even when the processes underlying the analyses are not fully understood. Providing (qualitative) information that correlates with the quantitative results would strengthen the findings of the test data. Furthermore, it could act as a safeguard to false-positive or false-negative findings.

Recommendations

As discussed in the preceding paragraph, including qualitative information about school and teacher effectiveness will serve to improve accountability systems because of the contextual information that it provides. This idea should be fervently advocated in school systems that already use test-based accountability systems.

This paper also can serve to encourage the discussion of the link between research and policy. For instance, Linn points out many shortcomings of using value added models (VAM) for measuring teacher effectiveness. However, these methods have been adopted by many states in recent years and even used to publish teacher scores online by the *Los Angeles Times* and *New York Times*. While this paper provides an excellent tour of how we have arrived at the current state of accountability, it also provides commentary on many methods of accountability, which seem to be at odds with some policy practices. The gaps between educational research and policy should be highlighted, and this paper serves as one method of doing so.

Mendoza-Denton, Rodolfo (2012). *A Social Psychological Perspective on the Achievement Gap in Standardized Test Performance between White and Minority Students: Implications for Assessment*.

Findings

Mendoza-Denton discusses the role that societal bias plays in differences in achievement test performance. Though the author acknowledges the research that has examined test bias in assessment, other factors including stereotype threat/social identity threat and status-based rejection sensitivity are explored with respect to student performance on educational assessments. The author points to societal messages that may influence student self-protective strategies and subsequent engagement in academics and performance on tests. The author ends the article with an eye toward possible changes in assessment practices, focusing on the role that test developers can play in closing the gap.

Commentary

The paper points to the importance of assessments within the context of other educational factors that play a role in student achievement. The author admits that the findings of the paper have more to do with the environment of examinees aside from the assessment; however, he provides clear suggestions for the purpose of closing the test gap. The implications that the author suggests are clear and very relevant to current assessment practices.

Recommendations

One aspect that is especially significant about the suggestions from the author is that many of them are directed at testing organizations. That is, the author provides recommendations that test developers can follow in an effort to close the achievement gap. These recommendations are important partly because they do not allow test developers to get “off the hook,” even though

the paper advocates for a shift away from viewing bias as within the test. The recommendations appear to be practical and straightforward, and should be promoted by the Gordon Commission.

I still believe that test bias is an aspect of assessment that is important to investigate now and in the future. However, the recommendations in the paper approach the achievement gap from a different direction than is traditionally considered in measurement, and are certainly worthwhile to explore.

Mislevy, Robert (2012). *Four Metaphors We Need to Understand Assessment*.

Findings

In his paper, Mislevy proposes that the discussion of how to improve assessment can be problematic because of different stakeholder values, and the language that each stakeholder is used to using may not be effective in communicating their ideas to others. Given this problem, he proposes four metaphors which can be helpful in thinking about assessments and policy: 1) Assessment as Practice, 2) Assessment as Feedback Loop, 3) Assessment as Evidentiary Argument, and 4) Assessment as Measurement. Additionally, Mislevy highlights four other additional metaphors that provide a “sharper focus” for understanding assessment: 1) Tests as Contests, 2) Assessment Design as Engineering, 3) Examination as the Exercise of Power, and 4) Assessment as Inquiry.

Commentary

This paper is incredibly useful to individuals in and out of the educational measurement community. It certainly can be agreed upon that stakeholders use different language to describe assessment depending on their interests. However, they have diverse areas of expertise, agendas, and goals that extend above and beyond different words for assessment; as such, the meanings and purposes of assessment vary as well. Furthermore, the metaphors described in the paper are multidimensional in that they present the perspectives of not only different stakeholders, but also those at varying stages in the assessment *process*, from development to decision making based on test results.

Throughout the paper, Mislevy incorporates and reincorporates examples that help to bring the metaphors into focus, and also provides implications for understanding assessment from the perspective of each metaphor. Overall, this paper is an excellent step forward in

understanding assessment now and in the future, especially given that the number of assessment stakeholders is not expected to reduce in size, but may grow to include more invested groups.

Recommendations

This paper frames the discussion of assessment by pointing out that assessment is viewed through multiple lenses. The paper can be used to point out how various groups hold a stake in educational assessments, but that those groups must be able to communicate with one another if improvement in certain areas is to occur (and, as pointed out by Mislevy, improvement in one area does not necessarily mean improvement across all areas).

Additionally, though not the heart of the article, the paper also introduces two concepts which are important in the current testing climate: Tests as Contests and Examination as the Exercise of Power. Given the importance that assessments hold in today's educational system, along with the varying uses of tests and test results, these metaphors are especially relevant to the discussion of testing. One future direction would be to expand on these specific metaphors and include the perspectives of and consequences/effects on different stakeholders.

Thurlow, Martha (2012). *Accommodation for Challenge, Diversity, and Variance in Human Characteristics*.

Findings

This paper describes the use of accommodations for various types of students, as well as a history of test accommodations over time. A comprehensive discussion of validity is included, and particular attention is paid to the type of validity evidence that has been collected for accommodations. The author also discusses the idea of “accessible assessment,” which leads to the discussion on how computer-based tests will likely lead the way in testing accommodations in the next century. This section also includes a description of “universal design.”

The paper describes students who have traditionally benefited from accommodations, and introduces the idea that students who have not officially been identified with a disability may benefit from accommodations. Finally, Thurlow concludes with some ideas for future directions, such as whether or not group characteristics should define accommodations rather than individual student needs, as well as the differences that students may face in accommodations at school and in post-secondary and work situations.

Commentary

After just finishing up a classroom assessment course for pre-service teachers, this paper reminds me of a worry that came up multiple times in class: are accommodations actually changing what the test is measuring rather than just eliminating construct-relevant variance? Also, how does one decide who should receive an assessment accommodation? This paper addresses those concerns, and therefore can be very useful not only to those in educational measurement, but also to teachers and administrators. With that in mind, this paper is written from a large-scale assessment perspective, but some discussion involving accommodations at the classroom level would also be useful. However, overall, the paper is very helpful in clarifying

misconceptions and vagueness surrounding assessment accommodations, and in pointing out the future directions that can be investigated by educational researchers.

Recommendations

In Thurlow's conclusion, she suggests that future research should examine the effects that accommodations have on students with and without the label of a disability. That is, she states that "students with and without disabilities do benefit from certain accommodations, just as students in both groups do not benefit from the same accommodations" (p.28). In line with that reasoning, within-group heterogeneity should be a factor when examining any group's performance on an assessment. That is, understanding and accounting for factors that may affect student performance beyond a disability, race, language, or gender may be one step in ensuring that assessments are able to capture what students know and are able to do. While not discounting past findings about group differences, innovative research should explore these peripheral factors more closely in the future.

Themes and Reflection on the Future of Assessment

Themes

The papers that were written for this project were comprehensive, insightful, and thought-provoking, each in different ways. Each paper brought forth distinct ideas and varying perspectives; however, there were a series of themes that appeared to overlap across many of the papers. Furthermore, these themes are not mutually exclusive. That is, where one theme comes into play, the others are intertwined. Some of these themes were explicit in the papers, while others were more subtle. For instance, the concept of validity was clearly a topic that was targeted by some authors. Recall that validity is defined as the “degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests” (APA, AERA, NCME, 1999, p. 9). Ho’s (2012) paper discussed varieties and drift of purposes of assessment, and Thurlow (2012) outlined the importance of maintaining test validity when making assessment accommodations. However, anytime a decision is meant to be made about a student based on test results, validity comes into play because of its direct relationship to the meaning of test scores (Messick, 1995). Whether an assessment is capturing novel skills that are meant to represent 21st century competencies (Bereiter and Scardamalia, 2012; Gorin, 2012), showing an achievement gap between two subgroups (Mendoza-Denton, 2012), or being used as an evidentiary argument (Gorin, 2012; Mislevy, 2012), validity is the central concern.

Another theme that was touched on in many of the papers, directly and indirectly, was accountability. Linn (2012) explicitly discussed past and present test-based accountability systems, while other papers referred to the responsibility that schools and teachers have for ensuring that learning is taking place (e.g., Armour-Thomas and Gordon, 2012; Ho, 2012; Varenne, 2012).

Finally, another major theme that emerged from many papers is the technological aspect of assessing students. For instance, Behrens and DiCerbo (2012) wrote thoroughly about the technological implications of assessment, Chung (2012) wrote about how technology can be used to inform at different levels of education, and Hill and Gates (2012) discussed how technology can be used for alternative methods of assessment and organizing student data. Dixon-Roman and Gergen (2012) and Bereiter and Scardamalia (2012) touched on the technological changes in society that will influence the future of assessment, Armour-Thomas and Gordon (2012) discussed how computers can be used to develop learning-centered assessments, and Thurlow (2012) suggested that computer-based assessments will allow for accommodations to be embedded in tests rather than using human intervention.

It also is essential to bear in mind that these three themes are considerably interrelated with one another. For instance, accountability and validity can be associated with one another if tests used for accountability may not have originally been meant for that purpose. If assessments are to be used for accountability, then validation for that purpose needs to occur. If one considers consequences to be a part of validity, then the relationship is all the more important. Technology and validity also are related to one another for similar reasons. As technology changes the way in which students may be assessed, validation practices may need to reflect those changes. Technology and accountability also can be related in that each may be used to inform and improve one another.

Missing Themes

In addition to the themes that were written about by most of the authors, there appeared to be some important ideas that were not explicitly covered, but did make brief appearances in some papers. These missing themes are important, nonetheless, in understanding the role that

assessment plays in the 21st century. For instance, the concept of transparency is an important factor that should inform assessments now and in the future. In many ways, measurement and testing are mysterious to those who are not involved in the development process. As discussed earlier, Dixon-Roman and Gergen (2012) point out that there is an assumption that educational assessments measure what they are meant to measure. Often, those assumptions are based on little more than blind faith. There are many examples of ways in which testing organizations and policy makers can increase transparency in testing. Take, for instance, how value-added (VA) is used in some school districts. The way in which student test scores are used to create a VA score for teachers is largely unknown to teachers, administrators, and policy makers. Informing these stakeholders, at least at a basic level, of how the scores are computed will result in better educated decisions in how to use them. Another idea of how transparency can be exercised in educational assessment is through making test items public prior to large-scale administration (E. Baker, personal communication, June 19, 2012). The results of this action could be the demystification of testing to students, parents, and teachers, in addition to discouraging cheating. As of yet, this has not been done in any large-scale testing situation. Commencing a dialogue about possible methods of transparency is an important step in improving the public's relationship with educational assessment.

Another theme that is missing from the papers is the use of student response processes and cognitive models to inform assessment design. Chung (2012) discusses using technology to understand student processes and Gorin (2012) points out various methods of collecting data on response processes. However, this theme was not addressed as comprehensively as it could be. The movement to have students think critically in school is growing. However, assessing critical thinking is a challenge (Cromwell, 1992), and in some cases, assessments intended to measure

critical thinking may not actually require students to do so (Leighton and Gierl, 2011). Despite this, Leighton and Gierl (2011) argue that cognitive models, which may be useful for ensuring that higher order thinking is used, are rarely employed in assessment design. Consequently, integrating cognitive models in assessment development may be one avenue to explore in the discussion of the future of assessment.

Future Directions

The various themes described above introduce some possible future directions for assessment in the 21st century. First and foremost, assessment specialists need to educate test users on the concept of validity and why it is so important in assessment. For example, consider the recent interest in quantifying teacher performance, as evidenced through both the *Los Angeles Times* and the *New York Times* publishing teacher VA scores online. While the actions of these two newspapers are questionable, it provides an example of how a couple of themes from the papers can be applied to the future of assessment. The VA scores were used in a way that many measurement scholars often warn against. For instance, they were used for high stakes purposes and little data was used to determine teacher scores. That is, the purposes of the assessments did not match this particular use. In addition to validity, transparency is a theme that can be applied to this situation. While some may argue that transparency was being demonstrated by publishing teacher names along with their effectiveness score, it likely would be difficult to find a teacher or policy maker who is able to explain how the VA models work. These two themes are important concepts to communicate with the public and policy makers. Using examples such as the one described above may be a useful opportunity for education and measurement researchers to spread their message.

Another future direction is to find ways to encourage the already growing movement of combining technology and assessment practices. For instance, policy makers could provide incentives for test designers to incorporate technology with their assessments, and testing companies could offer internships for video game developers. Additionally, graduate level measurement programs could offer courses that integrate technological advances and encourage students to consider ways in which to apply their own technological knowledge to their area of study.

Finally, another possible direction is to investigate the link between student response processes and assessments that are meant to measure student knowledge and abilities. Whether this is done through the use of technology, as Chung (2012) described in his paper, or through other methods, providing evidence for students' response processes should be encouraged, especially when complex thinking is required of students. Methods used to encourage including cognitive processes may vary, although developing standards within the field of measurement that require them for assessment design is one option. Alternatively, providing incentives for students and researchers to further study, use, or develop cognitive process models for assessment may be another method.

As discussed above, the papers that were prepared for the Gordon Commission each contain diverse and motivating ideas for the future of assessment. Though these ideas were put forth by leaders in the field of education, Sireci (2012) points out that "it is important to realize that the use of tests is initiated and managed by policy makers" (p.27). As such, the overall next step of this project should be to work toward the effective communication of these ideas and themes to those who ultimately decide what assessment will look like in the 21st century.

References

- AERA, APA & NCME (1999). *Standards for educational and psychological testing*. Washington, DC: Authors.
- Armour-Thomas, E. & Gordon, E. W. (2012). *Toward an understanding of assessment as a dynamic pedagogy*. Unpublished manuscript.
- Behrens, J.T. and DiCerbo, K.E. (2012). *Technological implications for assessment ecosystems: Opportunities for digital technology to advance assessment*. Unpublished manuscript.
- Bereiter, C. & Scardamalia, M. (2012). *What will it mean to be an educated person in mid 21st century?* Unpublished manuscript.
- Chung, G. (2012). *Toward the relational management of educational measurement data*. Unpublished manuscript.
- Cromwell, L. S. (1992). *Assessing Critical Thinking*. *New Directions for Community Colleges*, 77, 37-50.
- Dixon-Roman, E. J. & Gergen, K. J. (2012). *Epistemology and Measurement: Paradigms and Practices I*. Unpublished manuscript.
- Ercikan, E. (2006). *Developments in assessment of student learning*. In P.A. Alexander and P. H. Winne (Eds.), *American Psychological Association, Division 15, Handbook of educational psychology, 2nd edition* (pgs.929-953). Lawrence Erlbaum Associates.
- Dixon-Roman, E. J. & Gergen, K. J. (2012). *Epistemology and Measurement: Paradigms and Practices II*. Unpublished manuscript.
- Gorin, J. (2012). *Assessment as evidential reasoning*. Unpublished manuscript.
- Ho, A. (2012). *Variety and drift in the functions and purposes of assessment in education*. Unpublished manuscript.
- Koretz, D. (2008). *Measuring Up*. Cambridge, MA: Harvard University Press.
- Leighton, J.P. & Gierl (2011). *The Learning Sciences in Educational Assessment: The Role of Cognitive Models*. New York, NY: Cambridge University Press.
- Leighton, J.P. (In press). *A cognitive model for the assessment of higher order thinking in students*. In G. Schraw (Ed.), *Current perspectives on cognition, learning, and instruction* (Book series): *Assessments of higher order thinking skills*. Information Age Publishing.
- Linn, R. L. (2012). *Test-based accountability*. Unpublished manuscript.

- Mendoza-Denton, R. (2012). *A social psychological perspective on the achievement gap in standardized test performance between White and minority students: Implications of assessment*. Unpublished manuscript.
- Messick, S. (1995). *Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning*. *American Psychologist*, 50(9), 741-749.
- Mislevy, R. J. (2012). *Four metaphors we need to understand assessment*. Unpublished manuscript.
- Mislevy, R. J. (2006). *Cognitive psychology and educational assessment*. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed.) Westport, CT: American Council on Education and Praeger Publishers.
- National Research Council. (2001). *Knowing What Students Know*. Washington, DC: National Academy Press.
- Sireci, S.G. (2012). *Informing in the information age: How to communicate measurement concepts to education policy makers*. *Educational Measurement: Issues and Practice*, 31(2), 27-32.
- Thurlow, M. L. (2012). *Accommodation for change, diversity, and variance in human characteristics*. Unpublished manuscript.
- Varenne, H. (2012). *Education: Constraints and possibilities in imagining new ways to assess rights, duties and privileges*. Unpublished manuscript.
- Wilson, M. & Sloane, K. (2000). *From principles to practice: An embedded assessment system*. *Applied Measurement in Education*, 13, 181-208.