



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

Gordon Commission Fellow Synthesis Paper

Keena Arbuthnot

Louisiana State University

The Gordon Commission Synthesis paper is intended to provide an overview of the recommendations and findings put forth by the Gordon Commission authors. In choosing the articles for this task, it was my hope to be able to abstract information from the articles that spoke to my research interest and area of expertise. Ironically, the Gordon Commission papers have pushed me to think more critically about the future of assessment and the various aspects of assessment and assessment reform that should be considered. The papers that were reviewed varied in scope. The first section of the synthesis paper presents an overview of the findings and recommendations from each of the reviewed papers. The second part of this synthesis paper focuses on distilling the information from the reviewed papers and synthesizing the findings and recommendations from the entire body of research. Although the papers highlighted information about the history of assessment and the current prevailing issues and concerns in the field of assessment, the most important aspects of all the papers are the comments and proposals regarding the future of assessment. Presented below is a brief abstract and review of a select group of Gordon Commission papers.

Review of Select Gordon Commission Papers

The selected papers that will be discussed were chosen based on their substantive areas of interest. Each of the papers provides a different perspective and unique angle on issues and problems in the field of assessment. It is my hope not only to provide a brief synopsis of the paper, but also to highlight the findings and recommendations that I deemed were most powerful and informative to the future of assessment.

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

Linn (2012) provided a great overview concerning the history of accountability systems. The paper outlined several issues related to accountability, including why there has been such a

demand for accountability over the years. Who is inherently accountable in the area of education? In addition, the paper addresses the positive and negative effects of accountability systems. Most notably, the author provides commentary on the lessons learned from past experiences with test-based accountability systems including, 1) the same test form cannot be used year to year, 2) accountability systems need to include mechanisms to evaluate score inflation and guard against it, 3) low-stakes tests (i.e., NAEP) should be used to monitor progress on high-stakes tests to avoid test inflation, and 4) tests that include self-monitoring systems should be an effective approach in the future.

The author offers several recommendations and approaches to accountability that should be considered. First, the paper emphasizes the importance of usefulness of using both quantitative and qualitative approaches to accountability. The author provides examples of how these methods have been successfully used in places such as Great Britain and Finland. Linn (2012) states:

“A modification of traditional test-based accountability could use the student growth on achievement tests as a trigger to identify schools where inspections would be conducted to collect qualitative information that might explain low rates of student growth and suggest possibilities for improvement....”

Second, the paper highlights why accountability systems should focus on maximizing intended positive effects and minimize unintended negative effects. The author outlines some common pitfalls that should be avoided in future accountability systems. For example, he addresses the relative importance of making necessary accommodations for disabled students and addressing language issues related to English language learners. Third, Linn (2012) emphasizes that accountability systems should provide instructionally useful data to teachers and educational practitioners.

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

Ho (2012) begins the paper by stating the seven purposes of educational assessment as described by Haertel (2012). Haertel (2012) divides the purposes of assessment into two major categories that include *measuring* and *influencing* purposes. The author states there are four measuring purposes including instructional guidance, student placement and selection, informing comparisons among educational approaches, and educational management. Similarly, there were three influencing purposes that include directing student effort, focusing the system, and shaping public perceptions. Ho (2012) critiques Haertel's interpretation of these multiple purposes of educational assessment. The author emphasizes that Haertel's work should be expanded to produce a framework that spans beyond large scale assessment that would include classroom and formative assessment as well.

The second goal of the paper was to address the issue of "purpose drift," when an assessment is designed for a specific purpose and is ultimately being utilized for another purpose altogether. Ho (2012) states:

"Much of the struggle with purpose is captured by the rhetorical difficulty of communication that validity is not a property of an assessment but a use or interpretation."

The author highlights the fact that purpose drift presents validity issues in reference to the interpretation of test scores. Thus, Ho (2012) states, "If known forces cause the purposes of an assessment program to deviate from the purposes originally validated, then conventional validation approaches proposed in assessment literature are inadequate." The author highlights the need to focus on issues related to purpose drift and propose new and innovative ways to address validity issues as purposes shift or change.

Mendoza-Denton, Rodolfo (2012). *A social psychological perspective on the achievement gap in standardized test performance between White and Minority students.*

Medoza-Denton (2012) is intended to explore the social psychological issues that help to explain the achievement gap between minorities and White students. The article delves into how protection of self-esteem, conceptions of intelligence, social identity, and discriminatory experiences can all have an effect on test performance and consequently, the achievement gap. Based on the research findings in this area, the author proposes interventions with respect to schooling and the testing industry. First, in terms of schooling, the author posits that schools should promote environments that encourage students to understand that intelligence is malleable, not fixed. Research has shown that the view of intelligence as malleable allows students to discover and believe in the expansion of their intelligence, rather than feel that it is fixed and they have limited ability for growth. Next, the school environment should be one that values, accepts, and respects differing personal and social identities.

In addition, the author makes several recommendations for changes in the test development process and in structural changes and interventions that the testing industry can promote. In terms of test development, the author suggests that the current language of testing should be altered. Research has shown that language such as *cognitive*, *ability*, *aptitude* have been shown to activate stereotype threat. Next, Mendoza-Denton (2012) recommends that the test industry move away from tests that show group differences. Lastly, the paper proposes that the testing industry should advocate for the end of tracking and meaningfully diversify the industry at all levels.

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

Mislevy (2012) provides a framework that uses four overarching metaphors to help better understand assessments. The first is *assessment as practice*; this metaphor comes from a socio-cultural perspective in psychology. The author states, “Practices are recurring, organized activities that people become attuned to, learn their constraints and affordances, and use to interact with other people and situations.” The second is *assessment as feedback loop*. This metaphor focuses on the multiple uses of an assessment and the length of time the information can be given or shared with interested parties (i.e., teacher, administration, etc.). Third, *assessment as evidentiary argument* refers to providing evidence about what a student knows based on what he or she says, does, or makes in a limited set of situations or contexts. Lastly, as *assessment as measurement*, Mislevy states, “Educational measurement is better understood not as literally measuring existing traits, but as providing a framework to reason about patterns of information in context.” The author then provides an additional four metaphors that would be a more precise, or intricate, way to understand assessments. The following four additional metaphors were discussed: *test as contests*, *assessment design as engineering*, *examination as the exercise of power*, and *assessment as inquiry*. Mislevy (2012) proposes using this framework as a way to discuss and think about assessment issues in the future. The author states, “These metaphors make us aware of the conceptual frameworks we can take advantage of, and hook us into the experiences, the tools, and the wisdom of many disciplines. The metaphors do not resolve questions about assessment, but they do help us ask them sensibly.” Mislevy (2012) concludes that the use of the metaphors does not provide resolution to the problems related to assessment; however, it does provide a logical way to organize and conceptual the issues.

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

Gorin (2012) argues that the current state of assessment provides only a single source of evidence. Based on Mislevy's (2006) notion of assessment as evidentiary argument, Gorin (2012) calls for a much more comprehensive or expansive view of educational assessment, whereas multiple evidential sources are provided. Ultimately, this expanded view of educational assessment can provide additional information that will allow for a more refined assessment system that will be more reliable and valid in the manner in which it can make claims and decisions regarding student learning and instruction. The author provides examples of how multiple sources of evidence are used in other forms of assessment outside of education, including psycho-educational assessment, information technology management systems assessment, and medical/diagnosis assessment. In terms of the future of educational assessment, Gorin (2012) states four things that are needed from educational assessment to support evidentiary argument: 1) a clear definition of all possible states of knowledge; 2) a list of behavioral evidence that would illustrate a particular state of knowledge; 3) a set of data collection procedures that produce the relevant behavioral evidence; and 4) scoring rules for using behavioral evidence to determining individuals' state of knowledge from among the set of possible states.

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

Hill (2012) begins his paper with a brief overview of the origins of testing and the developments in testing over the years in Europe and the United States. The author proposes that the use of digital technologies can be instrumental in enhancing our current assessment system and can be used in helping us understand student performance more clearly. To illustrate this Hill (2012) provides a detailed description of the *Digital Testing Model/Digital Project Model* that he

and his colleagues implemented to a select group of students in the United States and in China. These models incorporate a process log that can assess how students spend their time while they work on tasks. Hill (2012) outlines some of the challenges facing this type of technology; however, the author is confident that these challenges can be addressed in the future and that new technologies will provide a better alternative to current assessments.

Kaestle, Carl (2012). *Testing Policy in the United States: A Historical Perspective*.

Kaestle (2012) emphasizes the importance of understanding the history of educational assessment in the United States. The author outlines all of the major developments in the field in the last two centuries. Kaestle (2012) states:

“I’m a historian, and my historical sense is that the assessment practices we have inherited, and which are deeply embedded in our educational system, are out of date. One of the uses of history can be to raise the issue of whether we should be thinking about how to apply our knowledge to the future, not just our traditions.”

The author brings up the point that possibly tradition has been an impediment to the development of a new assessment system. Kaestle (2012) argues that a new system of assessment should clearly define why the proposed system would be better than the current system, and be able to articulate how the new system can meet accountability goals. The author also stresses that a new system of assessment should be accessible not only to measurement professionals, but also to a lay person, and educational practitioners should be able to clearly understand the system as well. The author believes that this is a pivotal time in history to promote reform in the area of educational assessment. Although the author has described in some detail the durability of the current system, Kaestle (2012) highlights that in the last 12 years, there have been signs that change and reform will be more plausible, and he feels that the time for reform is at hand given the current advances in the area of assessment (i.e., technology).

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

Chung (2012) begins the paper by discussing how technological advances can have a significant effect on individualized instruction. Chung states, “A key issue is how to leverage these data to measure what students understand and can do, to derive meaningful measures of cognitive and affective processes, and to develop capabilities for precise diagnosis and targeting of instruction.” The author differentiates between outcome measures, which show if a student can finish a task or assignment, and process measures that show what a student does or tends to while completing a task. The process measures identify how a student completed a particular task. Process measures can provide more relevant diagnostic information in comparison to outcome measures. The author gives examples of how game play can be used to better understand the ways in which students process tasks. The author argues that adaptive systems can be instrumental in supporting teaching and learning. Lastly, Chung (2012) states that having transactional data that describes how a student processes tasks can be a key to transforming and improving our current assessment system.

Hakuta, Kenji (2012). *Assessment of Content and Language on the Heels of the New Standards: Challenges and Opportunities for English Language Learners*.

Hakuta (2012) begins by pointing out the current state of education for English language learners. Hakuta highlights three noteworthy areas that need to be addressed. First, the author highlights the difficulty with addressing the needs of long-term English language learners. Second, the author addresses the role that one’s native language should play in instruction. Lastly, the author highlights the potential for and advantages of incorporating bilingual education within the school systems. The author moves past the current state of ELL education and makes predictions about ELL education in the next five years. Where have the gains been made? What

has been accomplished? This prophetic dialogue highlights the successes and shortcoming of several assessment initiatives, including the Common Core, that have been implemented.

In terms of recommendations, Hakuta (2012) presents a list of opportunities that will be successful to the future of assessments. First, the author states that after the initial implementation of the Common Core assessments, researchers will be able to understand difficulties and issues related to ELL students more clearly. Second, he calls for implementation of a targeted assessment system for ELLs that ramps up the content-embedded language construct. Third, he proposes the use of dynamic formative assessment to give feedback to teachers and students. Lastly, Hakuta (2012) calls for bilingual assessments through authentic translation tasks. The author states:

“In order to support the development of assessments as envisioned here, a new field of applied scholarship that pays explicit attention to the relationship between language and content will need to be cultivated, in order to attract the best human talent available to maximize the potential. Traditionally, this is the represented in the interdisciplinary field broadly known as language and cognition, encompassing the fields of linguistics, cognitive psychology, sociocultural theory, philosophy, computer science, and the neurosciences.”

Synthesis of Gordon Commission Papers

The goal of the Gordon Commission is to formulate and conceptualize the future of assessment. In order to adequately address the future of assessment, it is first important to chronicle the history of assessment, as well as evaluate the weaknesses and prevailing issues of the current system. The Gordon Commission papers have provided some insight and recommendations of the possibilities and opportunities for the future of assessment. First, many of the papers provided a historical foundation of assessment. The insight from these papers can be used in helping to frame the discussion of the future of assessment. Next, the papers emphasized the current state of assessment. This information is quite helpful in the sense that it

provides a framework for the various components and aspects of our current assessment policies and practices that seem to be problematic. Most importantly, many of the authors were able to extend their work to discuss the future of assessment. Where should the field of assessment be moving? How should we strengthen assessments? What policies should be put in place that would help to move assessment into the future? Presented below is a synthesis of the issues and recommendations set forth by the Gordon Commission authors on the future of educational assessment.

Future in Educational Assessment

As discussed, all the Gordon Commission papers in some respect contributed to the notion of what the future holds for educational assessment. The recommendations and commentary were very diverse in nature and provided insight and reflection on several areas related to educational assessment. In order to synthesize the findings from multiple papers, I found it most practical to organize them based on the extent that the recommendation or proposal related to and/or had an impact on certain stakeholders. This organization is advantageous given that the goal of the Gordon Commission is to provide information and advice on the future of educational assessment to multiple stakeholder groups.

Testing Industry

The Gordon Commission papers make reference to the future role and responsibility the testing industry should play in reference to the future of assessment. The papers as a whole suggested that the testing industry become more responsible for the products that they produce and distribute. Several authors called for test developers to be more proactive in guarding against the unintended uses of tests. For instance, Ho (2012) highlighted the issues surrounding “purpose

drift.” Purpose drift is described as a situation in which test consumers are using tests for purposes other than those for which they were created. Ho (2012) states:

“...if known forces cause the purposes of an assessment program to deviate from the purposes originally validated, then conventional validation approaches proposed in assessment literature are inadequate.”

As previously mentioned, this can be a detriment to the validity of the interpretation of test scores for that particular assessment. It can be argued that the testing industry should be vigilant in regulating and ensuring that consumers of their products are not using them in ways that they were not intended to be used, especially when there are high stakes involved. Next, Mendoza-Denton (2012) raised the issue that the testing industry must be more responsible in their role of addressing issues related to the achievement gap and group differences. The authors contend that the testing industry cannot just stand back and say, “This is not my problem.” They must take on a leadership role not only in examining the issues more closely, but also in working to reduce the gaps between test takers from different groups. Lastly, many authors suggested that the testing industry explore and pilot new technologies that will provide a more comprehensive picture of what students know (Hill, 2012; Chung, 2012; Gorin, 2012). As several papers have shown, there are new technologies available that can be used in enhancing our current assessment delivery and process. In conclusion, the testing industry has a responsibility to play a leadership role in the reformation of assessment. Consequently, this will call for members of the testing industry to take on a much more intense role not only in the development of assessment, but also in regulating the proper use of their assessment products as well.

Policy Makers

When discussing the future of educational assessment, one of the key components of making significant changes in the field is promoting and drafting educational policy that will

encourage changes or modifications based on the suggestions that have been brought forth.

Several Gordon Commission authors discussed issues that would be relevant and useful points that should be considered when drafting future educational policies related to assessment. First, Kaestle (2012) states, that this time in history provides great opportunities for making progress in assessment reform. Given developments in recent years, it seems to be more political willingness to adhere to and potentially embrace a new assessment system. The question is raised, what educational policies should be put in place to reform or improve the current system? The authors reported the following issues should be considered when drafting future assessment policies. A new system of assessment should:

- focus on maximizing intended positive effects and minimize unintended negative effects (Linn, 2012)
- use a mixed methodological approach to accountability (Linn, 2012)
- use multiple measures should be used in lieu of one single assessment (Gorin, 2012; Linn, 2012)
- include provisions for making accommodations for the assessment of special populations of test takers including minority students, English language learners, and students with disabilities (Linn, 2012; Hakuta, 2012)
- use technology to provide a deeper understanding of student learning (Gorin, 2012; Chung, 2012; Hill, 2012)
- provide instructionally useful data to teachers and educational practitioners (Linn, 2012; Ho, 2012; Chung, 2012)

In order to successfully make lasting change in the field of assessment there must be a significant willingness on the part of policy makers to take drastic measures to revitalize and improve our current system.

Academicians and Researchers

The Gordon Commission papers have presented recommendations for further research and ways in which researchers and academicians in the field of assessment should understand assessment, as well as conduct additional research that will further contribute to the field. First,

Mislevy (2012) presents a framework that categorizes assessment using four different metaphors. In the future, using this new framework to dialogue about and understand assessment will promote a deeper and more refined way of discussing and interpreting the varying aspects and definitions of assessment. Similarly, Ho (2012) proposed that understanding and acknowledging the various purposes of assessment are among the first steps in understanding and scrutinizing current assessment policies and practices. Both Ho (2012) and Mislevy (2012) provide detailed information about the ways in which academics and researchers should conceptualize certain aspects of assessment. This common language and understanding is helpful to the field because it sets a common framework for those studying and critiquing the field to adhere to in the future.

Several of the Gordon Commission papers discuss how innovations in other fields of study can be instrumental in providing assessment solutions. For example, Gorin (2012) describes how assessments in the medical field and information technology management solutions can be used as models for the types of solutions that can be useful in educational assessment. In addition, several papers presented detailed information about how digital technologies can be used to revolutionize the field of assessment. Hill (2012) described how digital archives to store student assessment data can be a new, innovative alternative to our present system. Similarly, Chung (2012) describes how technological advances can be used to provide a more detailed account of how test takers process information while completing tasks. In sum, it is imperative to understand that the digital technologies, to some extent, must be part of the agenda for assessment in the future. Although there is no consensus on the precise way in which technology should be used in the future of assessment, there seems to be agreement that technology should be used in some form to help advance the field of assessment. Consequently,

further research needs to be conducted to identify technologies that would be best suited to help advance the field of assessment.

Practitioners

One of the most important consumers and stakeholder groups that educational assessments affect are practitioners. Several of the Gordon Commission papers highlighted the need to ensure that practitioners were considered in the discussion of the future of educational assessment. Many authors addressed issues about the feedback loop, and discussed that in the future information and feedback from educational assessments must be available to practitioners in real time (Ho, 2012; Mislevy, 2012). The current accountability system is criticized for the amount of time it takes to get the results of the assessment data to teachers so they can be effective tools in improving teaching and learning (Linn, 2012; Ho, 2012; Chung, 2012). Many also showed concern that current educational assessments do little to improve teaching and learning. Consequently, future assessment systems should emphasize the extent to which teachers and practitioners can understand and use the data to make necessary changes and modifications in the classroom. Second, some authors discussed how teachers can make changes to the classroom environment that could potentially affect the assessment outcomes of particular groups. Mendoza-Denton (2012) proposed that practitioners should create a school environment that values and respects differing personal and social identities. In addition, the environment should encourage the notion that intelligence is malleable, not fixed. It seems that there is consensus that a major priority of future assessment systems should be to address the needs of practitioners. Consequently, any new system should ensure that the results of assessments can be used by practitioners to improve teaching and student learning.

Conclusion

In sum, the group of Gordon Commission papers provided a very diverse and in-depth perspective about various aspects of assessment. As specified by the Gordon Commission, their goals are to present recommendations and insight into the future of assessment. What should our assessment system provide in the future? What should it look like in the future? Based on what the papers have presented, as well as my insight into the history and field of educational assessment, it seems timely and most beneficial that we continue to strive to make adjustments and positive changes to our current assessment system. It is in the best interest of all relevant parties to develop an advanced system that will be able to help those in the educational system better serve students and provide necessary information to stakeholders that provides an accurate and comprehensive evaluation of student learning and progress.