Educational Researcher

Mitigating the Dangers of a Single Story: Creating Large-Scale Writing Assessments Aligned With Sociocultural Theory

Nadia Behizadeh

The online version of this article can be found at:

A more recent version of this article was published on - Apr 11, 2014

Published on behalf of

American Educational Research Association

and

SAGE

Additional services and information for *Educational Researcher* can be found at:

Email Alerts: http://er.aera.net/alerts

Subscriptions: http://er.aera.net/subscriptions

Reprints: http://www.aera.net/reprints

Permissions: http://www.aera.net/permissions

Version of Record - Apr 11, 2014

>> OnlineFirst Version of Record - Mar 27, 2014

What is This?

Mitigating the Dangers of a Single Story: Creating Large-Scale Writing Assessments Aligned With Sociocultural Theory

Nadia Behizadeh[1]

The dangers of a single story in current U.S. large-scale writing assessment are that assessment practice does not align with theory and this practice has negative effects on instruction and students. In this article, I analyze the connections or lack of connections among writing theory, writing assessment, and writing instruction, critique the construct and consequential validity of direct writing assessment and portfolio assessment, and reframe reliability as local consensus among experts. A new vision of large-scale sociocultural writing portfolios in K–12 education is offered that builds on the practices of past large-scale portfolio assessment but also encourages students to write in multiple languages/dialects and modes for multiple purposes. Another key feature of sociocultural portfolios is that students are encouraged to write for impact. These additional components will mitigate the dangers of a single story by ensuring that (a) assessment practice matches sociocultural writing theory, resulting in high construct validity, and (b) teachers are encouraged to utilize culturally sustaining pedagogy, resulting in high consequential validity.

Keywords: assessment; literacy; meta-analysis; multiculturalism; social context; validity/reliability; writing

In her speech "The Danger of a Single Story," the writer Chimamanda Adichie (2009) demonstrated through her experience that there is not just one story to be told about a place or person, but many stories. She stated, "The single story creates stereotypes, and the problem with stereotypes is not that they are untrue, but that they are incomplete. They make one story become the only story." Her warning of a single story applies to education, and many scholars have emphasized the importance of multicultural education that recognizes and celebrates student diversity (e.g., Nieto & Bode, 2012), allowing many stories to be heard. However, in large-scale writing assessment in U.S. elementary and secondary schools, this danger of a single story is very real.

In many current large-scale writing assessments in the United States, such as those administered by districts and states to K–12 students, a student writes one timed essay on an assigned prompt, and then this one writing sample is used to evaluate that student's writing ability (Jeffery, 2009). This practice is known as direct writing assessment (DWA). On most DWAs, students can only write in Standard American English (SAE), a dialect of English that is the language of power in the United States (Gee, 2012; McWhorter, 2000). There are two ways this assessment

practice illustrates the dangers of a single story. The first danger is that because the particular type of story valued by current large-scale writing assessment is an academic essay written in SAE, when teachers teach to the test, they will be less likely to use culturally sustaining pedagogy (Paris, 2012) that values multiple forms of language. Assessments are needed that make "teaching to the test" less of an epithet; students and teachers deserve assessments that honor different ways of writing and connect to students' funds of knowledge (Gonzalez & Moll, 2002).

Secondly, if a child does not perform well on this timed essay, there will be a single story told about this student: he/she has below basic skills in writing, or maybe even far below basic skills. Yet this same student may be a brilliant poet or have a hundred pages of a first novel carefully stowed in his/her backpack. However, when a single story of deficiency is repeated again and again to a student, that student develops low writing self-efficacy and a poor self-concept of himself/herself as a writer (Ball & Ellis, 2008; Bandura, 1997). Thus, the second danger of the single story is the negative effect on students when one piece of

---

[1]Georgia State University, Atlanta, GA

Table 1
Alignment Between Writing Theory and Assessment Practice

	Current large-scale direct writing assessment	Large-scale portfolio assessment (as implemented in past)	New vision of large-scale sociocultural writing portfolio assessment
Theory	Writing is employing correct mechanics, conventions, and forms.	In addition to employing correct mechanics, conventions, and forms, writing is a process-based reflective practice.	In addition to being process-based and reflective, writing is a set of sociocultural practices enacted in varied contexts. Mechanics, conventions, and forms vary by context.
Assessment	Students write one timed essay on an assigned prompt.	Students draft, revise with assistance, and then reflect on multiple pieces of writing; different genres are represented.	Includes the practices of past large-scale portfolio assessment. Additionally, students are encouraged to write in multiple dialects/languages and multiple modes for multiple purposes, and to increase the social impact of their writing.
	The essay is scored by a third party outside of the school using a traditional rubric with categories for ideas, conventions, style, and organization.	Portfolios are scored by groups of teachers and administrators from the school or district. Rubric includes traditional categories from direct writing assessment with the addition of categories for process and reflection.	Portfolios are scored by groups of teachers and administrators, possibly including community members. Rubric includes traditional categories plus categories for multimodal effectiveness, process and reflection, writing in appropriate dialects/voices for different purposes, and writing for impact on an audience.
Reliability	Reliability defined as consistency of scores. By this definition, this assessment process is reliable.	Reliability defined as consistency of scores. By this definition, this assessment process may have issues with reliability.	Reliability defined as local consensus among experts. By this definition, this assessment process is reliable.
Validity	Assessment process has poor construct validity and poor consequential validity. Process does not align with sociocultural theories of writing and has negative washback for instruction.	Assessment process has moderate construct validity and moderate consequential validity. Process better aligned with sociocultural theories of writing but does not value multiliteracies; process has some positive washback on instruction, although it may not encourage culturally sustaining pedagogy.	Assessment process has high construct validity and high consequential validity. Process more closely aligned with conception of sociocultural writing that values multiliteracies; process has positive washback on instruction and encourages culturally sustaining pedagogy.

writing on a decontextualized prompt is used to represent writing ability.

Overall, two related main dangers of the single story of direct writing assessment exist: This practice has negative effects on instruction and negative effects on students. I argue that these dangers are a result of an assessment practice that is not aligned with prevailing theories of literacy developed by sociocultural literacy researchers. What is needed are large-scale sociocultural writing assessments that encourage culturally sustaining and meaningful classroom writing instruction. By critically examining the connections or lack of connections among writing theory, writing assessment, and writing instruction, in this article I propose a new vision for large-scale sociocultural writing portfolios in K–12 education. First, I explain how writing is a set of social and cultural practices, drawing on research from prominent literacy scholars. Second, I compare direct writing assessment with portfolio assessment and examine how each of these practices connects to theory. In this section, I also critique the traditional definition of reliability used in the measurement community for writing assessment and offer an alternative definition. Third, I consider the poor consequential validity of current large-scale writing assessments through the examination of evidence that these assessments are doing more harm than good. Finally, I discuss my vision of sociocultural portfolios as one

possible solution to the dangers of a single story in large-scale writing assessment. The major points of this article are summarized in Table 1.

Writing Defined as a Set of Sociocultural Practices

Over the last century, writing research has evolved from a mechanical view, to a cognitive model, and most recently to sociocultural theories of writing (Behizadeh & Engelhard, 2011). This evolution does not mean that mechanics and cognitive processes are no longer valued facets of writing, but rather that the task of writing is ultimately rooted in a writer's sociocultural background and the current sociocultural context in which the writer is creating text. As clearly stated by Prior (2006), "Sociocultural theories represent the dominant paradigm for writing research today" (p. 54). In its most basic sense, sociocultural theory considers varied contexts and defines writing as a contextual process. As the term "sociocultural" indicates, these contexts are comprised of social and cultural components.

In her review of sociocultural literacy theory, Perry (2012) presented three major subareas: literacy as a social practice, multiliteracies, and critical literacy. Considering Perry's divisions, I posit that an understanding of literacy as a social practice is the

foundation for all sociocultural literacy theories. Across these theories, a commonality is that literacy varies depending on the particular context in which a reading or writing event occurs. This context includes the communicative purpose for writing shaped by the intended audience and collaborating others, as well as the conventions within the cultures the writer is embedded (Perry, 2012). Because of the varied contexts influencing literacy events, sociocultural theory emphasizes the importance of cultural and linguistic variation in relation to defining writing.

Perry (2012) stated, "In answer to the question 'What is literacy?,' theorists of literacy as a social practice would say that literacy is what people do with reading, writing, and texts in real-world contexts and why they do it" (p. 54). Writing is defined by texts people create for authentic contexts and purposes. Therefore, using a sociocultural understanding of literacy, writing is the ability to use *multiple* modes and dialects/languages for varied sociocultural purposes; writing ability is a "set" of practices. To emphasize an important point for understanding this definition of writing, I employ Yancey's (1999) metaphor of "waves" of writing assessment theories and practices. Each new wave builds upon the past so that new understandings of writing assessment—or in this case, writing theory—do not negate past theories but expand upon them. Thus, when writing for a specific sociocultural purpose, written products are generated through complex cognitive processes and guided by context-appropriate conventions for writing (Behizadeh & Engelhard, 2011).

Culturally relevant pedagogy (Irvine & Armento, 2001; Ladson-Billings, 2009), although not exclusively a literacy theory, is one broad sociocultural theory critical to consider in the case for reforming large-scale writing assessment. Ladson-Billings (2009) detailed the characteristics of a culturally relevant teacher, including the teacher acknowledging the unique community and global identities of the students. Paris (2012) refined the idea of culturally relevant pedagogy by proposing the term *culturally sustaining pedagogy*. This refinement emphasizes that instruction should do more than relate to a student's culture; instruction should "support young people in sustaining the cultural and linguistic competence of their communities while simultaneously offering access to dominant cultural competence" (Paris, 2012, p. 95). Taking Paris's stance a step further, assessments encouraging culturally sustaining pedagogy must allow students to express their linguistic competence in ways resonant with both home and school literacy practices.

Street's (1984) conception of two opposing models of literacy illuminates the contrast between decontextualized models of writing and sociocultural models of writing. The first model is the autonomous model that proposes one vision of literacy assumed to be the same internationally. Street critiqued this model, stating, "What is taken in the 'autonomous' model to be qualities inherent to literacy are in fact conventions of literate practice in particular societies" (p. 4). Bakhtin's (1981) idea of unitary language is similar to Street's autonomous model. Bakhtin explained how unitary language cannot fully represent the infinite versions of language that exist in real life, what he calls "heteroglossia." Bakhtin's view is that the unitary or standardized way of writing or speaking is in itself a social construct that exists amongst many other possibilities for literate expression. Similar to the concept of heteroglossia, Street (1984)

proposed the ideological model, positing that the meaning of literacy is dependent on the social context. Echoing Bakhtin's view, Street (1984) stressed that "the meaning of literacy depends upon the social institutions in which it is embedded" (p. 8). In his later work, Street (2001) clarified that the ideological model "does not attempt to deny technical skill or the cognitive aspects of reading and writing, but rather understands them as they are encapsulated within cultural wholes and within structures of power" (p. 435). Thus, culture and power are primary considerations for defining writing, yet they do not eclipse the importance of acquiring skills and learning conventions.

Illustrating how writing varies, just over half a century ago, Diederich, French, and Carlton (1961) conducted a study on writing assessment in which 60 writing experts read 300 essays and scored them. The result was that 94% of the essays received at minimum seven unique scores. Often used to emphasize the need for standardized writing assessment, this study does just the opposite; it demonstrates the complexity of writing and how notions of writing quality range widely even among experts.

Because the meaning of literacy changes depending on sociocultural context, instead of referring to students being literate or illiterate, a more useful concept is the idea of "multiliteracies," which emphasizes the many ways students can be literate including diverse multilingual and multimodal forms of literacy (Cope & Kalantzis, 2000). Recent work has emphasized the importance of teaching and assessing students' abilities to compose in multiple modes, including linguistic (written), audio, spatial, gestural, and visual (Burke & Hammett, 2009; Cope & Kalantzis, 2000; Kimber & Wyatt-Smith, 2009). Regarding multilingual diversity, Heath's (1983) *Ways with Words* demonstrated how the issue was not whether students from different communities were literate but whether or not students were literate in a way that was valued by the school. Heath outlined the complex oral traditions of an African American community and how those traditions had less value in school. In fact, the African American practice of creative elaboration was construed by classroom teachers as lying. In addition, Gee (2001) presented the idea of primary and secondary discourse, where primary discourse is learned at home and secondary discourse is learned chiefly at school. For students whose primary discourse overlaps to a large degree with their secondary discourse, their transition from home to school is easier. Yet for students whose primary discourse differs greatly from their secondary discourse, success at school is more difficult.

These researchers have contributed to an understanding of writing as a set of sociocultural practices, and their work emphasizes that there is not just one correct version of literacy. Aligning writing assessment practice with theory is a requirement for construct validity. A test has construct validity if "it measures what it purports to measure" (IRA/NCTE, 2010, p. 52). As Jeffery (2009) stated, "A writing assessment is considered valid only when the constructs it measures are grounded in strong theoretical underpinnings" (p. 5). But it is not enough to have a strong theory on which to build an effective writing assessment system; we need an *accurate* theory. Scholars have noted that writing assessments at the state level often reflect a current-traditional rhetoric view of writing, a view that privileges conventions and organization over content and creativity and has been out of

vogue for many years now (Jeffery, 2009). By defining writing as a set of sociocultural practices enacted in varied settings, an accurate measurement of this process necessitates an assessment based on this definition.

A final consideration for understanding writing as a set of sociocultural practices is the purpose of writing, or the intended outcome. A number of scholars focus on the impact students achieve through their writing, including researchers utilizing participatory action research methods (Morrell, 2008; Torre & Fine, 2006) and researchers studying authentic writing (Behizadeh, 2012; Newmann, Marks, & Gamoran, 1996; Purcell-Gates, Duke, & Martineau, 2007). In order for writing assessment to be socioculturally relevant to students, the purpose and outcome of the writing needs to be considered in the assessment. For example, if students write a persuasive essay about why they deserve an expensive cell phone, the assessment should consider if the writing was effective in obtaining the phone. This is what real writing in society is about: writing to communicate ideas or to make change. Authentic writing, where students write meaningful texts that they perceive as connected to their lives, is a component of sociocultural writing theory (Behizadeh, 2012; Perry, 2012).

Comparison of Large-Scale Writing Assessment Practices

Considering the need for theory to support practice, a critical question to explore is, How do large-scale writing assessment practices align with sociocultural theories of writing? The three large-scale writing assessment practices most used in the United States in the last century are multiple-choice tests, direct writing assessments, and writing portfolios (Yancey, 1999). Although multiple-choice questions are a common assessment practice for writing, most states acknowledge the lack of construct validity in this practice and supplement multiple-choice items on writing with a direct writing assessment. On the other hand, portfolios are rarely used for large-scale writing assessment. In this article, I focus on direct writing assessments and portfolios. I describe each of these practices, examine their alignment with sociocultural theory to evaluate construct validity, and also critique their reliability claims.

Direct Writing Assessments

Many current large-scale writing assessments are direct writing assessments (DWAs). As already noted, DWAs are generally timed essays written by an individual without assistance in response to a standardized prompt. These essays are then sent to a testing center for evaluation where raters have been carefully trained to have the same concept of good writing.

The writing tests in Georgia are examples of this type of assessment. Students take the test in the 3rd, 5th, 8th, and 11th grades. In the 8th grade, students have 100 minutes to complete an essay on a prompt such as this one from 2012: "A family in your town has decided to donate a large amount of money to a person, charity, or cause. Choose a person, a group, or a charity that you think deserves the money" (Georgia Department of Education, 2012). At all grade levels, student work is sent to an evaluation center

where two trained raters rate each essay using an interpretive guide that includes ideas, organization, style, and conventions.

A major issue is that these assessments use a very narrow definition of reliability. Moss (1994) explained how historically, reliability has been determined "by examining consistency, quantitatively defined, among independent observations or sets of observations that are intended as interchangeable" (p. 6). Hillocks (2002) interviewed raters involved in the scoring process for a DWA in Illinois and reported,

> My source said that the scoring was a rigorous experience, because all knew that if they failed to remain reliable or if they failed to maintain the speed, they would be fired. Raters in this situation are not allowed to use their own judgment about a paper. They must try to get the "right" score using the official criteria. The effort is to minimize disagreement, that is, to remain reliable. (p. 120)

The solution in the measurement community to the issue of divergent scores was to establish a unitary view of good writing, in other words, to standardize the idea of good writing (Huot, 1990). In order to reliably assess writing on a large scale, measurement experts have continued to create writing assessments that are aligned with a unitary or autonomous view of literacy.

Regarding construct validity, as noted by Weigle (2002), a serious issue with current large-scale DWAs is "the fact that writing done under timed conditions on an unfamiliar topic does not accurately reflect the conditions under which most writing is done in non-testing situations or writing as it is taught and practiced in the classroom" (p. 197). This single story only represents what a student can do alone, without resources or extended time for reflection, in response to a particular prompt. Although this assessment may provide some information on a student's writing skills, it does not allow a student to demonstrate the range of genres, multiliterate capabilities, or impact on audience that an assessment technique that considered many writing samples and student reflection would allow. Hamp-Lyons (2002) noted how other dialects of English including Jamaican English, African American English, and Indian English, are devalued by DWAs that only privilege one way of writing. As it stands now, the current dominant writing assessment practice ignores both the social process of writing and different cultural ways of writing and does not align with current theories of writing.

On the other hand, one could argue that if writing is a set of sociocultural practices and these practices vary by context, then writing for a DWA could be seen as a particular context that demands a particular literacy practice. This argument would have some credence if writing in SAE to a random prompt was one of many social practices considered when issuing judgment on a student's writing ability. But this is the major issue; DWAs are purporting to represent writing ability writ large when they in fact use one narrow version of literacy to represent a broad construct. Additionally, a DWA serves no communicative function, and its sociocultural relevance is tautological; the test is creating a new context for writing that does not exist outside of testing. Therefore, writing for a DWA has low construct validity; students are not demonstrating their ability to employ a variety of modal and language resources for various authentic purposes.

Although I have focused on DWAs evaluated by human raters, I should note that automated essay scoring (AES) is becoming more prevalent and moving writing assessment practices even further from sociocultural theories (Condon, 2013; Deane, 2013). Referring to writing tests using AES, Condon (2013) concluded, "Because these tests underrepresent the construct as it is understood by the writing community, such tests should not be used in writing assessment, whether for admissions, placement, formative, or achievement testing" (p. 100). The same can be said for DWAs.

Writing Portfolios

Portfolios for writing assessment offer a better match between sociocultural theory and assessment practice than DWAs. Calfee and Freedman (1996) defined a literacy portfolio as "a folder with situated samples of student reading and writing performance" (p. 17). They specified that the contents of the portfolio should be driven by the goals of instruction; the teacher should be the primary evaluator with outside moderation; and the focus should be on capturing in-depth literacy processes. In their description, Calfee and Freedman emphasized that a portfolio is more than keeping work in a folder. Two critical elements of writing portfolios are that they contain evidence of the student's reflective process and that summative evaluation is delayed until the portfolio is completed (Hamp-Lyons & Condon, 2000; Weigle, 2002). For summative evaluation, flexible rubrics that can be applied to different genres of and purposes for writing have been used for successful large-scale programs. These rubrics include more traditional rhetorical elements of writing such as grammar, mechanics, and organization, as well as elements regarding critical analysis, reflection, and range of writing genres (LeMahieu, Gitomer, & Eresh, 1995; Weigle, 2002). Student reflections on the writing tasks offer critical contextualization for the assessment. Thus, portfolios exhibit evidence of student performance with social support in different genres and offer cultural and personal contextualization via student reflections, aligning this assessment practice closer to sociocultural theory than DWAs.

As an illustrative example, Kentucky's past writing portfolio assessment was administered in three grade levels. Students were not given prompts but were required to represent certain genres of writing including personal expressive, literary, and reflective pieces (Hillocks, 2002). In the scoring process, teams of teachers and administrators at the school evaluated portfolios using a holistic scoring guide. Two raters would read an entire portfolio, decide on a score, and then compare their decisions. If there was disagreement, the two raters could either discuss their rationale and try to arrive at consensus, or send the portfolio to a third rater. Similar portfolio systems were used in Vermont and Oregon (Elbow & Belanoff, 1997; Yancey, 1999).

Perceived issues with the reliability of portfolios for writing assessment have been major roadblocks for implementing them large-scale (Petruzzi, 2008; Yancey, 1999). The oft-cited reliability issue with the Vermont state-wide writing portfolio assessment where interrater reliability was low (Koretz, Stecher, Klein, & McCaffrey, 1994) is used as a rationale against large-scale portfolio assessment. In their review of portfolio implementation and design issues, Herman, Gearhart, and Aschbacher (1996) appear to be against portfolios for large-scale assessment because of reliability issues. As one solution to reliability concerns, the authors suggested making the contents of the portfolio uniform. However, this suggestion could prevent students from writing for authentic, sociocultural purposes. In fact, Calfee and Freedman (1996) shy away from fully endorsing portfolio assessment for large-scale high-stakes testing because of the knowledge that portfolios can be easily reduced to a series of decontextualized, rote writing.

Instead of rejecting large-scale portfolio assessment, the key is to avoid the conflation of reliability with standardization and to reframe reliability (Moss, 1994; Yancey, 2012). Drawing on Moss's (1994) work, I propose a more socioculturally oriented idea of reliability, where reliability is defined as local consensus among qualified evaluators. If reliability is understood in this way, then the method used in the past by Kentucky (Hillocks, 2002) and currently by the National Writing Project (2012) of local panels that reach a consensus is the best practice. Moss (1994) also noted that this consensus model of portfolio review is successfully used for peer review processes outside of K–12 school settings.

In a local consensus process for writing portfolios, qualified evaluators make independent observations on a collection of student writing and offer a rationale for judgments based on agreed upon standards and a consideration of the social and cultural contexts in which writing occurred, provided in part by reading student reflections on the process and purpose of each writing sample. Regarding consensus, Columbini and McBride (2012) argued that "writing assessment would benefit from seeking approaches to communal interaction that foreground the importance of discussion not only *preceding* but *throughout* assessment processes" (p. 195). Returning to the trope of a single story, a consensus process allows for various interpretations of a body of work, or multiple stories about a student's writing performance, to be voiced and considered, and then integrated into an agreed upon final judgment.

The qualified evaluators should include the English Language Arts (ELA) teacher of the student and other educators and administrators from the same school and perhaps also from other schools in the district, providing a balanced perspective. The ELA teacher most likely will conduct evaluations prior to group scoring in order to complete formative assessments. ELA teachers are qualified as experts by their credentials certifying that they have met professional standards, and also through the many hours spent interacting with the student, performing formative evaluation of his/her writing, and through the ongoing professional development in which teachers engage related to instruction and assessment. Yancey (1999) stressed that "teachers are experts in a local sense—authoritative about the relationship between a student and a specific course" (p. 496). As professionals with content, context, and student knowledge, the teacher provides a crucial voice in a socioculturally sound writing assessment. A number of scholars have advocated for the teacher as a primary rater of his/her students' writing (Darling-Hammond & Snyder, 1992; Freedman, 1993; Moss, 1994). Rather than posing a threat to reliability, the teacher provides more data to the team of evaluators, improving the accuracy of their judgment.

In order for the consensus model to be successful, educators need to engage in professional development to create a successful portfolio system. Fortunately, effective professional development models already exist such as the National Writing Project (NWP) and local versions such as the Bay Area Writing Project (Calfee & Freedman, 1996) and Pittsburgh's writing portfolio system (LeMahieu et al., 1995). An argument against portfolio assessment could be that the financial burden of this method is too great because of the investment in time and human resources, but college composition scholars who both research and participate in portfolio assessment disagree (Condon, 2013; Hester, O'Neill, Neal, Eddington, & Huot, 2007). Condon (2013) offered his experience, stating, "At my own institution, where faculty assess more than four thousand rising-junior writing portfolios every year, collecting and assessing these portfolios yields rich data about students' writing performances across the curriculum, about faculty expectations and standards, and about what kinds of assignments work best in various contexts" (p. 105). By building on models like the NWP and those implemented at the college level, and training local teachers and administrators to be raters of student work, portfolios can be cost-efficient and contribute to professional development.

Portfolios allow students to demonstrate their writing achievement in multiple genres for multiple audiences, to provide contextualization of their work through reflection, and to seek peer and teacher input during the process of writing. These practices better align portfolios with sociocultural theories of writing. However, past portfolio assessment practice did not explicitly encourage students to utilize different dialects, languages, or modes, or to reflect on impact, concerns I return to when I present my vision for sociocultural portfolios.

Poor Consequential Validity of Current Large-Scale Direct Writing Assessments

In the previous sections, I have established that portfolios are better aligned with sociocultural theories of writing than DWAs. In addition to poor construct validity, by valuing a single story written in Standard American English, DWAs have poor consequential validity. Beck and Jeffery (2007) define consequential validity as "the larger social consequences of using a particular test for a particular purpose" (p. 61). Supporting the importance of consequential validity, the most recent U.S. standards for educational assessment stress that it is not the test that is valid, but it is the use of the test that determines the validity (AERA, APA, & NCME, 1999). Similarly, recommendations for assessing literacy from the International Reading Association and National Council of Teachers of English (IRA/NCTE, 2010) state, "The consequences of an assessment procedure are the first and most important consideration in establishing the validity of the assessment" (p. 22). Aligning large-scale writing assessments with sociocultural constructs of literacy by shifting to portfolio assessment will not only improve the construct validity but also improve consequential validity. Focusing on linguistic and cultural issues, in this section I illuminate the ways in which the negative consequences of decontextualized tasks such as DWAs on poor, urban students are so dire that it is the civic responsibility of educators and policymakers to address these issues.

In K–12 education, an assessment is high-stakes when performance on the assessment determines grade advancement, high school graduation, or college admission. However, the stakes are also high for teachers who are more likely than ever to be evaluated by the performance of their students on these tests due to initiatives developed to receive federal Race to the Top funds (Department of Education, 2009). In general, most large-scale writing assessments administered by state boards of education are also high-stakes writing assessments.[1]

A plethora of studies illustrate that K–12 high-stakes writing assessment drives instruction, a phenomenon known as "washback" (Au, 2007; Dappen, Isernhagen, & Anderson, 2008; Hillocks, 2002; Jeffery, 2009; Messick, 1996; Moss, 1994). Washback occurs when the assessment influences how content is taught. Although intended outcomes should drive instruction (this is what is intended by backward design), negative washback occurs when assessments are poorly designed based on inaccurate understandings of a construct, leading to ineffective pedagogical practices such as rote memorization of decontextualized facts (Au, 2007). Negative consequences of high-stakes assessment are more prevalent in schools serving populations that generally do not perform as well on standardized high-stakes writing assessments, such as schools serving poor children of color or English Language Learners (ELLs; Ball & Ellis, 2008; Carbonaro & Gamoran, 2002; Inoue & Poe, 2012; McHenry & Heath, 1994; Solórzano, 2008; Winn & Behizadeh, 2011).

Ball and Ellis (2008) reviewed decades of writing research regarding teaching culturally and linguistically diverse students and concluded "that students of color are disproportionately relegated to classrooms using drill exercises rather than interactive, meaningful approaches that require extended writing, reflection, and critical thinking" (p. 507). Although Ball and Ellis did not explore possible reasons for an overemphasis on drills, research has demonstrated that teachers often assume that speakers of nonstandard dialects such as African American Vernacular English (AAVE) need remediation due to a perception that these students are behind in literacy acquisition (Balester, 2012; McWhorter, 2000; Smitherman & Villanueva, 2003). Furthermore, Winn and Behizadeh (2011) emphatically stated,

> [T]he devaluing of language is still a functioning practice in American schools. Those who will not conform, who will not accept that their history and language is deficient, all too often end up dropping out of school, or more accurately, being pushed out. Clearly, the right to literacy has been and is being violated for students of color. (p. 156)

I argue that large-scale tests that only value one dialect contribute to the devaluing of a student's primary discourse.

In part because of the lack of linguistic equality in schools, African American students consistently score lower on national reading and writing assessments than their White peers (Salahu-Din, Persky, & Miller, 2008; Vanneman, Hamilton, Baldwin Anderson, & Rahman, 2009). This has come to be "reliable," and a lot of focus has been on closing the Black–White achievement gap by raising test scores for students of color. However, one overlooked factor is that students' primary discourse is not valued in school (Gee, 2001; Heath, 1983). Although there may

be an achievement gap when only the dominant discourse is measured, there is no gap in literate practices.[2] As Heath (1983) and Gee's (2001) work demonstrate, students develop rich literate practices in their home communities even if they do not acquire the standard dialect of English.

A number of researchers have established the benefits of using students' primary language as a resource to help students become proficient in the dominant discourse; in a sense, using the primary as a bridge to the secondary without devaluing the primary (Gonzalez & Moll, 2002; Gutierrez, Baquedano-Lopez, & Tejada, 1999; Lee, 2006; Morrell & Andrade, 2002; Pacheco, 2012). Accessing and building on students' funds of knowledge (Gonzalez & Moll, 2002; Pacheco, 2012), connecting sonnets to hip hop (Morrell & Andrade, 2002), and creating a "third space" where students and teachers find common ground (Gutierrez, Baquedano-Lopez, & Tejada, 1999; Lee, 2006) can both honor primary discourses and teach the dominant discourse. Yet, because assessment practices drive instruction and *only* value the dominant discourse, teachers are less likely to engage in hybrid instructional practices, quite simply because stakes are high and time is limited. According to Lee (2006), with a few exceptions, "there are virtually no systematic supports for leveraging AAVE [African American Vernacular English] as a linguistic resource for academic learning broadly speaking, and school-based literacy learning in particular" (p. 307). Beyond the first year of intensive instruction for newcomers, ELLs in inclusion classes face a similar situation where their primary languages are not seen as linguistic resources in their own right, nor as linguistic foundations on which to build their knowledge of SAE (Kinloch, 2005; Smitherman & Villanueva, 2003).

Ultimately, "poor performance on high-stakes test disproportionately discourages linguistic- and ethnic-minority students from completing their high school course work and advancing to college" (Beck & Jeffery, 2007, p. 61). Instead of identifying the test as the problem, the students are seen as the problem. Returning to Adichie's (2009) concern with stereotypes, when this single story of failure is repeated again and again about certain groups of students who speak proficiently in different dialects of English or other languages, and are working to master Standard American English, this single story turns into a damaging discourse of deficiency about ELLs and students of color. It is important to note that negative washback from standardized large-scale writing assessments is not solely a problem for students of color and ELLs. White students whose primary discourse may be more aligned with school discourse (Heath, 1983) may perform adequately on large-scale writing assessments, but may not acquire higher-level and critical thinking skills that are associated with writing for expression or reflection rather than writing to demonstrate proficiency.

Some critics may argue that the intention of large-scale state writing assessments is not to evaluate the depth and breadth of student writing but to assess whether or not students are proficient in writing academic essays in SAE. Although students need to be able to write in SAE and know the conventions of academic essay writing, as has been often noted, "what you test is what you get" (Jeffery, 2009; White, 1985). In the case of DWAs, when only one way of writing is tested, and when that test has high stakes for teachers and students, instruction will focus on this one form of writing and the recognition and integration of multiliteracies in instruction will be less likely to occur. When the single story is tested, the single story will be valued.

Negative washback from theoretically misaligned assessments is why accurately measuring the construct of writing is so important. This section illustrated the negative effects of using a linguistically narrow definition of writing to develop writing assessments, a practice that has in part normalized the achievement gap. The next section outlines the possibility of portfolios for sociocultural large-scale writing assessments and the positive consequences for instruction.

High Consequential Validity: Toward a Vision of Sociocultural Large-Scale Writing Assessment

In terms of consequential validity, researchers have cited positive washback from assessment to instruction as a major rationale for portfolio assessment in writing (Weigle, 2002; Yancey, 1999, 2002). One positive consequence for instruction is that a range of writing is taught when portfolios include poetry, novellas, or reflections that employ a variety of modes in addition to more traditional persuasive and expository essays. By including these diverse genres of writing, students demonstrate the range of writing they are able to achieve. Also, portfolio assessments generally result in students having more choice of their writing and developing greater ownership and pride of their finished products (Calfee & Freedman, 1996). In addition, interactive student discussions are encouraged and process-oriented methods for instruction such as writing workshops are often utilized. Finally, students participate in the ongoing evaluation of their progress and are able through reflection on each piece of work to provide a rationale (Yancey, 2002, 2009). In addition to allowing students to write for authentic, socioculturally diverse purposes, these pedagogical practices align with strategies that have been shown to help struggling readers and writers develop their writing skills (Kamil, Borman, Dole, Kral, Salinger, & Torgesen, 2008).

Many excellent pedagogies from past portfolio use exist, but additional elements are required to synchronize portfolio assessment with sociocultural literacy theories. My vision of large-scale sociocultural writing portfolio assessment builds on past large-scale portfolio assessment but expands the assessment process to include three additional key components necessary for construct validity that may result in even higher consequential validity: honoring linguistic diversity, integrating multimodalities, and encouraging writing for impact.

The first and most significant change is that students should be encouraged to write in different dialects of English or other languages in which they desire to communicate. For example, students may choose to write a narrative in African American Vernacular English (AAVE) or a poem in Chicano English. English Language Learners (ELLs) can choose to write in their home language, whereas those students most comfortable with SAE can write another piece in this dialect. These pieces would be included along with writing samples demonstrating proficiency in SAE. Reflections in SAE would include the value of writing in different ways for different audiences. Although a writing sample in another language or dialect will be included,

because teachers will not have fluency in all the languages and dialects represented in their classrooms, only the reflection on the sample will be assessed. Allowing students to express their multilingual abilities would more closely align assessment practice with sociocultural theories of writing, improving construct validity.

Secondly, also included in the reflection should be student evaluation of the impact of their writing. Students will evaluate if their writing had the intended effect on their audience, which is important for the social purpose of writing. In their reflections, students can also provide a rationale for the modes they chose to effectively communicate their position or ideas. Encouraging students to write for impact is aligned with authentic writing instruction (Behizadeh, 2012) and critical text production for community change (Morrell, 2008; Pacheco, 2012). These two additional parameters for portfolio assessment systems are required to match practice to the full construct of writing.

A third and final area of alignment is multimodal composition, writing that includes more recent compositional forms such as blogs, wikis, or slideshow presentations, but also encompasses traditional written essays supplemented with tables, figures, and graphics. By encouraging students to integrate multiple modes including audio, gestural, spatial, and visual aspects (Cope & Kalantzis, 2000) in addition to written language, students will develop the skills necessary for navigating and creating content in the 21st century (Kimber & Wyatt-Smith, 2009; McGrail & Behizadeh, 2014). Past portfolios such as the "process-folios" from the Arts PROPEL project (Harvard University, 2013) that showcased in-process multimodal work with ongoing reflection are good starting points for evaluating compositions that integrate multiple modes.

A critical question to consider is what might this look like? What follows is my proposal based on past portfolio research and the three additional factors of honoring linguistic diversity, integrating multimodalities, and encouraging writing for impact. Although rubrics will be developed and/or adapted by individual schools, I offer an example here of what the practice of sociocultural writing portfolios might look like in an eighth-grade class:

In the second semester of instruction, after months of composing different texts for different purposes, students select three samples from their classroom portfolios (see Figure 1). Two of these are written in SAE and represent academic genres such as expository, narrative, or persuasive essays. For the third sample, students are allowed to choose a language/dialect and genre. All three pieces have a written reflection in SAE explaining the impact of the writing and the rationale for the language/dialect. Additionally, this reflection includes students' thoughts on their writing process and how they arrived at the final draft. Under impact, students are asked to explain the purpose of the writing and why they chose different modes, detail who read or heard the piece, and also present any impact their writing had. However, students do not lose points on the reflection for not achieving their intended impact; in the reflection they can evaluate why the intended impact did not occur and offer next steps for how they will make their writing more powerful.

Students then submit these samples from their portfolio for large-scale assessment at the school level. Depending on the decisions of the local scoring team, students may be required to

Genre: _____ (expository, narrative, persuasive, poetry, fiction, etc.)

Language\dialect: _____ (Standard American English, African American English, Spanish, Tagalog, etc.)

	Category	Rating
Writing Sample (used for SAE samples)	Ideas	1 2 3 4
	Organization	1 2 3 4
	Style	1 2 3 4
	Conventions	1 2 3 4
	Multimodal Effectiveness	1 2 3 4
Reflection	Impact	1 2 3 4
	Rationale for language/dialect	1 2 3 4
	Process	1 2 3 4

FIGURE 1. *Writing rubric for samples from sociocultural portfolios (eighth-grade example)*

submit drafts in addition to final pieces and reflections. After participating in 1 to 2 days of professional development and consensus building, all teachers and administrators participate in scoring these samples. Schools may choose to have community members participate in the training and scoring as well. For the scoring session, I have developed a possible rubric (Figure 1) that requires a traditional analytic rating on the two SAE essays in addition to evaluations of the three reflections. I have also added a strand for evaluating multimodality in which the raters determine if the chosen modes for expression are effective. Instead of asking the raters to judge the third piece, which may be in an unknown dialect or language, raters may only be able to review the selection and award points for the reflection. For the final score, the two essays are averaged, comprising 80% of the final grade (see Figure 2). Reflections are also averaged and comprise 20% of the final grade. This possible weighting is designed to make the essays in SAE count for the majority of the final score, but also allow the quality of the reflections to impact the final score. Because this process is "in-house" and serves as an important professional development opportunity for all teachers and administrators, the costs in terms of time and money are relatively low.

The above suggested assessment process focuses on Standard American English proficiency yet also allows space for students to present work in other languages and dialects alongside the standard dialect of English. Again, this is only a possible assessment process and rubric; local sites may develop a different rubric that they believe better matches their shared philosophy of education and aligns with standards. However, states and/or

Sample 1 in SAE		
Writing Sample	1 2 3 4
Reflection	1 2 3 4

Sample 2 in SAE		
Writing Sample	1 2 3 4
Reflection	1 2 3 4

Sample 3 in student's choice of language/dialect		
Reflection	1 2 3 4

Comments (optional):

Possible breakdown for final score:
Writing sample average comprises 80% of final score
Reflection average comprises 20% of final score

FIGURE 2. *Rubric for Sociocultural Portfolio Assessment*

districts should provide a template each school can adapt as long as their agreed upon process aligns with standards, encourages linguistic and multimodal diversity, and includes student reflection on impact.

Before implementation, this framework will need to be piloted, revised, and elaborated and decisions will need to be made about how to interpret various scores. Also, perhaps analytic scoring of two essays in addition to three reflections is too time-consuming, and holistic scoring will be a better option. Pilot studies may find that educators find four samples provide more information and help them arrive at a decision. However, the essential point here is that the assessment process is expanded to allow more than a single story to be evaluated, the students' reflection on process, impact, and language to be considered, as well as multilingual and multimodal proficiencies to be presented.

I foresee high consequential validity for this assessment process. In particular, I propose two specific positive effects of large-scale sociocultural portfolio assessment on instruction: (a) teachers will be more likely to use culturally sustaining pedagogy that values multiple literacies, and (b) students who speak dialects of English other than Standard American English will see that their literate practices are valued and be able to develop their home and community literate practices as well as develop more school-based literate practices (Delpit, 1995; Winn & Behizadeh, 2011). These effects support the idea that this iteration of portfolio assessment will have high consequential validity, in addition to high reliability through a process of evaluation through local consensus.

Conclusion

Writing instruction that fails to connect to students' funds of knowledge is not only conceptually unsound but pedagogically impotent. As established in cognitive science, "New information that cannot be tied to any prior knowledge is not learned well or at all" (Gee, 2008, p. 77). In order to write in the dominant discourse, students need to see how their out-of-school literate practices connect to in-school practices. Yet for instruction to change, assessment needs to change. What is untenable is to continue the cycle of inauthentic, standardized tests driving instruction, and then allow the resulting poor instruction to contribute to low test scores and detrimental labels for particular groups of students. Linking instruction to assessment, Moss (1994) stated over a decade ago, "Ultimately, the purpose of educational assessment is to improve teaching and learning" (p. 10). I believe sociocultural portfolios are one answer to the need for large-scale writing assessment aligned with sociocultural theories of writing.

If large-scale assessments are created that value a range of literacy practices, including dominant literacy practices, then instruction will be more student-centered and curriculum will be expanded to meet the diverse needs of a diverse student body. This is not to say that all students should not acquire the dominant discourse. To allow all students equal opportunities and access to wider career options, students need to understand the codes of power and language of power (Delpit, 1995). McWhorter (2000) acknowledges that schools should focus on aiding students in developing the standard dialect of English, yet clarifies that "the job of the school is to add a new layer to a child's speech repertoire, not to undo the one they already have" (p. 15). Sociocultural portfolio assessment honors students' primary discourse while supporting the development of SAE.

Recently, a movement has emerged in higher education to utilize writing portfolios aligned with sociocultural theories, but these attempts are generally restricted to individual institutions and do not affect K–12 education (Wardle & Roozen, 2012). As previously noted, pilot studies of sociocultural portfolios in K–12 schools are needed to examine the issues and benefits of this model implementation on a larger scale. In particular, cultural-historical activity analyses (Engeström, 2001) that consider the entire classroom system can provide detailed data on how to forge a "third space" of instruction and assessment where multiple ways of being literate intersect. This process of linking assessment practice with an understanding of writing as a set of sociocultural practices will take time. Freedman (1993) cautioned that for portfolios to work for large-scale testing, "Both testers and teachers will have to recognize that the process will be complex and the results of reciprocal relationships will not be immediate" (p. 48). However, it is worth the time and investment to ensure that instruction and students do not suffer from negative washback.

Missing from this article is a thorough consideration of how to assess the multimodal dimensions of writing. Although the tentative framework offered includes a strand on multimodal effectiveness, I did not include a detailed explanation on how this would be scored. The consideration of multimodalities is another important step in the future of writing assessment and some solid starting points do exist (Burke & Hammett, 2009; Harvard University, 2013; Kimber & Wyatt-Smith, 2009). One intriguing element from Kimber and Wyatt-Smith's (2009) multimodal assessment framework is their inclusion of a blank rubric strand they termed the "X factor." Their idea is that if we truly want students to be innovative, we must allow for the possibility that students will create something unprecedented that requires unforeseen criteria. What this might look like in practice needs to be investigated further.

As a final note, sociocultural writing portfolios not only align with prevailing sociocultural theories of writing but also can be linked to the Common Core State Standards (CCSS; Common Core State Standards Initiative, 2012). The CCSS include specifications for the production and distribution of writing, particularly focusing on the importance of audience. Hopefully, these additions will support instruction that encourages students to write for impact on a real audience. In addition, the college and career readiness (CCR) standards for language in the CCSS include a strand titled "Knowledge of Language" stating that K–12 students should "apply knowledge of language to understand how language functions in different contexts" (Common Core State Standards Initiative, 2012, pp. 25, 51). Although speaking or writing in different dialects or languages aside from SAE is not explicitly mentioned, this CCR standard can be more effectively realized by allowing students to consider a wide spectrum of language and the different conventions required for varying contexts.

States, districts, and teachers need large-scale writing assessments that provide useful and reliable information on student progress, align with writing as a set of sociocultural practices by valuing multiple forms of literacy, and encourage sound pedagogical practices. Although sound assessments will not prevent poor instruction, sound assessment will encourage culturally relevant and sustaining instructional practices. The proposed sociocultural writing portfolio assessment system is designed to capture the multi-voiced, dynamic nature of writing, including some of the context that enriches writing and makes it meaningful. Sociocultural portfolios will allow students to tell multiple stories and document their growth in many directions, mitigating the dangers of a single story.

NOTES

[1]Although reducing high stakes to moderate stakes is just as important for the effective functioning of our educational system in the United States as creating socioculturally sound writing assessment, it is outside the scope of this paper to fully explore the former. Please see Mintrop and Sunderman (2009) for a number of compelling reasons for reducing the stakes of large-scale testing.

[2]To clarify, a student's primary discourse is not his/her only discourse nor does a person's ethnicity indicate their preferred discourse. Specifically regarding AAVE and SAE, many African American families may feel equally comfortable in both dialects or prefer SAE to AAVE. The point is not to assume students' primary discourses but to allow students to express themselves in the discourses they choose.

REFERENCES

Adichie, C. (2009). The danger of a single story. Retrieved from http://www.ted.com/talks/chimamanda_adichie_the_danger_of_a_single_story.html

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education.
(1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, *36*, 258–267.

Bakhtin, M. M. (1981). *The dialogic imagination: Four essays by M. M. Bakhtin*. Austin: The University of Texas Press.

Balester, V. (2012). How writing rubrics fail: Toward a multicultural model. In A. B. Inoue & M. Poe (Eds.), *Race and writing assessment* (pp. 63–77). New York, NY: Peter Lang.

Ball, A. F., & Ellis, P. (2008). Identity and the writing of culturally and linguistically diverse students. In C. Bazerman (Ed.), *Handbook of research on writing: History, society, school, individual, text* (pp. 499–513). New York, NY: Lawrence Erlbaum.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W. H. Freeman.

Beck, S., & Jeffery, J. (2007). Genres of high-stakes writing assessments and the construct of writing competence. *Assessing Writing*, *12*, 60–79.

Behizadeh, N. (2012, November). *Making middle school writing meaningful: An analysis of factors increasing or decreasing perceptions of authenticity in writing*. Paper presented at the 2012 Annual Meeting of the Literacy Research Association (LRA), San Diego, CA.

Behizadeh, N., & Engelhard, G. (2011). Historical view of the influences of measurement and writing theories on the practice of writing assessment in the United States. *Assessing Writing*, *16*, 189–211.

Burke, A., & Hammett, R. F. (2009). *Assessing new literacies: Perspectives from the classroom*. New York, NY: Peter Lang.

Calfee, R., & Freedman, S. (1996). Classroom writing portfolios: Old, new, borrowed, blue. In R. Calfee & P. Perfumo (Eds.), *Writing portfolios in the classroom: Policy and practice, promise and peril* (pp. 3–26). Mahwah, NJ: Lawrence Erlbaum.

Carbonaro, W. J., & Gamoran, A. (2002). The production of achievement inequality in high school English. *American Educational Research Journal*, *39*, 801–827.

Columbini, C. B., & McBride, M. (2012). "Storming and norming": Exploring the value of group development models in addressing conflict in communal writing assessment. *Assessing Writing*, *17*, 191–207.

Common Core State Standards Initiative. (2012). *Common Core state standards for English language arts & literacy in history/social studies, science, & technical subjects*. Retrieved from http://www.corestandards.org/assets/CCSSI_ELA%20Standards.pdf

Condon, W. (2013). Large-scale assessment, locally-developed measures, and automated scoring essays: Fishing for red herrings? *Assessing Writing*, *18*, 100–108.

Cope, B., & Kalantzis, M. (Eds.). (2000). *Multiliteracies: Literacy learning and the design of social futures*. New York, NY: Routledge.

Dappen, L., Isernhagen, J., & Anderson, S. (2008). A statewide writing assessment model: Student proficiency and future implications. *Assessing Writing*, *13*, 45–60.

Darling-Hammond, L., & Snyder, J. (1992). Reframing accountability: Creating learner-centered schools. In A. Lieberman (Ed.), *The changing contexts of teaching (91st yearbook of the National Society for the Study of Education)*. Chicago, IL: University of Chicago Press.

Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing*, *18*, 7–24.

Delpit, L. (1995). *Other people's children: Cultural conflict in the classroom*. New York, NY: Free Press.

Department of Education. (2009). *Race to the top program executive summary*. Retrieved from http://www2.ed.gov/programs/racetothetop/executive-summary.pdf

Diederich, P. B., French, J. W., & Carlton, S. T. (1961). *Factors in judgments of writing ability* (Research Bulletin No. RB-61-15). Princeton, NJ: Educational Testing Service.

Elbow, P., & Belanoff, E. (1997). Reflections on an explosion: Portfolios in the 90's and beyond. In K. Yancey & I. Weiser (Eds.), *Situating portfolios: Four perspectives* (pp. 21–34). Logan, UT: Utah State University Press.

Engeström, Y. (2001). Expansive learning at work: Toward an activity theoretical reconceptualization. *Journal of Education and Work*, *14*, 133–156.

Freedman, S. W. (1993). Linking large-scale testing and classroom portfolio assessments of student writing. *Educational Assessment*, *1*, 27–52.

Gee, J. P. (2001). Literacy, discourse, and linguistics: Introduction. In E. Cushman, E. R. Kintgen, B. M. Knoll, & M. Rose (Eds.), *Literacy: A critical sourcebook* (pp. 525–554). Boston, MA: Bedford/St. Martin's.

Gee, J. P. (2008). A sociocultural perspective on opportunity to learn. In P. Moss, D. C. Pullin, & J. P. Gee (Eds.), *Assessment, equity, and opportunity to learn* (pp. 76–108). Cambridge: Cambridge University Press.

Gee, J. P. (2012). *Social linguistics and literacies: Ideology in discourses* (4th ed.). New York, NY: Routledge.

Georgia Department of Education. (2012). *Writing assessments*. Retrieved from http://www.doe.k12.ga.us/Curriculum-Instruction-and-Assessment/Assessment/Pages/Writing-Assessments.aspx.

Gonzalez, N., & Moll, L. (2002). Cruzando el Puente: Building bridges to funds of knowledge. *Educational Policy*, *16*, 623–641.

Gutierrez, K. D., Baquedano-Lopez, P., & Tejada, C. (1999). Rethinking diversity: Hybridity and hybrid language practices in the third space. *Mind, Culture, and Activity*, *6*, 286–303.

Hamp-Lyons, L. (2002). The scope of writing assessment. *Assessing Writing*, *8*, 5–16.

Hamp-Lyons, L., & Condon, W. (2000). *Assessing the portfolio: Principles for practice, theory, and research*. Cresskill, NJ: Hampton Press.

Harvard University. (2013). *Project Zero: Arts PROPEL*. Retrieved from http://www.pz.gse.harvard.edu/arts_propel.php

Heath, S. B. (1983). *Ways with words: Language, life, and work in communities and classrooms*. Cambridge: Cambridge University Press.

Herman, J., Gearhart, M., & Aschbacher, P. (1996). Portfolios for classroom assessment: Design and implementation issues. In R. Calfee & P. Perfumo (Eds.), *Writing portfolios in the classroom: Policy and practice, promise and peril* (pp. 27–59). Mahwah, NJ: Lawrence Erlbaum.

Hester, V., O'Neill, V., Neal, M., Eddington, A., & Huot, B. (2007). Adding portfolios to the placement process: A longitudinal perspective. In P. O'Neill (Ed.), *Blurring boundaries: Developing writing, researchers, and teachers* (pp. 61–90). Creskill, NJ: Hampton Press.

Hillocks, G. (2002). *The testing trap: How state writing assessments control learning*. New York, NY: Teachers College Press.

Huot, B. (1990). The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research*, *60* 237–263.

Inoue, A. B., & Poe, M. (2012). Introduction. In A. B. Inoue & M. Poe (Eds.), *Race and writing assessment* (pp. 1–11). New York, NY: Peter Lang.

IRA/NCTE Joint Task Force on Assessment. (2010). *Standards for the assessment of reading and writing* (Revised edition). Newark, DE: International Reading Association and National Council of Teachers of English.

Irvine, J. J., & Armento, B. J. (2001). *Culturally responsive teaching: Lesson planning for elementary and middle grades*. New York, NY: McGraw-Hill.

Jeffery, J. V. (2009). Constructs of writing proficiency in US state and national writing assessments: Exploring variability. *Assessing Writing*, *14*, 3–24.

Kamil, M. L., Borman, G. D., Dole, J., Kral, C. C., Salinger, T., & Torgesen, J. (2008). *Improving adolescent literacy: Effective classroom and intervention practices: A practice guide* (NCEE #2008–4027). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc.

Kimber, K., & Wyatt-Smith, C. (2009, September). *Rethinking quality assessment for 21st century learning: How students use and create knowledge online*. Paper presented at Assessment for a Creative World, 35th Annual Conference of International Association of Educational Assessment (IAEA), Brisbane, Australia.

Kinloch, V. F. (2005). Revisiting the promise of "students' right to their own language": Pedagogical strategies. *College Composition and Communication*, *57*, 83–113.

Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). The Vermont Portfolio Assessment Program: Findings and implications. *Educational Measurement: Issues and Practice*, *13*(3), 3–16.

LeMahieu, P., Gitomer, D., & Eresh, J. (1995). Portfolios in large-scale assessment: Difficult but not impossible. *Educational Measurement: Issues and Practice*, *14*(3), 11–16, 25–28.

Ladson-Billings, G. (2009). *The dreamkeepers: Successful teachers of African American children* (2nd ed.). San Francisco, CA: Jossey-Bass.

Lee, C. D. (2006). "Every good-bye ain't gone": Analyzing the cultural underpinnings of classroom talk. *International Journal of Qualitative Studies in Education*, *19*, 305–327.

McGrail, E., & Behizadeh, N. (2014). *Evaluating multimodal assessment frameworks: The interactive audience as a missing element of online multimodal composition*. Manuscript submitted for publication.

McHenry, E., & Heath, S. B. (1994). The literate and the literary: African Americans as writers and readers—1830–1940. *Written Communication*, *11*, 419–444.

McWhorter, J. (2000). *Spreading the word: Language and dialect in America*. Portsmouth, NH: Heinemann.

Messick, S. (1996). Validity and washback in language testing. *Language Testing*, *13*, 241–256.

Mintrop, H., & Sunderman, G. L. (2009). Predictable failure of federal sanctions-driven accountability for school improvement: And why we may retain it anyway. *Educational Researcher*, *38*, 353–364.

Morrell, E. (2008). *Critical literacy and urban youth: Pedagogies of access, dissent, and liberation*. New York, NY: Routledge.

Morrell, E., & Duncan-Andrade, J. M. (2002). Promoting academic literacy with urban youth through engaging hip-hop culture. *English Journal*, *91*, 88–93.

Moss, P. A. (1994). Can there be validity with reliability? *Educational Researcher*, *23*, 229–258.

National Writing Project (2012). National Writing Project offers high-quality writing assessment services. Retrieved from http://www.nwp.org/cs/public/download/nwp_file/15507/awc_brief.pdf?x-r=pcfile_d

Newmann, F. M., Marks, H. M., & Gamoran, A. (1996). Authentic pedagogy and student performance. *American Journal of Education*, *104*, 280–312.

Nieto, S., & Bode, P. (2012). *Affirming diversity: The sociopolitical context of multicultural education* (6th ed.). Boston, MA: Pearson.

Pacheco, M. (2012). Learning in/through everyday resistance: A cultural-historical perspective on community resources and curriculum. *Educational Researcher*, *41*, 121–132.

Paris, D. (2012). Culturally sustaining pedagogy: A needed change in stance, terminology, and practice. *Educational Researcher*, *41*(3), 93–97.

Perry, K. (2012). What is literacy? A critical overview of sociocultural perspectives. *Journal of Language and Literacy Education*, *8*(1), 50–71. Retrieved from http://jolle.coe.uga.edu/wp-content/uploads/2012/06/What-is-Literacy_KPerry.pdf

Petruzzi, A. (2008). Articulating a hermeneutic theory of writing assessment. *Assessing Writing*, *13*, 219–242.

Prior, P. (2006). A sociocultural theory of writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 54–66). New York, NY: Guilford.

Purcell-Gates, V., Duke, N. K., & Martineau, J. A. (2007). Learning to read and write genre-specific text: Roles of authentic experience and explicit teaching. *Reading Research Quarterly*, *42*, 8–45.

Salahu-Din, D., Persky, H., & Miller, J. (2008). *The nation's report card: Writing 2007* (NCES 2008–468). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://nces.ed.gov/nationsreportcard/

Smitherman, G., & Villanueva, V. (2003). *Language diversity in the classroom: From intention to practice*. Urbana, IL: National Council of Teachers of English.

Solórzano, R. W. (2008). High stakes testing: Issues, implications, and remedies for English language learners. *Review of Educational Research*, *78*, 260–329.

Street, B. V. (1984). *Literacy in theory and practice*. Cambridge: Cambridge University Press.

Street, B. V. (2001). The new literacies studies. In E. Cushman, E. R. Kintgen, B. M. Kroll, & K. M. Rose (Eds.), *Literacy: A critical sourcebook*. Boston, MA: Bedford/St. Martin's.

Torre, M., & Fine, M. (2006) Resisting the passive revolution: Democratic, participatory research by youth. In D. Carlson (Ed.), *Democratic Leadership: Keeping the Promise*. Retrieved from http://web.gc.cuny.edu/psychology/socpersonality/Fine/Mfine/res_resisting.pdf

Vanneman, A., Hamilton, L., Baldwin Anderson, J., & Rahman, T. (2009). *Achievement gaps: How Black and White students in public schools perform in mathematics and reading on the National Assessment of Educational Progress* (NCES 2009–455). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Wardle, E., & Roozen, K. (2012). Addressing the complexity of writing development: Toward an ecological model of assessment. *Assessing Writing*, *17*, 106–119.

Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.

White, E. (1985). *Teaching and assessing writing*. San Francisco, CA: Jossey-Bass.

Winn, M. T., & Behizadeh, N. (2011). The right to be literate: Literacy, education, and the school-to-prison pipeline. *Review of Research in Education*, *35*, 147–173.

Yancey, K. B. (1999). Looking back as we look forward: Historicizing writing assessment. *College Composition and Communication*, *50*, 483–503.

Yancey, K. B. (2002). Portfolios in the writing classroom: A final reflection. In K. B. Yancey (Ed.), *Portfolios in the writing classroom: An introduction* (pp. 102–116). Urbana, IL: National Council of Teachers of English.

Yancey, K. B. (2009). Reflection and electronic portfolios: Inventing the self and reinventing the university. In D. Cambridge, B. Cambridge, & K. Yancey (Eds.), *Electronic portfolios 2.0: Emergent research on implementation and impact* (pp. 5–16). Sterling, VA: Stylus.

Yancey, K. B. (2012). College admissions and the Insight Resume. In A. B. Inoue & M. Poe (Eds.), *Race and writing assessment* (pp. 171–185). New York, NY: Peter Lang.

AUTHOR

NADIA BEHIZADEH, PhD, is an assistant professor of adolescent literacy at Georgia State University, Department of Middle and Secondary Education, 30 Pryor Street, Atlanta, GA, 30303; *nbehizadeh@gsu.edu*. Her research focuses on authentic literacy instruction and sociocultural writing assessment.