

Righting Technologies: How Large-Scale Assessment Can Foster a More Equitable Education System

Nadia Behizadeh^{a1} and Tom Liam Lynch^b

^a Georgia State University

^b Pace University

Abstract

For the last century, the quality of large-scale assessment in the United States has been undermined by narrow educational theory and hindered by limitations in technology. As a result, poor assessment practices have encouraged low-level instructional practices that disparately affect students from the most disadvantaged communities and schools. In this historical and theoretical review, we examine the misalignment between educational theory and large-scale assessment practices that rely upon technology, using writing assessment as a case in point. Drawing upon sociocultural theory and critical software studies as conceptual frameworks, we find that today's software-powered technologies, although capable of taking progressive educational ideals to scale, have not been used for these purposes. Our proposed solution is to shift from using technologies to assess predetermined samples of evidence of learning to using technologies to facilitate complex and negotiated models of assessment. This solution would require policy shifts that honor the needs of various stakeholders in the assessment process. We offer a power-sharing concept called negotiated control that engages policymakers, educators, researchers, and community members in the assessment process.

Keywords: large-scale assessment, education technology, writing assessment, sociocultural theory, software

For over 20 years, U.S. presidential administrations—from Clinton to Obama—have made concerted attempts to use large-scale assessment to overhaul public K–12 education in the name of equity. Through the Goals 2000: Educate America Act in 1994, the Clinton administration sought to renovate large-scale assessment nationwide and invest in school-technology infrastructure. In 2009, the Obama administration's Race to the Top program earmarked over \$4 billion to entice states and districts to reform schools from

¹ Correspondence concerning this article should be addressed to: Dr. Nadia Behizadeh, College of Education and Human Development, Middle and Secondary Education, 30 Pryor Street, 6th floor, Atlanta, GA, 30303. Email: nbehizadeh@gsu.edu.

the ground up by adopting reform strategies that included common assessment systems. More recently, the Every Student Succeeds Act in 2015 continued the requirement of annual large-scale assessment for most grade levels in multiple subject areas. At the heart of these initiatives is a century-old belief that rigorous, large-scale assessment is key to ensuring that schools provide equal opportunities for all children to learn. Yet, time and time again, such initiatives seem to fall far short of their goals, utilizing a great deal of public funding and incurring incalculable opportunity costs (Darling-Hammond, 2010; Noguera & Wells, 2011).

Despite rhetoric to the contrary, policymakers—and many of the philanthropic institutions and companies whom they enlist as partners—define what it means to assess children’s learning in ways that are deeply rooted in turn-of-the-century United States. We argue that one of the causes of policymakers’ failure to realize 21st century education reform lies in their own 19th century paradigm concerning large-scale assessment models. More specifically, we argue that what it means to assess children at scale was narrowly defined by industrial-age technologies and that policymakers have ironically calcified education in the name of unbridled innovation. Highlighting the interplay of theory, technology, and policy in creating an equitable large-scale assessment system, in this historical and theoretical review, we explore the history of large-scale assessment technologies and theories, new software capabilities, and policy supports for equitable, large-scale assessment. We then focus on the role of large-scale writing assessment practices as a case in point for how large-scale assessment might look different today, providing federal and state officials the accountability they demand while also being more equitable and sustainable at the local level.

Thorndike Versus Dewey: A Battle of Epistemologies

Current theoretical tensions in large-scale assessment in the United States mirror historical tensions from almost a century ago. The history of American education has been framed as an early 20th century battle between Edward Thorndike and John Dewey, in which Thorndike wins (Lagemann, 1989; Tomlinson, 1997). Despite both scholars identifying with progressivism, their pedagogical stances were fundamentally divided. Thorndike believed learning was *epistemologically predetermined*, whereas Dewey viewed learning as *epistemologically negotiated*. In epistemological negotiation, learning is process-based, and meaning is socially constructed in real-world settings via systematic problem-solving approaches. Epistemological negotiation is a key feature of sociocultural learning theories that situate the learner and the learning process within a unique social and cultural context (Behizadeh & Engelhard, 2011; Perry, 2012; Prior, 2006; Street & Lefstein, 2007). This context may include different languages (Behizadeh, 2014; Shohamy, 2013), different modes of communication (McGrail & McGrail, 2013; New London Group, 2000), and different cultural backgrounds (Irvine & Armento, 2001; Ladson-Billings, 2009; Paris, 2012). Sociocultural approaches to teaching and learning and increased educational equity can be linked by epistemological negotiation. In this negotiation, students bring their own knowledge, experiences, languages, and identities as valued resources for interpreting, critiquing, and making sense of texts and perspectives. Dewey (1938) explained this negotiation of meaning, stating:

As an individual passes from one situation to another, his world, his environment, expands or contracts. He does not find himself living in another world but in a different part or aspect of one and the same world. What he has learned in the way of knowledge and skill in one situation becomes an instrument of understanding and dealing effectively with the situations which [*sic*] follow. The process goes on as long as life and learning continue. (p. 42)

What Dewey describes as an individual's responsiveness to one's world passing from "one situation to another" is fundamental to sociocultural learning. Sociocultural learning is the foundation for culturally relevant and sustaining pedagogy (Ladson-Billings, 2009; Paris, 2012), problem-posing education (Freire, 1970/2000), project- and problem-based learning (Markham, Larmer, & Ravitz, 2003), social justice education (Adams, 2016), and many more pedagogical approaches that have been lauded by the educational research community as a way to realize equitable education. In all of these pedagogies, learners actively make sense of the world and create new knowledge in relation to themselves and their lived experiences, requiring a process of epistemological negotiation.

In contrast to epistemological negotiation, epistemological predeterminism is a view of learning that focuses on gaining particular skills and knowledge without considering process or context. Epistemological predeterminism positions learning as individualistic, decontextualized, and product-based, where learning can be empirically sampled with psychometric instruments. In epistemologically predeterministic assessment practices, the "right" answer precedes student inquiry and defies critique—a view of knowledge creation that contrasts with Dewey's view. The correctness of learning is often evaluated via simple scoring mechanisms, such as multiple choice questions or externally applied rubrics to timed essays. This view of learning corresponds with the banking method of education that treats students as empty vessels to be filled with knowledge (Freire, 1970/2000), rather than as active producers of knowledge.

The differing epistemologies of Dewey and Thorndike, which we simplify for heuristic purposes, relate to the *cult of efficiency* (Holt, 1994; Yengo, 1964), a movement in education that drew upon business and management practices, as well as emerging methods of intellectual and mental achievement tests. According to Yengo:

It is quite apparent that by 1928 John Dewey was disturbed and dissatisfied with the kind of science of education that was developing. The efficiency movement was no longer rendering more effective the mechanical aspects of the school as it appeared in 1917. Rather, a complete school of educational thought and practice was being established upon the principles of educational measurement and scientific management and was dignifying itself as being the science of education. (p. 40)

Dewey was concerned that the cult of efficiency only measured existing knowledge and limited the creation of new knowledge (Yengo, 1964). Thorndike, however, embraced the efficiency movement and believed that creating efficient, objective measures for learning would improve U.S. education (Clifford, 1984).

The cult of efficiency was necessarily concerned with large-scale measures for learning. However, when taking a measure to scale, it is necessary to employ technology. If Thorndike “won” the battle for U.S. education, it is, at least in part, because his epistemologically predetermined, product-driven view of learning aligned easily with technologies or “conditioning elements” (Levin, 1956, p. 124) available at the time. For instance, intelligence testing relied on psychologists creating “expert” questions and answers that required only mass paper-based printing technologies, which were in abundance in the early 20th century (Giordano, 2007). As the introduction of basic computing devices in the mid-20th century allowed for more efficient forms of epistemologically predetermined testing instruments, their use continued to spread.

If Dewey “lost,” it is partially due to the fact that available technologies in the 1910s and 1920s could not support taking his process-based sociocultural model to scale. The steely analogue machines of the early modern era could not handle Dewey’s pedagogical approach, which was one of epistemological negotiation and was fundamentally communicative in nature. Taking Deweyan sociocultural processes of negotiation and knowledge creation to scale did not translate easily at a time when communication technologies like telegraphs and telephones were just slowly taking root, and the country’s infrastructure was racing to catch up.

It was this key epistemological difference between Thorndike and Dewey that led one historian to claim that “Thorndike is the historical starting point for any study or analysis of modern educational technology” (Saettler, 2004, p. 56). Specifically, Saettler lists Thorndike’s scientific and technological achievements, including the performance of “extensive scientific studies of mental tests, scales of achievement, and textbooks” (p. 56). It is worth noting that Saettler refers to two kinds of technology. Explicitly, he refers to tests, scales, and textbooks as technologies. Implicitly, such technological instruments are themselves the products of industrial technologies like the printing press. Dewey, according to Saettler, was “destined for disappointment” because he “had too hastily destroyed the traditional instructional pattern without replacing it with something better” (p. 58). Saettler’s “something better” refers to Dewey’s inability to harness technology for the purpose of taking work to scale as Thorndike did. As the public school system continued to solidify in the new century, the need to assess students at scale only grew, and Thorndike’s testing solution directly affected the trajectory of American education (Giordano, 2007).

In the next sections, we explore policies and practices of large-scale assessment as they relate to available technologies, in order to suggest possibilities for aligning large-scale assessment with Dewey’s view. However, before turning to policy, it is necessary to consider that the meaning of *technology* today is not what it was a century ago. Today, to refer to technology is to refer to software.

How Technology Has Changed, and Why Dewey Needed Software

In the latter part of the 20th century, the history of educational technology underwent a revolutionary shift, as personal computers and the Internet became popularized. From 1995 to 2014, Internet access amongst American adults increased dramatically as virtually all aspects of society adapted to online and digital technologies (Fox & Rainie, 2014). In an effort to theorize and examine this digital revolution, media theorists and scholars from several fields recommended a focus not on *technology* but on *software*. Although software can be a complex term to pin down (Frabetti, 2015), for present purposes, software refers not to digital products, applications, and services per se, but, rather, to software as interwoven human and programming languages executed by computers that make digital products, applications, and services possible. Although the development of software dates back to at least the 1950s (Tukey, 1958), widespread access to software via user-friendly devices and interfaces is much more recent (Manovich, 2001, 2013). Unlike older analogue technologies (like the industrial printing press), digital technologies, or more precisely *software-powered technologies* (Lynch, 2015b), have an active and agentic quality (Berry, 2011). They can be programmed to behave in ways that previous technologies could not. For example, consider the spellcheck functionality in word processing applications. When a user writes a word that does not exist in the computer's dictionary, the application places a red squiggly line beneath the offending word. The agentic behavior occurs because the application is powered by software written to identify users' incorrect spellings, using standard dictionaries that can be customized over time, and immediately flags perceived errors. Such errors can even be automatically corrected. This is a far cry from non-digital technologies that are used to correct spelling, which consist of paper, ink, and a great deal of human intervention.

In addition to having an agentic quality, today's software-powered technologies have made complex social communication ubiquitously available. Instead of the film projectors and room-sized computers of the past, we have software-powered technologies, such as interactive whiteboards, mobile phones and tablets, sophisticated information systems, and data dashboards. Software-powered technologies can now support the process of learning, not just the product. It is possible for educators to formatively assess collaborative learning, not just the accuracy of answers to multiple-choice questions. However, many of the most fervent and sincere calls for using software-powered technologies to reform education today—from policymakers, philanthropists, companies, and educators—continue to perpetuate the century-old paradigm that positions technology in epistemologically predetermined ways. This positioning contradicts intentions to prepare students for a 21st century world that requires a more Deweyan form of epistemological negotiation inherent in collaborative, critical, and creative learning environments (Darling-Hammond, 2010; Trilling & Fadel, 2009; Wagner, 2012).

Whereas the analogue technologies of a century ago demanded predetermined assessment practices (i.e., paper-based exams requiring single correct answers), today's software-powered technologies can deftly support negotiated learning at scale, which we can see when classrooms on different sides of the world video conference with each other to collaborate on projects. These new software-powered technologies are both ubiquitous and at times invisible. In terms of ubiquity, the use of new devices, software applications, and web-based tools has been clearly spreading in schools, with some high-profile initiatives in urban districts like Los Angeles, Newark, and New York City (Lynch, 2015a; Selwyn, 2014). At the same time, software-powered technologies are positioned less explicitly as tools to address district, state, and federal policymakers' demands for quantitative data about students and teachers, which has led to a mostly hidden intricate world of information systems (Lynch, 2013, 2016; Ravitch, 2013; Taubman, 2009).

But there is a catch. Although software excels at facilitating negotiated interpersonal communication, it can also be used to reinforce predeterminism. One observes predeterminism in assessment as multiple-choice examinations and so-called personalized-learning products, which are increasingly being offered via computer labs. However, although the ontology of software is capable of aligning with sociocultural models of learning, this does not imply that this is how software is positioned by policymakers or by private companies that develop products for districts. Rather, researchers of educational technology frequently conclude that the kinds of products used in schools perpetuate an approach to instruction and assessment that values single right answers, skill-and-drill activities, and low-level cognitive engagement (Buckingham, 2008; Cuban, 2001; Meier, 2005; Philip & Garcia, 2013; Selwyn, 2014). As the nature and ubiquity of technology has grown over the last two decades, the paradigm that frames how technology is actually used in instruction and assessment has hardly changed at all. It is important to note that a paradigm shift to align assessment with Deweyan approaches to learning requires a concurrent policy shift that reduces the stakes associated with large-scale assessment and includes more stakeholders in the assessment process.

A Policy Framework for Equitable Large-Scale Assessment Systems

These theoretical origins of large-scale assessment and available industrial technologies have ingrained certain assumptions about how we assess students in public education. Although assessment practices have been the subject of critique for decades, seldom do critics fixate on the role that technologies play in the implementation of particular assessment approaches and the ideologies that undergird the assessment systems themselves (Giordano, 2007; Taubman, 2009). Based on our analyses thus far, theory and technology are available to support large-scale sociocultural assessment. However, in addition to sound theory and technology, there is a third requirement for socioculturally sound assessment: enacting policy supports for epistemological negotiation in assessment. Building on past theory and research on assessment policy, we discuss two critical policies for sociocultural assessment that allow for epistemological negotiation: reducing stakes and negotiated control.

A Call for Reducing the Stakes

High stakes means that assessment results are being used to make critical decisions about students or teachers, such as student retention, teacher evaluations, and/or merit pay. A large body of literature has documented the negative effects of high-stakes, standardized tests on student learning (Amrein & Berliner, 2003; Au, 2007; Au & Gourd, 2013; Bauer & Garcia, 2002; Haertel, Moss, Pullin, & Gee, 2008; Ketter & Pool, 2001; Madaus, 1994; Mintrop & Sunderman, 2009; Nichols, Berliner, & Noddings, 2007), calling into question the consequential validity of these assessment practices. Consequential validity is determined by the impact of using a particular assessment practice in actual settings (Behizadeh & Engelhard, 2015). Mintrop and Sunderman (2009) thoroughly reviewed studies examining sanctions-driven accountability practices under No Child Left Behind, concluding that there were strong links between sanctions and narrowing the curriculum to test-taking preparation. They further concluded that these practices contributed to teachers focusing more on students nearing the proficiency category and less on those far below the proficiency category.

Even after No Child Left Behind, the stakes remained high. Despite the Obama administration's support of innovative technology use for non-assessment purposes, such as 21st century skill development (e.g., Computer Science for All) and whole-school reform (e.g., Digital Promise, The League of Innovative Schools), the Race to the Top program continued to forcefully promote high-stakes, large-scale assessment practices (Lynch, 2015b), particularly through assessment requirements for states that adopted the Common Core State Standards (Common Core State Standards Initiative, 2017). The more recent Every Student Succeeds Act (ESSA, 2015) has not done much to reduce the frequency of high-stakes, large-scale assessment. ESSA established minimum requirements for annual testing: annual reading and math testing for students in grades 3–8; science assessment once during elementary, middle, and high school; and at least one interdisciplinary assessment in high school. Additionally, ESSA mandated that accountability systems be in place by the 2017–2018 school year (Sharp, 2016).

However, due to ESSA's redistribution of authority from federal to state governments, large-scale assessment policy in the United States is currently in flux (Behizadeh & Pang, 2016). Thus, there is an opportunity to interrupt large-scale testing's long history of objectifying students and teachers, from the use of IQ exams to classify and track students at the turn of the century, through Chapter I testing provisions in the 1960s (Madaus, 1994), to more recent waves of high-stakes testing. Despite policymakers' best intentions, the past century of U.S. education is one in which large-scale assessment practices have perpetuated rather than mitigated inequity in our schools, in part due to the corrupting influence of high stakes. Although software-powered technologies can now support epistemologically negotiated large-scale assessment, without reducing stakes, assessment systems will risk being corrupted just as past systems have been.

A Call for Negotiated Control

Once stakes have been reduced and students and teachers can focus on meaningful negotiation without fear of sanctions, we propose extending this negotiation process to those stakeholders—primarily educators and policymakers—who use assessment data to make decisions. A key question arises: How can we ensure that epistemologically negotiated assessment practices are legible (Scott, 1998) to stakeholders when the process of reducing complex and complicated phenomena is often destructive (Foucault, 1975/1995; Stimson, 2000) at the local level? First, we have to reframe how we think about reliability. Reliability need not be defined as a quantitative measure of the correlation among independent scores. Instead, reliability may be reframed as *situated*, such as through local consensus of qualified evaluators, including the peer-review process used in the academy (Moss, 1994). Second, extending this peer-review process to K–12 education, we envision teachers, instructional coaches, and potentially students participating in the evaluation process and providing input. Although epistemologically negotiated assessment practices that are supported by software can reduce learning to numbers and provide an analytic shortcut to representing student achievement—thus increasing legibility of outcomes for policymakers—these numbers will not be psychometrically reliable. Rather, they are *hermeneutically* reliable (Moss, 1994; Petruzzi, 2008), meaning that they represent a consensus derived from negotiation among qualified experts, primarily teachers.

We understand that this reframing of the assessment process requires extended timelines for discussion and negotiation of meaning (Freedman, 1993), yet the diversity of opinion generated through meaningful analysis of student work offers a richness of educational possibilities and opportunities for epistemological negotiation. When applied to the exemplar of writing assessment, negotiated reliability can be achieved through methods used in the past by Kentucky and Vermont for large-scale portfolio assessment, and used currently by the National Writing Project (2017), which consists of panels of local stakeholders reaching a consensus on writing quality.

We believe that advances in software-powered technologies can support large-scale epistemologically negotiated assessment systems so that rich processes of learning are accessible to all stakeholders, including policymakers. Also, epistemologically negotiated assessment systems that are supported by software can provide more legible assessment data for policymakers by allowing them to generate numerical representations of student learning without losing access to the negotiated processes of meaning-making that isolated numbers often obscure.

In addition to rebalancing the power of teachers and policymakers, negotiated control requires that the influence of profit-driven corporations be minimized. As Picciano and Spring (2013) make clear in their structural analysis of the relationship between public educational agencies and private enterprise, public agencies have grown accustomed to outsourcing the expert guidance and technical logistics needed for large-scale assessment. Over time, the direction of assessment policy has tended to defer to whatever products companies can quickly provide, rather than to current educational theory and research on learning and assessment, hence supporting Thorndike's alignment with the cult of efficiency. The result is that state and district officials who lack the expertise to design

and implement large-scale assessment systems become dependent on corporate partners with off-the-shelf products.

Picciano and Spring (2013) refer to the “education-industrial complex,” which they define as “a series of networks and alliances that strive to influence the creation or modification of policies at all levels of government consistent with views and ideas that support extensive uses of technology and are profitable for its members” (p. 8). The result is a mutually beneficial deference between policymakers and companies, where the policies created are circumscribed by the technologies that companies can provide at scale. In particular, it is especially profitable for companies to sell those products and systems, which they already have in place. This is especially true of software-powered technologies, in which profitability only occurs once they are developed and tested. After that, an increase in users only further increases company profits. If software is capable of more, its potential is seldom realized not only because the policies do not promote more socioculturally responsive uses, but also because the software-powered products that have already been developed are inherently more profitable than investing in developing new ones. Thus, corporations should be given much less control and educators much more.

Other scholars have proposed assessment procedures that connect to our idea of negotiated control. For example, in writing assessment, Huot (2002) reviewed proposals of multiple scholars who advocated for evaluation of writing using locally agreed-upon standards with some degree of external review or oversight (Allen, 1995; Berlak, 1992; Moss, 1994). Referring to Allen’s (1995) study, Huot noted, “His use of electronic communication points out the vast potential the Internet and the Web have in providing the linkage and access necessary to connect site-based, locally controlled assessment programs from various locations” (p. 106). Notice that the use of technology that Huot mentions is one in which software facilitates communication, an epistemological negotiation. In another proposal, Petruzzi (2008) suggested that “accountability reports to stakeholders could be ‘interpretive summaries,’ that use rhetorical reasoning rather than quantitative measures of objective data” (p. 239). We believe that software-powered technologies can help facilitate these proposals for negotiated control and that current federal policy under the Every Student Succeeds Act may support our proposals.

Linking theory, technology, and policy, a major shift in large-scale assessment practices is possible with the dominance of sociocultural theory, the emergence of software-powered technologies, and the passage of power-sharing federal legislation that may lead to reduced stakes and increased local participation. In the next section, we examine the intersections of theory, technology, and policy through a case study of large-scale writing assessment. We close with examples of what negotiated control of writing assessment that is theoretically and technologically aligned with epistemological negotiation could look like.

An Illustrative Case: Large-Scale Writing Assessment

A comparison between two kinds of writing assessment technologies will help illustrate the problems and possibilities before us. Writing assessment offers an appropriate site of study for exploring large-scale assessment systems, in part due to the Common Core State Standards' unique emphasis on literacy practices across the content areas. In fact, writing argument was one of the first pedagogical emphases to emerge from the initial Common Core rollout (Common Core State Standards Initiative, 2012). We consider portfolio assessment and on-demand essays assessment, also known as direct writing assessment (DWA), as realized in paper-based and software-powered versions in the United States.

Portfolio Assessment and E-Portfolios

Portfolios were one of the major alternative assessment practices used in the 1990s for U.S. writing assessment. Bauer and Garcia (2002) summarized four major points of alternative assessment methods such as portfolios: Students and teachers should collaborate to select evidence of learning, multiple meanings should be honored, varied responses should be valued, and a student's individual approach should be recognized. Based on these features, portfolio assessment is more likely to align with epistemological negotiation.

According to Madaus (1994), equity for diverse students was one rationale for a switch from multiple choice assessment and DWA to so-called alternative assessment. One reason for increased equity is the close alignment between a sociocultural construct of writing and portfolio assessment, which then encourages culturally relevant and engaging instructional practices (Behizadeh, 2014; Gordon, Engelhard, Gabrielson, & Bernknopf, 1996; Murphy & Yancey, 2008). Bauer and Garcia (2002) explained:

The equity claim is based on the hope that students will have better access to instruction that meets their literacy needs (educational equity), to literacy assessment tools that reveal what they can and cannot do (assessment equity), and to greater voice in their literacy development (empowerment equity). (p. 464)

In these statements, the authors posited that positive washback—the effects of an assessment on instruction—from alternative assessment, such as portfolios to instruction, is a major rationale for use, a finding echoed by Murphy and Yancey (2008) in their review of writing assessment research.

Unfortunately, despite claims for increased equity, the use of portfolios for large-scale assessment has decreased dramatically in the United States, in large part due to concerns with reliability of scores (Koretz, Stecher, Klein, & McCaffrey, 1994; Murphy & Yancey, 2008). This issue directly connects to negotiated control. For example, in Vermont and Kentucky, the state governments enacted statewide portfolio assessment systems in the 1990s, yielding multiple positive results, including increased writing instruction, more attention to the process of writing, and increased quality of student writing (Callahan, 1999; Gearhart & Wolf, 1994; Koretz et al., 1994). Yet in both states, external review of the systems focused on low quantitative reliability statistics, meaning

there was a low correlation between independent scores by raters (Evaluation Center of Western Michigan University, 1995; Koretz et al., 1994). This concern overshadowed the positive effects of portfolio assessment, suggesting that the primary purpose of the assessment was as an accountability system serving policymakers rather than as a teaching and learning tool serving teachers and students. This conflict of purposes is embedded in a theoretical conflict between sociocultural and skill-oriented views of writing, yet is also a technological limitation in that past portfolios did not generate legible data on student achievement for stakeholders. In other words, the numbers generated by portfolio assessment could not easily be interpreted out of context. Yet recent advances in software-powered technologies paired with a heuristic conception of reliability could potentially resolve these issues by providing rich nuanced qualitative and quantitative data that can serve multiple purposes. We argue that these technologies can produce data that is legible to stakeholders and meaningful to educators and students.

More recently, in addition to paper-based portfolio systems, software-powered writing portfolios have become more heavily researched in schools and districts, although not at a state-wide level (Cambridge, Cambridge, & Yancey, 2009; Lorenzo & Ittelson, 2005; Yancey, 2009). Hamp-Lyons (2002) noted that computer-based portfolio assessment not only provides autonomy to writers, but also provides “multiple pathways for writers through the many pitfalls of tests” (p. 11). More importantly, Hamp-Lyons added, “The possibilities for computer-based writing assessment are not limited by the computer software but in [*sic*] what ‘the system’ (educational, financial, political) will allow” (p. 11). In historical practice, it seems e-portfolios have often served as digital filing cabinets. Advances in software-powered technologies, however, make it more possible than ever for a school to use digital writing portfolios as a way for students to submit multiple drafts of a composition—potentially even multimodal pieces that include video and visuals along with written text—and receive feedback from various family or community members active in these students’ lives.

A policy of negotiated control may be applied to writing assessment by, for example, inviting parents and community members to read and submit feedback on selected student writing that is digitally shared through an e-portfolio. This feedback could then be accessible to the student along with self-evaluations, reflections, and teacher and peer feedback. Connecting policy to technology, we envision software being employed to render this broad spectrum of meaningful feedback in visual displays that do more than reduce nuanced feedback and diverse perspectives to a number. Important for legibility, evaluators could still assign numbers to different portfolios in order to facilitate a quick review of a large data set by district, state, and federal policymakers. However, these numbers would be locally determined using local consensus, and stakeholders (again, representing a wide range of individuals) would have access to the full body of a student’s work, which can then be analyzed and discussed to generate better supports for teachers and students.

On-Demand Essay Assessment and Automated Essay Scoring

Writing as a contextualized sociocultural practice is firmly established as a dominant theory in the writing-research community, and alternative assessment such as portfolios had a brief moment of ascendance in the 1990s. Yet on-demand essay assessment, also referred to as DWA, has been the dominant form of writing assessment in the United States for many decades (Behizadeh & Pang, 2016; Hamp-Lyons, 2002). In on-demand essay assessment, students are given a random prompt and a set amount of time, and essays are externally scored at a testing center on a rubric that is either holistic (i.e., one number represents an overall impression of the piece) or analytic (i.e., a set of numbers is used to represent different aspects of writing) (Hillocks, 2002; Huot, 2002). Although they admit that the practice is superior to multiple-choice tests, critics of on-demand essay assessment practices articulate two main issues. The first issue is lack of authenticity/validity or, as it is sometimes called in the field of educational measurement, ecological validity. The second issue is washback, referring to the effects of on-demand essay assessment practices on teaching (Hamp-Lyons, 2002).

In terms of negotiated control, the entire on-demand essay assessment process is outsourced, either to private assessment companies or to teachers who have no relationship with students, thus limiting teachers' control of the system. Behizadeh and Pang (2016) found that 46 out of 50 states used external scoring processes for writing assessment with no involvement of the classroom teacher. In terms of instructional benefits, on-demand essay assessment does not purport to benefit instruction in the same way as portfolio assessment. Because on-demand essay assessments are a one-time, impromptu, timed test, the focus is not on the process of writing, thoughtful revision, and reflection. Rather, an on-demand essay assessment is separated from the curricular, instructional, and sociocultural context of a student's learning. As a result, it captures only what an individual student can produce on a random topic in a limited amount of time, aligning this method more with epistemological predeterminism than epistemological negotiation.

Regarding technology, the software-powered counterpart to on-demand essay assessment is automated essay scoring (AES). AES received much media attention recently when reports emerged that AES could quickly and cheaply score large numbers of student essays with accuracy rates comparable to human scorers (Shermis, 2014; Shermis & Burstein, 2003), an affordance highly valued by policymakers with limited funds at their disposal. In AES, software is used to read student writing, develop a quantitative and categorical profile of the written work based on a bank of human-scored writing samples, and render a numerical score and pre-written qualitative feedback to students. Referring to a study by Shermis (2014) on the performance of nine AES vendors in the United States, Perelman (2014) critiqued Shermis' claim that AES can reliably replace human scorers in high-stakes assessment contexts. In Perelman's analysis, he found that "the study's raw data provide clear and irrefutable evidence that Automated Essay Scoring engines grossly and consistently over-privilege essay length in computing student writing scores" (p. 104). This conflation of essay length with writing skill could have enormous negative washback to writing instruction, including encouraging "bloated and vapid prose" (p. 110). Other automated scoring processes that rely on more complicated analyses, such as latent semantic analysis (Landauer,

McNamara, Dennis, & Kintsch, 2013), may offer an approach for automated scoring that, to some extent, takes meaning into account. However, these scoring processes are essentially mathematical programs analyzing a writer's text. In AES, software is generally being harnessed to scale epistemological predeterminism, functioning as little more than a linguistic bean counter.

Righting Theory and Technology: Negotiation Through Software-Powered Technologies

The two cases of large-scale writing assessment above offer timely and relevant sites of study to better understand how the theoretical stances of epistemological predeterminism and epistemological negotiation operate in education. Both epistemological predeterminism and negotiation honor the pragmatic need to identify the essential knowledge and skills that students are expected to learn in advance. As a result, neither stance disaffirms the use of quality learning standards that are created by expert educators at the state and local level. The heart of the distinction between the two concepts lies in what they accept as sufficient evidence of student learning.

For epistemological negotiation, in the case of portfolio assessment and e-portfolios, students are assessed based on their ability to negotiate meaning over a sustained writing process, and the writing process includes examination of audience, genre, and purpose. What students express is not a single decontextualized response, but evidence from multiple kinds of writing assignments over time. The process of portfolio assessment includes frequent forms of low-stakes feedback on drafts where students can take risks because doing so is a valued element of negotiation. The use of technology to support portfolio assessment, which we refer to as e-portfolios, can be used to accentuate the negotiative quality of this assessment practice. When software-powered technology is used to improve the efficiency of students' writing processes, including the receipt of frequent low-stakes feedback, its use supports epistemological negotiation.

In contrast, on-demand essay assessment and AES place lesser value on the process and greater value on the final product as assessed through rubrics. The use of rubrics embodies epistemological predeterminism when used in the rapid assessment of on-demand writing, as is so often the case in large-scale writing assessment. For instance, Hillocks (2002) recounts stories of on-demand essay assessment raters who felt that they needed to undermine their true evaluation of student writing in order to conform to the rigid guidelines imposed by the testing center. In this way, the human rater is acting as little more than a machine. In such cases, the rigidity of the rubric's language and the numerical values associated with such language creates a situation where students cannot be assessed on their ability to negotiate meaning but rather must comply with narrowly defined predetermined expectations in a single final product.

AES can be used to make predeterministic assessment more efficient in terms of time and money. Software can be programmed to read students' typed essays, compare them against a bank of pre-scored essays, and algorithmically generate a score. It is important

that the prompts offered to students are consistent in order for such assessment systems to work. That is, students receive identical questions, and their answers must be similar. Epistemological predeterminism is operationalized not only in the standardization of the prompts, but also through the rubrics used and eventually through the algorithms encoded into software (Lynch, 2015a).

When one considers the emphasis today on 21st century skills for college and career readiness, the above examples illuminate how far current assessment practices are from assessing such skills. In the past, on-demand essay assessment and AES gained prominence because they were perceived to be the best technologies available to assess writing at scale. Such assessment practices promoted epistemologically predeterministic approaches to education, albeit to a lesser degree than multiple choice tests. On-demand essay assessment perpetuates a Thorndikean assessment philosophy that can be executed at scale, rather than a Deweyan ideal that has historically defied operationalization. Nevertheless, on-demand essay assessment was the best assessment technology available at the time.

Today, we have available software-powered technologies that can facilitate the assessment of epistemological negotiation at scale. In the case of digital writing portfolios, software could be positioned to facilitate communication between different stakeholders in a manner that reinforces the integrity of the writing and learning processes as one of negotiated meaning-making. We have the opportunity to finally break from 19th century models of learning and assessment that we have inherited and perpetuated for over a century. Based on our analyses, we suggest software-powered e-portfolios as a future direction for large-scale writing assessment and potentially other content areas as well.

Promising Software-Powered Assessment Systems

Specifically looking at portfolios or rich tasks that utilize software-powered technologies, such as in the New Basics project (Klenowski, 2011; Lingard & McGregor, 2014), we highlight a few examples of promising assessment systems. One is the Scholar system developed by Bill Cope (Cope, 2013; Cope & Kalantzis, 2017; see <http://info.cgscholar.com/>) that includes a community space similar to popular social media sites; a creator space for writers to create work and review others' work; a publisher function that has revising and publishing tools that can be applied to multiple products; and an analytics function for data-mining and assessing the process and products of writing. Although it is currently being used for college writing, Scholar could be a possibility for large-scale portfolio assessment in elementary through high school and in multiple content areas.

A second promising example is the work of Patrick Griffin on the Assessment and Teaching of 21st Century Skills project (AT21CS; Assessment and Teaching of 21st Century Skills, 2012; Griffin, 2015), which is currently in a final development stage. According to Griffin (2015), this project will use technology in innovative ways that support collaboration, creativity, criticality, and problem-solving. Additionally, Griffin states that AT21CS will provide “detailed time-stamped data capturing the activities of collaborators” and “log stream and chat stream data for modeling and evaluating student activity” (p. ix). Yet the success of these new software-powered technologies when used

in large-scale K–12 school settings will depend on the quality of their use. Without negotiated control and lowered stakes, these systems could become yet another top-down reform that does not allow for epistemological negotiation for students. Therefore, any promising technology that allows for epistemological negotiation needs to be paired with negotiated control and stakes reduction.

In addition to these systems that are designed for widespread use, as briefly referenced earlier, there are also simpler software-powered technologies that could be used by teams of teachers or individual schools to design their own negotiated assessment system. These include cloud services like Google Docs (Dunn, Luke, & Nassar, 2013; Godwin-Jones, 2008) and manuscript review systems, such as ScholarOne (see <http://scholarone.com/products/manuscript/>), that allow for collaborative writing and review, which could include writing across content areas, such as descriptions of problem-solving approaches in math. Districts and states could then use a sampling system to review student work and provide feedback to professional development to strengthen these homegrown systems and share outstanding examples with other educators.

Although the bulk of research on large-scale portfolio assessment centers on writing, scholarship indicates similar promise for supporting epistemological negotiation in science (Chang & Tseng, 2009; Gunay & Ogan-Bekiroglu, 2014; Valdez, 2001), social studies (Ugodulunwa & Wakjissa, 2015), and math (Helton, 1994; National Council of Teachers of Mathematics, 2016). For example, in their experimental study of physics portfolios, Gunay and Ogan-Bekiroglu concluded that portfolio assessment was not only effective for assessing student growth in physics, but that it also helped students learn physics. However, there appeared to be scant research on e-portfolios in these other content areas, and future research is needed to design and study the effects of large-scale e-portfolio assessment in multiple content areas and in interdisciplinary ways that combine content areas.

Using already available and easily accessible tools, a next step in our research is to develop and pilot an alternative system that builds on past national and international writing assessment systems (Calfee & Perfumo, 1996; Callahan, 1999; Cambridge et al., 2009; Cumming & Maxwell, 2004; Gearhart & Wolf, 1994; Koretz et al., 1994; Lorenzo & Ittelson, 2005; Mills & McGregor, 2016) and that meets our recommendations for socioculturally sound and technologically powerful assessment. We hope other researchers in science, math, social studies, foreign language, and other content areas will take up this charge to create e-portfolio assessment systems that are powerful supports for teaching and learning, and persuasive and legible to district and state policymakers.

Conclusion

Twenty years ago, equity concerns were a primary force for revising assessment (Madaus, 1994), and equity concerns should still be the guiding force behind assessment reform. The disparate impact of high-stakes testing on minoritized groups is a civil rights

issue, one that this nation must attend to if we are to live up to our democratic ideals. As Darling-Hammond (2010) has argued, “low-income students of color have been the primary victims of high-stakes testing policies” (p. 74). If there is reason to hope, and we believe there is, it is because the kinds of pedagogies and assessment practices that software-powered technologies can support at scale align with calls for 21st century skills, the philosophies of Dewey, and the work of other researchers who have built upon Dewey’s research. Epistemological negotiation of learning supported by software, freed by reduced stakes, and supported by negotiated control, may be a way to address the continued inequities of current testing practices.

Although it is critical to recognize students’ lived realities through open-ended assessment structures, and to allow students and other stakeholders to participate in the assessment process, it is important that these changes, and all of the suggested assessment reforms for which we advocate, be paired with social supports for high-poverty students and families and investments in high-poverty communities (Darling-Hammond, 2010; Noguera, 2003). Although we center our work in this piece on addressing current inequities of U.S. schooling from within, we cannot emphasize enough the importance of federal and state investments in alleviating conditions associated with poverty. Until the United States can narrow the gap between the rich and the poor, closing the opportunity gap between the high- and low-performing will continue to be a very difficult if not insurmountable task.

Reconfiguring responsibilities among administrators, teachers, district and state officials, students, and community members, through assessment policies built on the concept of negotiated control, is just one positive step that must be paired with a holistic approach to supporting high-poverty communities. As Noguera (2003) stressed in his research on urban schools, “we must stop pretending that we can avoid confronting and addressing urban conditions as we try to devise strategies for improving urban schools” (p. 144). As educational researchers, we need to be careful not to get stuck flying at too high of an “algorithmic altitude” (Lynch, 2015b, p. 94), where we forget that the real work of education reform is in having our feet on the ground and our actual bodies in spaces where we can engage with those whom we aim to serve.

In order for the United States to be a global leader, its public schools must provide more than basic skills education. We need all students to be prepared for life and work in the 21st century, which requires critical thinking, collaboration, and problem-solving, all of which are processes that are connected to negotiating meaning rather than regurgitating facts. Anyone can look up a fact in seconds using the affordances of various search engines and applications. Yet humanity faces a multitude of serious problems for which readily available answers do not exist and must be negotiated on a national or even global scale. These problems include continuing conflict in war-torn regions, and climate change due to an increasing population and unfettered consumption of fossil fuels. These complex issues will require the next generation to work together to balance the sometimes contradictory needs and desires of multiple nations, states, and communities. Shifting writing assessment practice from DWA to writing portfolios that rely on negotiated control of content and assessment processes could seem like a distant solution to the pressing problems of today. However, assessment can be a powerful tool for reshaping how students learn and teachers teach. It is high time to leverage software-

powered technologies to bring to scale the assessment practices, and consequently the pedagogies, that are more socioculturally responsive to our students and communities, and, as Dewey argued over a century ago, truer to the democratic ideals that should drive education. It is Dewey's century to win, computer in hand.

Author Biographies

Dr. Nadia Behizadeh is Assistant Professor of Adolescent Literacy in the Department of Middle and Secondary Education at Georgia State University. Her research interests include authentic and culturally sustaining writing instruction for youth, problem- and project-based literacy learning, sociocultural writing assessment, validity of writing assessment, and social justice teacher education. Dr. Behizadeh's current research project examines access to powerful writing pedagogy in urban, public schools, and is funded by the Spencer Foundation and the Conference on English Education.

Tom Liam Lynch is Assistant Professor of Educational Technology at Pace University in Manhattan. A former English teacher and school district official for the New York City Department of Education, Dr. Lynch led the implementation of a \$50 million online/blended learning program in over 100 schools called iLearnNYC. He also designed and guided the initial implementation of WeTeachNYC, a digital resource repository and learning environment for the city's approximately 80,000 teachers. Dr. Lynch researches the relationship between software theory, literacy, and education, most recently examining the relationship between K–12 computer science education and literacy studies.

References

- Adams, M. (2016). Pedagogical foundations for social justice education. In M. Adams & L. A. Bell (Eds.), *Teaching for diversity and social justice* (pp. 27–53). New York, NY: Routledge.
- Allen, M. (1995). Valuing differences: Portnet's first year. *Assessing Writing*, 2, 67–90. [https://doi.org/10.1016/1075-2935\(95\)90005-5](https://doi.org/10.1016/1075-2935(95)90005-5)
- Amrein, A. L., & Berliner, D. C. (2003). The effects of high-stakes testing on student motivation and learning. *Educational Leadership*, 60(5), 32–38.
- Assessment and Teaching of 21st Century Skills (2012). *About the project*. Retrieved from <http://www.atc21s.org/about.html>
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36, 258–267. <https://doi.org/10.3102/0013189X07306523>
- Au, W., & Gourd, K. (2013). Asinine assessment: Why high-stakes testing is bad for everyone, including English teachers. *English Journal*, 103(1), 14–19.
- Bauer, E. B., & Garcia, G. E. (2002). Lessons from a classroom teacher's use of alternative literacy assessment. *Research in the Teaching of English*, 36, 462–494.

- Behizadeh, N. (2014). Mitigating the dangers of a single story: Creating large-scale writing assessments aligned with sociocultural theory. *Educational Researcher, 43*, 125–136. <https://doi.org/10.3102/0013189X14529604>
- 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. <https://doi.org/10.1016/j.asw.2011.03.001>
- Behizadeh, N. & Engelhard, G. (2015). Involving diverse communities of practice to minimize unintended consequences of test-based accountability systems. *Measurement: Interdisciplinary Research and Perspectives, 13*, 26–30. <https://doi.org/10.1080/15366367.2015.1016320>
- Behizadeh, N., & Pang, M. E. (2016). Awaiting a new wave: The status of state writing assessment in the United States. *Assessing Writing, 29*, 25–41. <https://doi.org/10.1016/j.asw.2016.05.003>
- Berlak, H. (1992). The need for a new science of educational of assessment. In H. Berlak, F. M. Newmann, E. Adams, D. A. Archbald, T. Burgess, J. Raven, & T. A. Romberg (Eds.), *Toward a new science of educational testing and assessment* (pp. 1–23). Albany, NY: SUNY Press.
- Berry, D. (2011). *The philosophy of software: Code and mediation in the digital age*. New York, NY: Palgrave Macmillan. <https://doi.org/10.1057/9780230306479>
- Buckingham, D. (2008). *Beyond technology: Children's learning in the age of digital culture*. Malden, MA: Polity Press.
- Calfee, R., & Perfumo, P. (1996). A national survey of writing portfolio practice: What we learned and what it means. In R. Calfee & P. Perfumo (Eds.), *Writing portfolios in the classroom: Policy and practice, promise and peril* (pp. 63–81). Mahwah, NJ: Lawrence Erlbaum.
- Callahan, S. (1999). All done with the best of intentions: One Kentucky high school after six years of state portfolio tests. *Assessing Writing, 6*, 5–40. [https://doi.org/10.1016/S1075-2935\(99\)00005-7](https://doi.org/10.1016/S1075-2935(99)00005-7)
- Cambridge, D., Cambridge, B., & Yancey, K. (2009). *Electronic portfolios 2.0: Emergent research on implementation and impact*. Sterling, VA: Stylus.
- Chang, C., & Tseng, K. (2009). Use and performance of web-based portfolio assessment. *British Journal of Educational Testing, 40*, 358–370. <https://doi.org/10.1111/j.1467-8535.2008.00885.x>
- Clifford, G. J. (1984). *Edward L. Thorndike: The sane positivist*. Middletown, CT: Wesleyan University Press.
- 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
- Common Core State Standards Initiative (2017). *Development Process*. Retrieved from <http://www.corestandards.org/about-the-standards/development-process/>
- Cope, B. (2013). Towards a new learning: The ‘Scholar’ social knowledge workspace, in theory and practice. *E-Learning and Digital Media, 10*, 334–358. <https://doi.org/10.2304/elea.2013.10.4.332>

- Cope, W., & Kalantzis, M. (2017). *E-learning ecologies: Principles for new learning and assessment*. New York, NY: Routledge.
- Cuban, L. (2001). *Oversold & underused: Computers in the classroom*. Cambridge, MA: Harvard University Press.
- Cumming, J. J., & Maxwell, G. S. (2004). Assessment in Australian schools: Current practice and trends. *Assessment in Education: Principles, Policy & Practice, 11*, 89–108. <https://doi.org/10.1080/0969594042000209010>
- Darling-Hammond, L. (2010). *The flat world and education: How America's commitment to equity will determine our future*. New York, NY: Teachers College Press.
- Dewey, J. (1938). *Experience and education*. West Lafayette, IN: Kappa Delta Pi.
- Dunn, J. S., Luke, C., & Nassar, D. (2013). Valuing the resources of *infrastructure*: Beyond from-scratch and off-the-shelf technology options for electronic portfolio assessment in first-year writing. *Computers and Composition, 30*, 61–73. <https://doi.org/10.1016/j.compcom.2012.12.001>
- Evaluation Center of Western Michigan University. (1995). *An independent evaluation of the Kentucky Instructional Results Information System (KIRIS)*. Frankfort, KY: Kentucky Institute for Education Research. Retrieved from <http://files.eric.ed.gov/fulltext/ED394971.pdf>
- Every Student Succeeds Act, 20 U.S.C. § 1001 *et seq.* (2015).
- Foucault, M. (1995). *Discipline and punish: The birth of the prison*. (A. Sheridan, Trans.). New York, NY: Vintage Press. (Original work published 1975)
- Fox, S., & Rainie, L. (2014). *The web at 25 in the U.S.* Washington, DC: Pew Research Center. Retrieved from <http://www.pewinternet.org/2014/02/27/the-web-at-25-in-the-u-s/>
- Frabetti, F. (2015). *Software theory: A cultural and philosophical study*. New York, NY: Rowan & Littlefield Education.
- Freedman, S. W. (1993). Linking large-scale testing and classroom portfolio assessments of student writing. *Educational Assessment, 1*, 27–52. https://doi.org/10.1207/s15326977ea0101_3
- Freire, P. (2000). *Pedagogy of the oppressed* (M. B. Ramos, Trans.). New York, NY: Continuum. (Original work published 1970)
- Gearhart, M., & Wolf, S. (1994). Engaging teachers in assessment of their students' narrative writing: The role of subject matter knowledge. *Assessing Writing, 1*, 67–90. [https://doi.org/10.1016/1075-2935\(94\)90005-1](https://doi.org/10.1016/1075-2935(94)90005-1)
- Giordano, G. (2007). *How testing came to dominate American schools: A history of educational assessment*. New York, NY: Peter Lang.
- Godwin-Jones, R. (2008). Web-writing 2.0: Enabling, documenting, and assessing writing online. *Language Learning and Technology, 12*(2), 7–13.
- Gordon, B., Engelhard, G., Gabrielson, S., & Bernknopf, S. (1996). Conceptual issues in equating performance assessments: Lessons from writing assessment. *Journal of Research and Development in Education, 29*, 81–88.

- Griffin, P. (2015). Preface. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods & approach* (pp. v–ix). Dordrecht, Netherlands: Springer.
- Gunay, A., & Ogan-Bekiroglu, F. (2014). Impact of portfolio assessment on physics students' outcomes: Examination of learning and attitude. *Eurasia Journal of Mathematics, Science, and Technology Education*, *10*, 667–680.
- Haertel, E. H., Moss, P. A., Pullin, D. C., & Gee, J. P. (2008). Introduction. In P. A. Moss, D. C. Pullin, J. P. Gee, E. H. Haertel, & L. J. Young (Eds.), *Assessment, equity, and opportunity to learn* (pp. 1–16). Cambridge, England: Cambridge University Press.
- Hamp-Lyons, L. (2002). The scope of writing assessment. *Assessing Writing*, *8*, 5–16. [https://doi.org/10.1016/S1075-2935\(02\)00029-6](https://doi.org/10.1016/S1075-2935(02)00029-6)
- Helton, J. (1994). Appropriate strategies for improving math portfolios: A comparison of self-assessment versus peer conferencing. *UKERA Occasional Papers*. Retrieved from <http://files.eric.ed.gov/fulltext/ED414191.pdf>
- Hillocks, G. (2002). *The testing trap: How state writing assessments control learning*. New York, NY: Teachers College Press.
- Holt, M. (1994). Dewey and the “cult of efficiency”: Competing ideologies in collaborative pedagogies of the 1920s. *Journal of Advanced Composition*, *14*, 73–92.
- Huot, B. (2002). *(Re)Articulating writing assessment for teaching and learning*. Logan, UT: Logan State University Press.
- Irvine, J. J., & Armento, B. J. (2001). *Culturally responsive teaching: Lesson planning for elementary and middle grades*. New York, NY: McGraw Hill.
- Ketter, J., & Pool, J. (2001). Exploring the impact of a high-stakes direct writing assessment in two high school classrooms. *Research in the Teaching of English*, *35*, 344–393.
- Klenowski, V. (2011) Assessment for learning in the accountability era: Queensland, Australia. *Studies in Educational Evaluation*, *37*, 78–83. <https://doi.org/10.1016/j.stueduc.2011.03.003>
- 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. <https://dx.doi.org/10.1111/j.1745-3992.1994.tb00443.x>
- Ladson-Billings, G. (2009). *The dreamkeepers: Successful teachers of African American children* (2nd ed.). San Francisco, CA: Jossey-Bass.
- Lagemann, E. C. (1989). The plural worlds of educational research. *History of Education Quarterly*, *29*, 184–214. <https://doi.org/10.2307/368309>
- Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2013). *Handbook of latent semantic analysis*. New York, NY: Psychology Press.
- Levin, S. M. (1956). John Dewey's evaluation of technology. *The American Journal of Economics and Sociology*, *15*, 123–36. <https://doi.org/10.1111/j.1536-7150.1956.tb01090.x>
- Lingard, B., & McGregor, G. (2014). Two contrasting Australian curriculum responses to globalisation: What students should learn or become. *The Curriculum Journal*, *25*, 90–110. <https://doi.org/10.1080/09585176.2013.872048>

- Lorenzo, G., & Ittelson, J. (2005). Demonstrating and assessing student learning with e-portfolios. *Educause Learning Initiative*. Retrieved from <https://net.educause.edu/ir/library/pdf/eli3003.pdf>
- Lynch, T. L. (2013). The secretary and the software: On the need for integrating software analysis into educational spaces. In J. Gorlewski & B. Porfilio (Eds.), *Left behind in the race to the top* (pp. 129–142). Charlotte, NC: Information Age Publishing.
- Lynch, T. L. (2015a). Holy interfaces and the sanctity of software: A critical software analysis of rubrics as vehicles of conservative reform. In M. Tenam-Zemach & J. Flynn (Eds.), *A rubric nation: A reader on the utility and impact of rubrics in education* (pp. 125–141). Charlotte, NC: Information Age Publishing.
- Lynch, T. L. (2015b). *The hidden role of software in educational research: Policy to practice*. New York, NY: Routledge.
- Lynch, T. L. (2016). As blankness falls: How data shape the relationship among students, teachers, and the public. *English Journal*, 105(3), 98–100.
- Madaus, G. F. (1994). A technological and historical consideration of equity issues associated with proposals to change the nation's testing policy. *Harvard Educational Review*, 64, 76–95. <https://doi.org/10.17763/haer.64.1.4q87663r0j76rww1>
- Manovich, L. (2001). *The language of new media*. Cambridge, MA: MIT Press.
- Manovich, L. (2013). *Software takes command*. New York, NY: Bloomsbury Academic.
- Markham, T., Larmer, J., & Ravitz, J. L. (2003). *Project based learning handbook: A guide to standards-focused project based learning for middle and high school teachers* (2nd ed.). Novato, CA: Buck Institute for Education.
- McGrail, E., & McGrail, J.P. (2013). Preparing young writers for invoking and addressing today's interactive digital audiences. In K. E. Pytash & R. E. Ferdig (Eds.), *Exploring technology for writing and writing instruction* (pp. 54–76). Hershey, PA: IGI Global.
- Meier, E. B. (2005). Situating technology professional development in urban schools. *Journal of Educational Computing Research*, 32, 395–407. <https://doi.org/10.2190/5KCQ-5VKQ-380X-JYND>
- Mills, M., & McGregor, G. (2016). Learning not borrowing from the Queensland education system: Lessons on curricular, pedagogical and assessment reform. *The Curriculum Journal*, 27, 113–133. <https://doi.org/10.1080/09585176.2016.1147969>
- 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. <https://doi.org/10.3102/0013189x09339055>
- Moss, P. A. (1994). Can there be validity without reliability? *Educational Researcher*, 23, 5–12. <https://doi.org/10.3102/0013189X023002005>
- Murphy, S., & Yancey, K. B. (2008). Construct and consequence: Validity in writing assessment. In C. Bazerman (Ed.), *Handbook of research on writing: History, society, school, individual, text* (pp. 365–385). New York, NY: Routledge.

- National Council of Teachers of Mathematics (2016). *Large-scale mathematics assessments and high-stakes decisions: A position of the National Council of Teachers of Mathematics*. Retrieved from <http://www.nctm.org/Standards-and-Positions/Position-Statements/Large-Scale-Mathematics-Assessments-and-High-Stakes-Decisions/>
- National Writing Project. (2017). *About NWP*. Retrieved from <https://www.nwp.org/cs/public/print/doc/about.csp>
- New London Group. (2000). A pedagogy of multiliteracies: Designing social futures. In B. Cope & M. Kalantzis (Eds.), *Multiliteracies: Literacy learning and the design of social futures* (pp. 9–42). New York, NY: Routledge.
- Nichols, S. L., Berliner, D. C., & Noddings, N. (2007). *How high-stakes testing corrupts America's schools*. Cambridge, MA: Harvard Education Press.
- Noguera, P. A. (2003). *City schools and the American dream: Reclaiming the promise of public education*. New York, NY: Teachers College Press.
- Noguera, P. A., & Wells, L. (2011). The politics of school reform: A broader and bolder approach for Newark. *Berkeley Review of Education*, 2(1), 5–25.
- Paris, D. (2012). Culturally sustaining pedagogy: A needed change in stance, terminology, and practice. *Educational Researcher*, 41(3), 93–97. <https://doi.org/10.3102/0013189x12441244>
- Perelman, L. (2014). When “the state of the art” is counting words. *Assessing Writing*, 21, 104–111. <https://doi.org/10.1016/j.asw.2014.05.001>
- Perry, K. (2012). What is literacy? A critical overview of sociocultural perspectives. *Journal of Language and Literacy Education*, 8, 50–71. Retrieved from http://jolle.coe.uga.edu/wp-content/uploads/2012/06/What-is-Literacy_KPerry.pdf
- Petruzzzi, A. (2008). Articulating a hermeneutic theory of writing assessments. *Assessing Writing*, 13, 219–242. <https://doi.org/10.1016/j.asw.2008.10.005>
- Philip, T. M. & Garcia, A. D. (2013). The importance of still teaching the iGeneration: New technologies and the centrality of pedagogy. *Harvard Educational Review*, 83, 300–319. <https://doi.org/10.17763/haer.83.2.w221368g1554u158>
- Picciano, A. G., & Spring, J. (2013). *The great American education-industrial complex: Ideology, technology, and profit*. New York, NY: Routledge.
- 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 Press.
- Ravitch, D. (2013). *Reign of error: The hoax of the privatization movement and the danger to America's public schools*. New York, NY: Knopf.
- Saettler, P. (2004). *The evolution of American educational technology*. Greenwich, CT: Information Age Publishing.
- Scott, J. C. (1998). *Seeing like a state: How certain schemes to improve the human condition have failed*. New Haven, CT: Yale University Press.
- Selwyn, N. (2014). *Distrusting educational technology: Critical questions for changing times*. New York, NY: Routledge.
- Sharp, L. A. (2016). ESEA reauthorization: An overview of the Every Student Succeeds Act. *Texas Journal of Literacy Education*, 4, 9–13.

- Shermis, M. D. (2014). State-of-the-art automated essay scoring: Competition results and future directions from a United States demonstration. *Assessing Writing*, 20, 53–76. <https://doi.org/10.1016/j.asw.2013.04.001>
- Shermis, M. D., & Burstein, J. C. (Eds.). (2003). *Automated essay scoring: A cross-disciplinary perspective*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Shohamy, E. (2013). The discourse of language testing as a tool for shaping national, global, and transnational identities. *Language and Intercultural Communication*, 13, 225–236. <https://doi.org/10.1080/14708477.2013.770868>
- Stimson, S. (2000). Rethinking the state: Perspectives on the legibility and reproduction of political societies. *Political Theory*, 28, 822–834. <https://doi.org/10.1177/0090591700028006006>
- Street, B. V., & Lefstein, A. (2007). *Literacy: An advanced resource book*. New York, NY: Routledge. <https://doi.org/10.4324/9780203463994>
- Taubman, P. M. (2009). *Teaching by numbers: Deconstructing the discourse of standards and accountability in education*. New York, NY: Routledge.
- Tomlinson, S. (1997). Edward Lee Thorndike and John Dewey on the science of education. *Oxford Review of Education*, 23, 365–383. <https://doi.org/10.1080/0305498970230307>
- Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times*. San Francisco, CA: Jossey-Bass.
- Tukey, J. W. (1958). The teaching of concrete mathematics. *The American Mathematical Monthly: The Official Journal of the Mathematical Association of America*, 65, 1–9. <https://doi.org/10.2307/2310294>
- Ugodulunwa, C., & Wakjissa, S. (2015). Use of portfolio assessment technique in teaching map sketching and location in secondary school geography in Jos, Nigeria. *Journal of Education and Practice*, 6(17), 23–30.
- Valdez, P. S. (2001). Alternative assessment: A monthly portfolio project improves student performance. *Science Teacher*, 68(8), 41–43.
- Wagner, T. (2012). *Creating innovators: The making of young people who will change the world*. New York, NY: Scribner.
- 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.
- Yengo, C. (1964). John Dewey and the cult of efficiency. *Harvard Educational Review*, 34, 33–53. <https://doi.org/10.17763/haer.34.1.17835n27v7772901>