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

Gordon Commission
Fellow Synthesis Paper

Catherine Kramarczuk Voulgarides

New York University

Hakuta, Kenji (2012). *English Language Learners and Non-Standard Language Speakers.*

Findings

Professor Hakuta's paper is a practical, yet forward thinking paper about the role of assessment for English Language Learners (ELLs). He writes about a futuristic situation in the year 2017 when the Common Core Standards have been implemented and lessons can be gleaned. He focuses on describing how the Common Core have affected English Language Learners, how they can be improved, and how more research is needed to understand assessment of ELLs.

In setting up his futuristic scenario, Hakuta brings attention to the fact that the Common Core Standards rely upon and employ discipline-specific language. This is problematic because the standards encourage interdisciplinary construction of arguments without providing linguistic scaffolding for ELLs. Hakuta highlights that this reality needs to be considered by PARC and SBAC as they develop assessments. He also proposes that these assessments be formative, continuous, and interactive. He argues that this is necessary because it will allow for language to be understood as something that exists within a specific domain or context that is embedded within larger systems of meaningful discourse.

Hakuta hopes that in 2017 assessment for ELLs not only will capture academic language knowledge acquisition, but also basic language acquisition. Thus, he foresees moving away from one-time, arbitrary cut-off scores of language mastery, a practice embedded within NCLB, to ongoing formative assessment.

Recommendations

Hakuta recommends that the future of assessment consider creating an interdisciplinary and applied field that addresses the complex needs of ELLs. He imagines the field being called "epistemological linguistics" that is composed of several disciplines: linguistics; cognitive psychology; socio-cultural theory; philosophy; computer science; and neuroscience. He foresees the goals of the field as threefold: 1) to be able to describe language in the context of epistemological domains; 2) to be able to understand the role of language in learning, individual knowledge representation, and the social life of knowledge in classrooms and beyond; and 3) to be able to explain the causal connections that may exist between enriched language representations and enhanced cognitive, socio-cognitive, and socio-cultural exchanges of knowledge domains.

Commentary

I appreciate Hakuta's piece and think it is relevant for the future of assessment because it proposes practical and much needed solutions for current policy. I think the changes he would like to see for 2017 must be clearly heard by and shared with PARC and SBAC. However, I fear that the current policy climate might not be responsive or knowledgeable enough about research on ELLs to effectively apply what he proposes. Therefore it seems important that the Gordon Commission highlight Hakuta's recommendations in public reports.

Varenne, Hervé (2012). *Education: Constraints and Possibilities in Imagining: New Ways to Assess Rights, Duties and Privileges.*

Findings

Hervé Varenne begins his essay by highlighting how schooling in America has largely been understood as a state-granted and legislated individual right. This right allows for individuals to gain privilege and rewards through receipt of education. He argues that because schooling is understood in this way, those who succeed are deemed meritorious of their success. However, this is problematic because there is a persistently strong tie between family background and educational attainment and achievement that puts notions of meritocracy into question. Thus, reform strategies centered on the language of "success" and "failure," such as NCLB, will rarely improve schools because no matter how well intentioned, the policy always will be subject to a process of human interaction that gives privilege to some over others.

According to Varenne, in order to achieve more egalitarian schooling, we must move away from Horace Mann's notion of the "educated" and "non-educated." Rather, we must embrace the ubiquity of education in all facets of life (p. 13) and think systematically about schools and reform. This will remove privilege from the educative process and make it more inclusive. To do this, Varenne argues that the life of schooling should be understood as an ongoing, interactive, and changing process. This perspective implies that assessment also must be conceptualized as ongoing.

Recommendations

Varenne argues that there is a political need to "re-imagine assessment" (p. 23). The future of assessment must not be time bound, static, bound in the language of success or failure, or exclusionary. He would like to see assessment move from being something that is rewarded in the market to something that betters the human experience.

Commentary

I think this paper is very important for the future of assessment because it highlights human interactions within a social context. Assessment does not happen in a black box. It happens within a society, a community, a school, a classroom, a place where many competing interests are present. If the future of assessment does not take into consideration human interaction, then it will miss the opportunity to make assessment more productive rather than reproductive of privilege, success, and failure.

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

Findings

In this paper Cauce and Gordon try to develop a useful "multidimensional view of human agency" (p. 2) that has relevancy for assessment. The definition that they develop relies heavily upon the economist Sen (1985, 1987, 1999 a, b). Sen argues that social structure and agency affect an individual's ability to act upon the world and to freely make choices to act upon that world (p. 5). In conjunction with this definition Cauce and Gordon argue that agency is also the ability to choose between alternatives and to be able to fully act on those with personal values intact.

Cauce and Gordon apply their notion of human agency to assessment by drawing upon the work of Bandura (2006). They agree that his four properties of self efficacy: intentionality; forethought; self reactiveness; and self reflectiveness, (p. 11) are closely related to agency. But, they caution that these characteristics should be considered in conjunction with social structure. This is because they interact with multiple social identities and contexts that affect agency and hence achievement.

The authors end the article suggesting that developing human agency become an end in itself for education and assessment. They believe that agency should be more directly developed in the educative process because it allows for personal empowerment, increased creative capacity, and social progress.

Recommendations

For the future of assessment Cauce and Gordon recommend that the measurement community find ways to incorporate human agency and the "propensity to display it" (p. 14) into assessments. They make it clear that the future of assessment should consider agency as a product of education and not just a process.

Commentary

I think that this paper is extremely provocative for the future of assessment. By proposing that developing human agency become central to education, the authors seem to be realigning the focus of assessment. I think that this concept could be further developed if it was grounded within classroom practices and then expanded outward.

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

Findings

Bereiter and Scardamalia outline what they believe are important skills for humans to possess in the 21st century. They do this by first critiquing how the digital age and social media have influenced students. They argue that the "fleeting" nature of "sound bites" of information that are available on the Internet have created superficial learners. Although they recognize that new media has allowed for grand narratives of knowledge to be questioned, they worry that students are not being asked to sufficiently synthesize knowledge. The authors think students must be pushed to develop theories and ideas with new media. However, they worry that presently, students are not critical enough nor are they placing readily available information within larger political, historical, economic, or social contexts.

With these concerns in mind, the authors explore the implications of "dual process theory," (Stanovich and West, 2000) which understands learning as occurring within two systems. System 1 is a reactionary system that immediately triggers emotions when information is received. System 2 is more sequential and discerning. The authors think that new media initiates huge System 1 reactions that lead to shallow System 2 justifications (p. 9). In order to push back against this type of learning the authors propose several "core competencies" they see as necessary 21st century skills. They are as follows:

1. Students should be able to build, amend, and create knowledge.
2. Students should be able to work with abstractions and convert them to real world applications.
3. Students should be able to engage with systems thinking that recognizes the complexity of the world.
4. Students should be able to engage in collective work that is collaborative.

Recommendations

Based on the core competencies the authors recommend the following for the future of assessment:

- develop assessments that foster creativity; i.e., a portfolio assessments
- allow students to explore information in ways that promote agency and creativity
- encourage systems level thinking that calls for students to discern usefulness of knowledge and to be able to place knowledge within context

Commentary

I think this article is an important forward-looking piece for the future of assessment because it provides a conceptual base from which to build concrete recommendations. However, I would appreciate more conversation about, or development of, the notion of dual process theory in conjunction with a more extensive discussion on “bounded rationality” (Simon, 1957). These seem to be important concepts to bring together because they engage with the complexities associated with processing vast amounts information in possibly complementary ways.

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

Findings

Greg Chung argues that increasing technological information associated with the digital age presents a unique opportunity to develop more individualized instruction and learning. His paper assumes that there will only be greater interest in digital innovation and individualized learning. Based on this assumption he argues that the rise in availability of technical data about student learning, at the fine-grained and transactional level, should be used in conjunction with existing data to advance the field of assessment and measurement.

Chung distinguishes between different types and levels of data that have traditionally been used to understand student learning. First he makes the distinction between coarse-grained and fine-grained data. Coarse-grained data tends to capture aggregate trends associated with student learning. Fine-grained data involves the more nuanced aspects of learning that occur within classroom assessments and interactions. Second, Chung argues that there are three levels of data associated with assessment: systems, individual, and transactional. Demographic data is often captured at the system level. Individual level data, which Chung argues is most often used by the measurement community, reports on things like student mastery of learning targets. Transactional data captures a student's interaction with and processing of a learning task or experience. It is the availability of transactional data, Chung argues, that is growing with the digital age but is not being used enough in assessment.

With these data distinctions in mind, Chung asks whether or not fine and coarse-grained data can be combined to provide a more comprehensive assessment of student learning. He is intrigued by this possibility and offers a couple examples of programs that have tried to do this. One is the learning analytics approach (van Barneveld, Arnold and Campbell, 2012, p. 8), which combines macro and micro data to predict academic risk of students in university. The second is relational data management, which uses data gathered for one purpose for a different one.

Chung concludes that the future of assessment should focus on conducting research that facilitates meaningful use and integration of fine-grained data to existing data sources.

Recommendations

Chung recommends looking to the following for the future of assessment:

- Signals Project at Purdue (Campbell and Oblinger, 2007)
- game play behavior and the use of cluster analysis (Antonenko, Toy and Niederhauser, 2012)
- Quality School Portfolio QSP (Baker, 2000)
- Transmedia apps→ USDOE Ready to Learn Initiative

Commentary

I think it is very important for the future of assessment to seriously consider how fine and coarse-grained data can be melded. I believe that if transactional and individual level data were matched with coarse grained, the assessment community would have a chance not only to understand how students learn, but also how student learning may be related to disparate aggregate patterns.

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

Findings

In this paper Ho highlights how the measurement community is concerned with validation, and how this concern is inherently tied to the purpose of assessment. He argues that if purpose is not central to an assessment's validation, then there can be costly unintended consequences (p. 2). But, maintaining the test maker's purpose is difficult when the test enters a public space where its use and meaning can be manipulated. He calls this phenomenon "purpose drift." This leads him to suggest that test makers should have "purpose drift" in mind as they develop new tools of measurement.

Ho uses Haertal and Herman (2006) to illustrate how assessment was initially designed to serve two goals: to assess where individuals are to be placed and to improve the quality of education (p. 3). However, after the publication of *A Nation at Risk*, assessment began to directly inform public opinion. Ho uses Haertal (2012) to point out that when assessment enters public consciousness it takes on another role beyond measurement. It begins to influence society and policy.

Ho argues that the influencing nature of assessment can have unintended consequences that vary in their severity. These depend upon what type of assessment is used, what the results indicate for different groups, how much time it takes for the results to influence practice, and how high stakes is the assessment. Ho believes that purpose drift creates an ethical dilemma and responsibility with which the testing community must grapple.

Recommendations

Ho calls for the testing community to have increased “foresight” on the consequences of purpose drift. He believes that potential routes of purpose drift should be thought of and explicitly taken into account when tests are developed.

Commentary

I think that Ho’s notion of “purpose drift” is useful for the future of assessment. Ho has provided an analytical tool that can be used to both look backward and forward at the testing industry. I think the concept is useful for understanding how policy, practice, and assessment interact and how mistakes made in the past might be avoided in the future.

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

Findings

According to Robert Linn, test-based accountability has been the most persistent policy initiative in education for the past 50 years (Elmore, 2004). He argues that this has been especially true lately, because American students are performing well below their international counterparts and because of the persistence of the racial/ethnic achievement gap. In this article he briefly charts how testing and accountability has grown to become an influential education reform strategy.

Linn argues that test-based accountability (especially NCLB) was premised on the goals of increasing student achievement and equity in sub-populations. However, these goals rest on several assumptions. One such assumption is that teachers know how to improve student achievement. The second is that the sanctions and rewards associated with achievement can be adequately linked to teacher performance. The third is that achievement tests correctly measure student learning and that the tests can't be manipulated in practice. These assumptions have proved problematic and unrealized.

With this in mind, Linn provides a brief history of how test-based accountability gained prominence. He argues that assessment was always part of education, but that it gained public attention with the passing of the Elementary and Secondary Education Act and the publication of *A Nation at Risk*. He claims that after these events, all 50 states responded by creating criterion and standards-based tests to measure student achievement. It was not until the 1990s, when rewards and sanctions were first applied to assessment, that test-based accountability became a prominent reform strategy that only intensified with the passage of NCLB.

Linn is skeptical of the future of test-based accountability as a means to raise student achievement. He is also cautious in lauding attempts to refine the current system with developments such as value added and teacher evaluations.

Recommendations

Linn offers the following suggestions for the future of assessment:

- further develop the common core and the support efforts of PARC and SBAC
- push to include qualitative approaches to accountability (i.e., self evaluations)
- Measures of Effective Teaching (MET) project (Kane and Staiger, 2012)
- UTeach Observation Protocol (UTOP)

interim assessments and formative assessments

Commentary

I find this piece useful if it can be put into a productive conversation with Ho's notion of "purpose drift." Linn paints a broad picture of how assessing students has become politically fraught and relatively divorced from improving student learning. If purpose drift can be applied to Linn's paper the measurement community might be able to figure out how to approach predicting possible influencing features of assessment.

Mendoza-Denton, Rodolfo (2012). *Cultural Identity, Existential State and the Measurement of Human Performance.*

Findings

Rodolfo Mendoza-Denton attempts to shed light on the importance of bias in testing. However, he departs from traditional discussions of test bias and proposes that bias within the testing environment is just as important. He says that individual psychology is inseparable from social context (p. 5) and that the self and one's environment are intricately intertwined. However, he clearly states that participants are not passive; rather, they are active and interactive (p. 7). This is important because certain environments may facilitate productive or destructive reactions to failure. These environments have "attributional ambiguity" where participants are unable to understand if a negative outcome is attributed to one's own shortcomings or to the bias in people around them (p. 9). Mendoza-Denton believes that ambiguous environments may contribute to the achievement gap more so than test bias.

According to Mendoza-Denton the education system contains contradictory hegemonic messages that allow for ambiguity to arise (i.e., celebrate diversity in a colorblind community). He argues that racial discrimination has become more covert, which allows for ambiguous signals of unequal treatment to trigger attributional ambiguity (p. 12). Research shows that ambiguity can be a negative influence on performance because it promotes uncertainty about one's perceptions and ability (p. 19). While ambiguity can generate self protective tendencies it can also lead to disengagement because feedback is not trusted. This phenomenon is related to stereotype threat, social identity threat (Steele, Spencer, Aronson, 2002), and situation specific psychological disengagement (Major et al., 1998).

Mendoza Denton critiques the testing community for placing more emphasis on test bias rather than considering how educational environments affect students. He believes social context is an extremely important factor to consider because interpersonal interactions affect how students will do, how they will perceive their efforts, and how much they will invest in their work.

Recommendations

For the future of assessment Mendoza-Denton admits that his paper is hard to translate to practice. However, he offers a few suggestions and recommendations:

- create tests that assess a wider variety of skills than currently being focused on. He cites a study about LSAT scores and effective lawyers. (Shultz and Zedeck, 2011)
- move away from dissemination of ideas of fixed intelligence to ones that highlight incremental and malleable intelligence
- use explicit anti tracking propaganda
- increase diversity in the testing industry
- recognize that there is “a strong need to create environments in which people can trust the fairness of the feedback they receive, as well as their own belongingness within these environments (Mendoza-Denton et al., 2010; Cheryan, Plaut, Davies and Steele, 2009; Walton and Cohen, 2007, 2011)

Commentary

I think this piece lays groundwork for how the measurement community can think about connecting assessment to broader social issues. This seems like an important intersection to explore when thinking through issues of equity and excellence in assessment. It also is related to both Varenne and Cauce and Gordon’s papers because it stresses the importance of human interactions and how engagement or disengagement with learning can transpire.

Gergen, Kenneth and Dixon-Roman, Ezekiel (2012). *Epistemology in Measurement – Part II. Social Epistemology and the Pragmatics of Assessment.*

Findings

In this paper Gergen and Dixon-Roman use a constructivist approach to explore how positivist and empiricist ideologies have shaped the measurement community. They argue that although positivism and empiricism are often understood as value free and objective, they are not. They say that the empiricist focus on things like validity, reliability, and scaling make claims to truth that privileges one set of knowledge over another. They use a constructivist perspective along with Kuhn's notions of scientific revolutions (1968) to trace the ideological underpinnings of assessment and measurement and suggest that a paradigm shift may be underway.

The authors believe that the measurement community has placed too much emphasis on the cost and benefits of education. They fear that the intrinsic desire to learn or to teach has been muted by an individualistic focus that is not in tune with a world that demands collaboration. They argue that currently, education and measurement divides and stratifies society and creates arbitrary hierarchies that have serious social implications. Additionally they critique the current measurement system for not allowing local differences and human characteristics to be considered. The authors believe that measurement, as it now exists, is counterproductive to the needs of the 21st century.

Recommendations

The authors propose that the future considers using assessment strategies that run counter to positivist/empiricist trends. They provide several examples:

- empowerment evolution (Fetterman, 2000, 2004, 2007)
- dialogic evaluation (Ryan and Stefano, 2000; Schwandt, 2005)
- appreciative evaluation (Preskill and Catsambas, 2006; Coghlan, Preskill and Catsambas, 2003)
- socio-cultural or situative assessment (Moss et al., 2008; Gee, 2008a)

Commentary

I think that this paper's discussion on individualism and neoliberalism is important to put into conversation with the Cauce and Gordon paper. Gergen and Dixon-Roman's critique on individualism strengthens Cauce and Gordon's suggestion that developing human agency be central to education, rather than individual achievement. If these two papers could be combined around this subject, I think it may be possible to make an even stronger case for placing the development of human agency at the center of education.

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

Findings

Federal policy dictates that state and local education agencies must provide accommodations in assessments for students with disabilities. Martha Thurlow's article discusses how accommodations have evolved over time. She describes how they have been received and viewed by the measurement community and how they have grown to be inclusive of both students with and without disabilities and English Language Learners (ELLs).

Research on the effects of accommodations for students with disabilities has been conducted on a range of topics, from college entrance exams to state level tests. Initially these studies focused on understanding whether or not accommodations leveled the playing field for students with disabilities. As time passed, research began to ask whether or not tests remained valid when accommodations were used. This research relied upon the "interaction hypothesis," (Sireci, Scarpati, and Li, 2005) which posited that accommodations would not change test scores unless a person has a disability. This hypothesis eventually gave way to the "differential boost hypothesis," (Fuchs and Fuchs, 1999, Sireci et al 2005) which posits that students with disabilities benefit more from accommodations than those without a disability.

However, Thurlow notes that the increase in availability of computer-based assessments has led to recent developments in accommodations research known as "universal design" (Thompson, Thurlow, and Malouf, 2004; Johnstone, Altman, and Thurlow, 2006). "Universal design" is an inclusive form of accommodation that allows students to show relevant knowledge and skills through "multiple means of representation, action/expression and engagement" (p. 14). It also is a way to reduce variation in test administration. Thurlow argues that "universal design" has made accommodations a part of the process of assessment rather than separate from it, and thus has made them relevant for all types of learners.

Recommendations

Thurlow recommends the following for the future of assessment:

- Move away from theories of differential boost and think of students as being defined not by their disability, but by their learning needs.
- Increase access to college by creating more congruency between K-12 settings and assessment and college entrance exams and settings.

Commentary

I think that Thurlow's article places emphasis on increasing access to fair assessment. However, I think that a focus on access is not enough for the future of assessment. I am preoccupied with the fact that the article rarely discusses how access intersects with privilege and varying social contexts. I think this article could continue to add to discussions on the future of assessment if it engaged with Varenne's focus on the importance of human interactions within the schooling process.

Reflection

In reflecting on the papers, I am intrigued by how several of the authors have tried to contextualize assessment within the current political and policy landscape and within varying social contexts. This has led me to think about how the future of assessment could benefit from engaging in a more systemic approach for understanding how students learn. I think that Chung's proposal for combining both coarse and fine-grained data has enormous potential. Not only will it further personalize education, but it also can also highlight how micro instances of learning and engagement are related to larger macro and structural facets of society.

I imagine that if data were used more systematically it could not only show disparities between sub groups, but it also could provide insight into how these disparities manifest. Systematic use of data could show how varying outcomes converge or diverge around social and economic resources and points of privilege. Additionally, I imagine that if multiple layers of data are used to describe the same learning experience or learner, the assessment community could see how particular constructs under an empiricist/positivist approach are value laden and privilege some knowledge's over others. This idea builds off of the findings in Gergen and Dixon-Roman's paper.

Lastly, I believe that my participation in the Gordon Commission allowed me to grow as a scholar. The process of collaboration in which the other fellows and I contributed was enormously helpful in processing the Commission's papers. Each fellow approached the papers with a different perspective that furthered my understanding of and engagement with the material. Additionally, over the course of our meetings I was able to better understand how my own interests in the intersection between policy and practice could be applied to assessment and measurement.

References

Antonenko, P. D., Toy, S. & Niederhauser, D. S. (2012). *Using cluster analysis for data mining in educational technology research. Educational Technology Research and Development, 60*, 383–398.

Baker, E. L. (2000). *Understanding educational quality: Where validity meets technology*. In William Angoff Memorial Lecture Series. Princeton, NJ: Educational Testing Service.

Bandura, A. (1986). *Social Foundations of thought and action: A social cognitive theory*. New Jersey: Prentice-Hall, Inc.

Bandura, A. (1991). *Self-regulation of motivation and action through anticipatory self reactive mechanisms.* Nebraska Symposium on Motivation.

Bandura, A. (1997). *Self-Efficacy: The exercise of control*. USA: W.H. Freeman and Company.

Bandura, A. (2001). *Social cognitive theory: An agentic perspective. Annual Review of Psychology, 52*, 1-26.

Bandura A. (2002). *Social cognitive theory in cultural context. Journal of Applied Psychology*: *An International Review, 51*,269-290.

Bandura, A. (2005). *Guide for constructing self-efficacy scales. In Self Efficacy Beliefs of Adolescents*, 307-337.

Bandura, A. (2006). Towar

Campbell, J. P. & Oblinger, D. (2007). *Academic analytics*. Washington, DC: EDUCAUSE Center for Applied Research.

Cheryan, S., Plaut, V. C., Davies, P. G., & Steele, C. M. (2009). *Ambient belonging: How stereotypical cues impact gender participation in computer science. Journal of Personality and Social Psychology, 97,* 1045-1060. doi:10.1037/a0016239

Coghlan, A.T., Preskill, H. & Catsambas, T.T. (2003) *An overview of Appreciative Inquiry in evaluation. New directions for evaluation*. *100*, Winter.

Elmore, R. F. (2004). *The problem of stakes in performance-based accountability systems.* In S. H, Fuhrman & R. F. Elmore (eds.), *Redesigning accountability systems for education*. New York, Teachers College Press.

Fetterman, D.M. (2000) *Foundations of empowerment evaluation*. Thousand Oaks, CA: Sage. Fetterman, D.M. & Wandersman, A. (Eds.) (2004) *Empowerment evaluation, principles in practice*. New York: Guilford. (2007) *Empowerment evaluation, yesterday, today, and tomorrow*. *American Journal of Evaluation*, *28*, 179-198.

Foucault, M. (1978) *The history of sexuality, Vol. 1, an introduction*. New York: Pantheon (1980) *Power/knowledge*. New York: Pantheon.

Fuchs, L.S., & Fuchs, D. (1999). *Fair and unfair testing accommodations. School Administrator, 56*(10), 24-29.

Gee, J.P. (2008). *A sociocultural perspective on opportunity to learn.* In Moss, P.A., Pullin, D.C., Gee, J.P., Haertel, E.H. & Young, L.J. (eds.). *Assessment, Equity, and Opportunity to Learn*. New York NY: Cambridge University Press.

Haertal (2012)

Haertal and Herman (2005)

Johnstone, C.J., Thompson, S.J., Miller, N.A. & Thurlow, M.L. (2008). *Universal design and multi-method approaches to item review. Educational Measurement: Issues and Practice, 27*(1), 25-36.

Kane, T. J. & Staiger, D. O. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation*. NBER Working Paper No, 14607.

Kuhn, T.S. (1970) *The structure of scientific revolutions*. 2nd ed. Rev. Chicago: University of Chicago Press.

Major, B., Spencer, S., Schmader, T., Wolfe, C. & Crocker, J. (1998). *Coping with negative stereotypes about intellectual performance: The role of psychological disengagement.* Personality & Social Psychology Bulletin, 24, 34-50.

Mendoza-Denton, R., Goldman-Flythe, M., Pietrzak, J., Downey, G. & Aceves, M. J. (2010). *Group-Value Ambiguity: Understanding the Students' Self-Esteem. Science*. doi:10.1177/1948550609357796.

Moss, P.A., Pullin, D.C., Gee, J.P., Haertel, E.H. & Young, L.J. (2008). *Assessment, Equity, and Opportunity to Learn*. New York NY: Cambridge University Press.

Preskill, H. & Catsambas, T.T. (2006) *Reframing evaluation through appreciative inquiry*. Thousand Oaks: CA: Sage.

Ryan, K.E. and DeStefano, L. (Eds.) (2000) *Evaluation as a Democratic process: Promoting inclusion, dialogue, and deliberation.* San Francisco: Jossey-Bass.

Sen, A. (1985) Well-being, agency, and freedom. The Dewey Lectures 1984. *The Journal of Philosophy*, *82*, 169-221.

Sen, A. (1987). *The standard of living.* In G. Hawthorn (ed). *The Standard of Living*. Cambridge: Cambridge University Press, pgs. 1-38.

Sen, A. (1992). *Inequality re-examined*. Oxford: Clarendon Press.

Sen A. K. (1999a). *Democracy as a universal value. Journal of Democracy, 10*, pgs. 3-17.

Sen, A. (1999) *Development as Freedom*. Oxford: Oxford University Press.

Sen, A. (October 28, 2003). *The Importance of basic education*. Speech delivered at the Commonwealth Education Conference, Edinburgh (accessed through http://people.cis.ksu.edu/~ab/Miscellany/basiced.html)

Simon, H. A. (1957). *Models of man: Social and rational: Mathematical essays*. New York: Wiley.

Shultz, M.M. & Zedeck, S. (2011). *Predicting lawyer effectiveness: broadening the basis for law school admission decisions. Law & Social Inquiry, 36*, 620-661.

Stanovich, K. E. & West, R. F. (2000). *Individual differences in reasoning: Implications for the rationality debate. Behavioral and Brain Sciences, 23*, 645–665.

Steele, C. M., Spencer, S. J. & Aronson, J. (2002). *Contending with group image: The psychology of stereotype and social identity threat.* In M. P. Zanna (Ed.), Advances in experimental social psychology (Vol. 34, pgs. 379–440). San Diego, CA: Academic.

Schwandt, T.A. (2005) *A diagnostic reading of scientifically based research for education. Educational Theory*, 55, 285-305.

Sireci, S.G., Scarpati, S.E. & Li, S. (2005). *Test accommodations for students with disabilities: An analysis of the interaction hypothesis. Review of Educational Research, 75*(4), 457-490.

Thompson, S.J., Thurlow, M.L. & Malouf, D. (2004). *Creating better tests for everyone through universally designed assessments. Journal of Applied Testing Technology, 10*(2).

U.S. Department of Education (DOE). (2010). *Transforming American education: Learning powered by technology*. Washington, DC: Author.

van Barneveld, A., Arnold, K. E. & Campbell, J. P. (2012). *Analytics in higher education: Establishing a common language*. Washington, DC: EDUCAUSE Center for Applied Research.

Walton, G. M. & Cohen, G. L. (2007). A question of belonging: race, social fit, and achievement. *Journal of Personality and Social Psychology*, *92,* 82–96.

Walton, G. & Cohen, G. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. *Science*, *331*, 1447–1451. doi:10.1126/science.1198364