The University of Texas at Austin
Texas Education Review
College of Education

Texas Center for Education Policy

Journal Homepage: Texas Education Review
Published online: August 2021
Submit your article to this journal

Shooting for the STAAR: An Authentic Assessment Pilot Proposal to Replace Inequitable High-Stakes Accountability

ALEJANDRO MADRIGAL III
The University of Texas at Austin

ELIZA EPSTEIN
The University of Texas at Austin

To cite this article: Madrigal, A. & Epstein, E. (2021). Shooting for the STAAR: An authentic assessment pilot proposal to replace inequitable high-stakes accountability. *Texas Education Review*, *9*(2), 44-68. http://dx.doi.org/10.26153/tsw/13910

Shooting for the STAAR: An Authentic Assessment Pilot Proposal to Replace Inequitable High-Stakes Accountability

ALEJANDRO MADRIGAL III
The University of Texas at Austin

ELIZA EPSTEIN
The University of Texas at Austin

Texas has long been considered both a model and menace of education reform (Haney, 2000; Kuhn, 2013). With the introduction of standards-based education reforms in the state in the 1990s, school leaders touted high test scores and closed achievement gaps, cementing the Texas model as successful and thus replicable (Haney, 2000). But what began as a narrative of "the Texas miracle" was later revealed to be myth based on suspect sources, missing students, and mirage—a miracle that was "more hat than cattle" (Haney, 2000, p. 124). Despite three decades of questions about the legitimacy and value of high-stakes standardized testing both within Texas and across the nation, these costly exams continue to dominate the educational assessment landscape (Kamenetz, 2014).

Over the last decade, calls to divest from high-stakes, standardized assessments grew profoundly (Hagopian, 2015). In 2012, a majority of Texas school districts signed a resolution denouncing the excessive use of high-stakes tests (Scott, 2012). Since the resolution, parents and community members, exhausted by excessive testing and "teach-to-the-test" instruction, increasingly opt their children out of standardized tests (Michels, 2014; Curtis, 2019). A 2019 poll conducted by Raise Your Hand Texas (2020) found that over 70% of respondents opposed STAAR (State of Texas Assessment of Academic Readiness) as a measure of accountability for Texas public schools. The mayor of Devers, a city east of Houston, went as far as to ban STAAR within the city limits in peaceful protest (Horelica, 2019). As of June 2021, the "Texas Parents Opt Out of State Tests" Facebook group has over 40,000 followers actively organizing against high-stakes assessment (Texas Parents Opt Out of State Tests, n.d.). Many members of this group, alongside others, decried the demoralizing and damaging impacts of the "drill and kill" model of schooling demanded by STAAR in over six hours of impassioned testimony before the House Committee on Education during the 86th legislative session in 2019 (Swartz, 2019a; Swartz, 2019b).

Opposition to Texas' high-stakes testing system is not restricted to community members. Lawmakers and school stakeholders alike increasingly vocalize their opposition to the use of high-stakes standardized tests like STAAR (Carpenter, 2019; Raise Your Hand Texas, 2020). While the COVID-19 pandemic influenced recent bipartisan calls from 68 legislators to suspend the STAAR for 2020-2021 (Carpenter, 2020), the Texas State Teachers Association argued, "even under normal circumstances, STAAR exams and test prep waste millions of tax dollars and rob students and teachers of valuable classroom time for real teaching and learning" (Texas State Teachers Association, 2020). Given Texas's history as arbiter for nationwide high-stakes testing policies (Haney, 2000), as well as influential purveyor of textbooks (Crocco, 2014; Davies, 2020), we believe the state is in position to once again lead the way in transforming assessment.

The Cases Against Standardized High-Stakes Testing

This article is organized into five main sections. We first discuss the inadequacy, inefficacy, and bias of standardized, high-stakes assessments. The second section identifies unintended deleterious

effects of high-stakes assessments on classroom instruction and curriculum.[1] The third section discusses alternative assessments and examples of their successful implementation at multiple national sites. In the fourth section, we lay out House Bill 1867—a bill that would establish a Texas Commission on Assessment and Accountability tasked with recommending a high-quality accountability system. The purpose of this new system would be to move beyond the—we argue—costly, racist, inequitable, and punitive system currently plaguing the state, its students, and its educators. We conclude with a proposed amendment to HB 1867 that includes the development of a district-level pilot program to explore and develop a system of authentic assessments to replace STAAR.

Failed Assessments

Continued calls to scrap STAAR come from concerned scholars, researchers, and community stakeholders. Repeated revelations about the weaknesses of high-stakes testing—from technical shortcomings to racist roots—demonstrate the numerous harms of the current system. These revelations include the reproduction of schooling and socioeconomic inequalities (Au, 2016; Kendi, 2016a), gaming the system (Amrein-Beardsley et al., 2010; DeMatthews & Knight, 2019; Deming et al., 2016), increased stress on students and school officials (Putwain & Remedios, 2014; Putwain & Symes, 2011; Thibodeaux et al., 2015; Walker, 2014), and negative consequences for students overall (Au, 2016; McNeil, 2005), but particularly students of color, emergent bilingual students, and students with disabilities (Darling-Hammond, 2007; Heilig & Darling-Hammond, 2008; Kendi 2016a; Ladson-Billings, 2006; Jones, 2007). Stakeholder pushback against high-stakes testing echoes the critiques of the educational research community in such a way that the above-described consequences are taken as emblematic of high-stakes testing itself. While opposition commonly targets profit-driven corporations like Pearson (Au & Gourd, 2013; Blakeslee, 2013; Guisbond, 2014), continued investment in high-stakes testing as a means of education reform reveals the shared interests of politicians and corporations alike in supporting the *assessment industrial complex* (Conn & Tenam-Zemach, 2019). The term assessment industrial complex describes the interconnected interests of neoliberal policy makers, corporations, and education reformers who support not only high-stakes tests, but myriad other test-support products and projects like educator professional development and testing prep materials (Conn & Tenam-Zemach, 2019). Additionally, and arguably more powerfully, the assessment industrial complex forecloses dialogue and imagination about the purposes of education and co-opts popular opinion about improving education in order to line corporate wallets (Conn & Tenam-Zemach, 2019).

STAAR: A Flawed and Expensive Instrument

Technical critiques point to the fact that standardized-test questions are sourced from obscure material and constructed in intentionally challenging ways in order to generate variability in responses and scores that legitimize the tests themselves (Cheek, 1993). And if by some (Texas) miracle all students achieved 100% proficiency, accusations of cheating or test illegitimacy would abound (Au, 2016). Broader economic critiques recognize the way that "knowledge is purchased, not made" (Conn & Tenam-Zemach, 2019, p. 24) in high-stakes, test-based accountability systems, resulting in an assessment industrial complex that chains states and districts to branded materials hawked by

[1] We use a non-parenthetical 'unintended' in this section of the brief but later trade its use for a suggestive parenthetical '(unintended)'. The choice is not solely stylistic; interrogation of the histories of standardized exams prompts us to question the intentions of current test authors and promoters, especially provided they also know these tests' histories.

multinational corporations (Apple, 2006; Conn & Tenam-Zemach, 2019; Fabricant & Fine, 2013). Political appointees determine confidential, arbitrary cut scores, which determine who passes and who fails (Kuhn, 2015). In Texas, testing materials have been replete with errors (Ayala, 2016; Chang, 2016) and misaligned questions where students are expected to know content that falls outside of the grade-level expectations (Szabo & Sinclair, 2012; Szabo & Sinclair, 2019). Standardized tests purport to provide information about student learning, but actually reveal more about other non-school factors (Amrein & Berliner, 2002; Au, 2016). For example, researchers in New Jersey were able to accurately predict student standardized test scores using racial and demographic data (Tienken, 2017).

Students Don't Learn in a Day

Even if STAAR were capable of providing a quality measure, which we argue it does not, STAAR scores only deliver an abstracted snapshot of student performance on *one* test (a flawed and racist one), on *one* day, in *one* subject. While standardized test scores do provide a number, these numbers are simplified, at best, and are incapable of providing rich information about whether, or what, a child has learned (Gagnon & Schneider, 2017; McNeil, 2005). Unlike educators, STAAR tests do not know how to take a different approach, rethink pedagogy, or provide more scaffolding if a student is sick, hungry, tired, or just misunderstanding a question.

Let's think about how STAAR might translate to the NBA. Former Houston Rocket[2] James Harden is a phenomenal basketball player, earning MVP honors for the 2018-2019 season, and leading the NBA in free throw attempts—and makes—every year since the 2014–15 season. For his exceptional performance, he signed a four-year contract worth over $42 million dollars annually. But what if the Rockets evaluated his play based on one day, rather than on seasonal or career averages? For example, while Harden hit 44% of field goal attempts and averaged 36.1 points per game during the 2018-2019 season, on April 20th, 2019, in a critical playoff game with season-ending implications, he shot only 15% from the field and scored only 22 points. This example reflects the problematic nature of evaluating anyone on anything through a single-day snapshot of performance (see Rothstein, 2000).

There is More than One Right Answer

Standardized testing in the United States finds lineage in eugenics work and theory (Stoskopf, 2012).[3] Lewis Terman, a contributor to the design of the Stanford-Binet IQ test, and Carl Brigham, the designer of the Alpha Military test (which later became the SAT), situated their exams in eugenicist thought with the intent of establishing an empirical measure of White supremacy (Kendi, 2016a; Rosales & Walker, 2021). These psychologists, alongside their peers, argued their tests scientifically proved intelligence to be hereditary. The racial and social class bias of the tests' questions unsurprisingly yielded Brigham and Terman's desired result—the demonstration of Whites as intellectually

[2] By the time of publication, Harden had been traded to the Brooklyn Nets, though the analysis of his contract remains relevant.

[3] Eugenicists claimed certain groups of people were predisposed to "defective genes" (Stoskopf, 2012, p. 34). Much eugenicist rhetoric harped on the inferiority of non-Whites, though eventual calls for sterilization included non-Whites, persons identified as having disabilities, and anyone deemed to be a detriment to society (Kevles, 1999; Stoskopf, 2012).

superior (Au, 2016; Karier, 1972).[4] W.E.B. DuBois observed that standardized IQ test questions were structured and "adjusted so as to put black folk absolutely beyond the possibility of civilization" (cited in Guthrie, 1988, p. 55).

The tests soon underpinned the creation of the first gifted and talented programs, as well as the eventual tracking and sorting of students into segregated education courses (Stoskopf, 2012). Leta Hollingworth, a professor at Teachers College and a pioneer of gifted and talented programs, echoed both Terman's eugenicist and IQ test-committed logic, decrying the reproduction of the "stupid, the criminal, and other mentally, physically, and morally deficient" (Stoskopf, 2012, p. 37). As to whether Terman realized or intended this connection between his IQ tests and the segregation of students, we are guided by his words:

> Among laboring men and servant girls there are thousands like them… as far as intelligence is concerned, the tests have told the truth… No amount of school instruction will ever make them intelligent voters or capable voters in the true sense of the word… children of this group should be segregated in special classes… (Stoskopf, 2012, p. 36)

By 1930, standardized test results justified sorting students by purported intelligence and potential in elementary, secondary, postsecondary, and military spaces (Rosales & Walker, 2021). By the 1950s, some U.S. universities employed the SAT in admissions processes in explicit attempts to exclude Black students (McCardle, 2020; Price, 2019). In this way, the SAT normalized a decades-long belief in standardized tests as having "told the truth" about intelligence and potential (Terman, 1916, p. 92). The education system's commitment to evaluation grounded in eugenics and holding it as scientific truth all but guaranteed the marginalization of non-White students, students identified as having disabilities, and any student deemed unfit for advanced learning.[5] The high stakes nature of these exams can be summed up in the eventual calls to sterilize these groups—a low test score not only meant that one was lesser, but that one should not be allowed to procreate (Kevles, 1999; Stoskopf, 2012).

Despite aptitude tests like the SAT finally seeming to be losing their grip on college admissions (Tugend, 2019), for decades the SAT and its attendant "merit" scholarships have determined who is

[4] Throughout this paper, we capitalize White. We are guided in this temporally-situated decision by Eve Ewing (2020), who wrote: "As long as White people do not ever have to interrogate what Whiteness is, where it comes from, how it operates, or what it does, they can maintain the fiction that race is other people's problem, that they are mere observers in a centuries-long stage play in which they have, in fact, been the producers, directors, and central actors" (n.p.). As she also argued, there are reasons not to capitalize it, but this need to demonstrate its existence, to elucidate the colonial logics foundational to this nation and its oppressive institutions, are critical to this paper.

[5] It must be made exceedingly clear that support for eugenics and its derivative instruments in the United States was not fringe nor extremist. Eugenics courses were common curricula at the postsecondary level, with the number of offered courses expanding from 44 to 376 between 1914 and 1928 (Stoskopf, 2012). Farber (2008) describes the rise of pro-eugenics organizations in the United States, beginning with the Eugenics Records Office in 1910. He writes of Charles Davenport, then American-based zoologist and eventual international eugenics leader, creating The Office with assistance from philanthropic donors, including one Harvey Kellogg. Two more pro-eugenics organizations, the Eugenics Research Association and the American Eugenics Society, soon followed in influence throughout the United States and Europe (Farber, 2008). Farber goes on to highlight H.H. Laughlin, co-founder of the American Eugenics Society and superintendent of the Eugenics Records Office, one of whose publications included drafting of what would become model law for compulsory sterilization in the United States. In the realm of non-academics, eugenics theorizing materialized in "Better Baby" and "Fitter Family" contests (Chen, 2009).

worthy of acceptance to — and funding for — colleges nationwide (Rosales & Walker, 2021; Troy, 2016). Brigham, the architect of what would become the SAT, himself denounced the test later in his life — admitting that the SAT test revealed nothing about intelligence, but rather provided "a composite including schooling, family background, familiarity with English and everything else, relevant and irrelevant" (Brigham, 1930, cited in Lemann, 2000, p. 34). These words, found in Brigham's unpublished works, would not reach or influence those advocating for the SATs, nor the hundreds of universities who were basing admissions and distributing "merit" scholarships on student scores (Lemann, 2000).

The SAT and the Stanford-Binet are aptitude tests are not criterion-based tests like the STAAR, yet the two types of tests are connected by a strong ideology bound to the notion that these exams tell us an objective truth about student learning or academic potential.[6] This narrative holds strong today, through tests like STAAR, and drives a national dogma to classification by test score. But where low SAT scores slam the door on opportunity for higher education, low scores on high-stakes assessments like STAAR often lead to further student surveillance (Au, 2016; Grady et al., 2012), school reconstitution (Elmore, 2002), and narrower curriculum (Darling-Hammond, 2007), pushing out students and killing their educational dreams —at least within public schools (McKay et al., 2015), as we discuss further in the next section. Academic tracking abounds as students are segmented into "high" and "low" academic tracks (National Association of Secondary School Principals, 2006; Oakes, 2005), sifting students into categories which determine the nature of their schooling experience and their opportunities for the future (Au, 2106; Oakes, 2005). Students identifying as African American, Black, or Latino are also more likely to be placed into low-track courses compared to White students, spaces often provided with fewer resources (Knoester & Au, 2017; Oakes, 2005). Though missing the explicit rhetoric of eugenics, functions of high-stakes testing systems today are frighteningly reminiscent of the eugenicist projects of the early 20th century. Segregating students in this way continues the project of Terman and others in practice, if not in spirit.

Standardized Testing Perpetuates Racism

Standardized testing thus began as and continues to be "a racial project in the United States" (Au, 2016, p. 43). Identifying the racist, sexist, nativist, ableist roots of the high-stakes standardized assessment project demands a reassessment of what theories of change are driving today's educational reformers (Stoskopf, 2012). Continued allegiance to White-normed accountability systems creates an educational landscape that, we argue, contributes to the erasure of non-White ways of being and knowing, and produces racialized outputs to match the racialized inputs. As a result of the assessment industrial complex (Conn & Tenam-Zemach, 2019), "communities of color have lost a say in what their children learn and how they get to learn it" (p. 130). As such, the system of high-stakes testing disproportionately negatively impacts the material conditions, lives, and life opportunities of Black, Brown, and Indigenous students (Au, 2016; Conn & Tenam-Zemach, 2019). We assert that interrogating standardized, high-stakes testing systems in this way is pivotal to rectifying inequities in the assessment industrial complex. It is precisely this kind of conversation deepening we advocate for at the K-12 level through authentic, rather than standardized, assessments.

[6] Definitionally, aptitude tests like the SAT and Stanford-Binet are norm-referenced tests. These tests make comparisons between individuals. Criterion-referenced tests, like the STAAR, measure a test taker's performance compared to a specific set of standards or criteria (Burkett, 2018). We argue that either iteration, especially when used in a high-stakes system, replicates inequity and perpetuates harm to students.

It is a fallacy that multiple choice, standardized tests are objective instruments (Au, 2016). Some test advocates even argue that high-stakes accountability systems serve as civil rights protections for students,[7] particularly those from vulnerabilized populations (Derstine, 2015; Leadership Conference on Civil and Human Rights, 2015).[8] Instead, these tests evaluate students on racialized and classed knowledge (Au, 2016; Conn & Tenam-Zemach, 2019; Kendi, 2016a; Weiner, 2014). High-stakes accountability systems do not deliver racial justice by closing achievement gaps. Instead, they police boundaries and maintain the inequitable status quo (Au, 2016). We argue that attention to vulnerabilized students is critical, but that policing and punishment are not the answer to issues of schooling inequity. We need to think differently about how to invest our education energy, or else risk continuing the reproduction of schooling inequities perpetuated by standardized, high-stakes exams. We suggest looking to broader abolition movements, which recognize the fundamentally racist and oppressive nature of any policing mechanism; we the authors dream of freer futures (Boggs et al., 2019, Davis, 2005; Love, 2018) where schools generate growth, learning, and collective strength—futures made possible by the legislation proposed in Texas that we lay out later in this work. We urge all education stakeholders, but especially policymakers, to step beyond the hierarchy and sorting that are endemic to the logics of no child (being) left behind (Amrein & Berliner, 2002). The idea that no human is disposable requires a different approach to learning, one that is collective and transformative (Love, 2019; Shalaby, 2020).

Consequences for Teachers, Classrooms, and Students of Color

This second section identifies (unintended) deleterious effects of standardized assessments on curriculum and classroom instruction. Failing to deliver on promises made by proponents, these assessments actually do long-term harm to teachers and students, with a disproportionate negative impact on communities of color (Amrein & Berliner, 2002; Kendi, 2016a; Ladson-Billings, 2006).

Inequity Maintained

Assessment measures that consistently reflect inequities rather than mitigate them cannot be used for equitable ends. Shining a light on a problem does nothing to solve that problem. And, if forty years of shining the lights hasn't solved the problem, then the light is shining in the wrong direction. High-stakes testing policies have "a disproportionate negative impact on students from racial minority and socioeconomic backgrounds" (Amrein and Berliner, 2002, p. 11). Continued allegiance to standardized tests as an accountability tool belies racist beliefs that students of color and low-income students do not measure up, when in fact, the tests and accountability norms are the problem—not the students (Au, 2016). Continued use of these high-stakes tests perpetuates trauma and racism at the expense of authentic learning and equity, *de facto* contributing to the creation and reinforcement of a racial hierarchy (Desai & Sanya, 2016; Kendi, 2016b). Because high-stakes assessment and their

[7] As Au (2016) points out, some organizations like the NAACP, National Council of La Raza, and LULAC (who recently published and then retracted a letter asking the Biden Administration to reject requests for testing exemptions in the spring on 2021) have flip-flopped on the importance of high-stakes testing for civil rights protections. While not arguing causality, Au points out the large sums of money provided to these organizations by foundations like Gates and Walton, both strong supporters of high-stakes testing.

[8] We use 'vulnerabilized' here and elsewhere in this brief as a means to differentiate persons from the political, social, economic, and environmental conditions that surround them. We note that they have been made vulnerable by the forces of neoliberalism and racial capitalism and how those forces operate within schools (Au, 2016; Braginsky, 2020; Kelley, 2002; Tuck, 2009).

derivative accountability systems function in this way, we assert the use of these measures to itself be a racist act.[9]

Do (no) Harm

Despite visionary goals to close the "achievement gap," year after year the schools under the most pressure to perform are disproportionately attended by students of color and students from low-income households (Au, 2016). The pressure and cut-scores of high-stakes tests and accountability have driven up the dropout rate and reports of stress-induced illness in students and educators in both elementary and secondary schools (Au, 2016; Counsell & Wright, 2018; Darling-Hammond, 2007; Lobman, 2014; Nichols et al., 2006). There is evidence of both pushout—schools pressuring lower-performing students to leave or holding them back to drive scores up—as well as students leaving school because they do not achieve passing scores or are frustrated by being held back (Advancement Project, 2010; Darling-Hammond, 2007; Glennie et al., 2012; McNeil et al., 2008). Dropout and pushout disproportionately impact students of color, while also creating "much less engaging, and even hostile" (Advancement Project, 2010, p. 5) school environments for students identified as having disabilities, students from low-income households, and emergent bilingual students (Au, 2016; Darling-Hammond, 2007; Palmer & Rangel, 2011; Sunderman & Kim, 2004). Census data shows that Black students in particular are pushed out/dropout at higher rates when schools require exit exams (Dee & Jacob, 2006). Research around stereotype threat demonstrates the ways that students in high-poverty schools and students of color, particularly Black males, bear the brunt of the negative impact from test pressure (Holme et al., 2010; Steele, 1999). Additionally, educators and administrators may be inclined to use "fear-appeals," or language that attempts to motivate by warning students of the consequences of not passing or doing well on tests, which subsequently increases student stress and lowers their performance (Putwain & Remedios, 2014; Putwain & Symes, 2011). An email published in the New York Times, sent from a charter school educator, exposed the intensity of the pressure, warning students of the risks of not following the "plan of attack" for high achievement on the state assessment. The email message stated: "Any scholar who is not using the plan of attack will go to effort academy, have their parent called, and will miss electives. This is serious business, and there has to be *misery felt* for the kids who are not doing what is expected of them" (Taylor, 2015, para. 5, emphasis added). This email demonstrates the ways that both the schooling system and students are surveilled and punished as a result of high-stakes assessment. And that this email came from an educator at Success Academy in Harlem, a school where 98% of attendees are identified as students of color, further demonstrates the disproportionate impact of testing surveillance for that same student group (Au, 2016; Public School Review, n.d.). In Texas, State Senator Jose Menendez highlighted last year that the tests are doing "unnecessary harm" to our state's students and communities (April 13, 2019, para. 6).

This harm extends beyond schools. The introduction of standards-based exit exams is correlated with an increase in the incarceration rate (Baker & Lang, 2013). A Chicago Public School student

[9] We name the testing industrial complex racist. We do this in solidarity with activists, parents, scholars, and students who have pushed to end high-stakes testing for decades. The authors acknowledge that we are members of the education system, and have participated in upholding the testing system in our roles as educators. We are not pointing fingers and placing blame. In this paper, we are focused on identifying the misguided theories of change that underpin testing regimes as mechanisms of equity and challenge ourselves, policy makers, legislators, and others to both imagine more humanizing and meaningful assessment practices and make steps in that direction.

who "blamed herself" for not passing a high-stakes exam shared her thoughts, revealing the ongoing harm of these tests:

> Because it does not...test my knowledge, because I'm getting all these honors classes and A.P. [Advanced Placement]...it does not say who I am or what is my strength, it just doesn't say anything about me. It's just a stupid number that they put on your forehead. It's injustice. It's a stupid way to...decide whether a student should pass or stay. (Lipman, 2003)

When students are forced out of education system before completing high school, they face more obstacles to entering higher education or securing jobs with livable wages (Au, 2016; Fine & Pryiomka, 2020). High-stakes testing creates categorical sorting beyond internal school tracking (Oakes, 2010); student opportunities for future success in college, career, and life is put at risk if they leave the education system (at least in part) because of high-stakes testing (Au, 2016; Fine & Pryiomka, 2020).

What Gets Tested is What Gets Taught

A growing body of literature points to standardized, and especially high-stakes, testing as contributing to over-alignment of curriculum-to-tests, pedagogical shifts towards test prep, and the exacerbation of schooling inequalities (Au, 2016; Darling-Hammond, 2007; Holme et al., 2010; Neill, 2003). In terms of curriculum covered, teachers are forced to prioritize tested subjects like math and literacy, putting science and social studies on the back burner and, commonly, skipping arts, music, and physical education altogether (Au, 2016; Holme, 2008). One Texas educator shared their experience with curriculum changes as a result of high-stakes assessment:

> As part of U.S. history for my fifth graders, I used to teach about what happened in the Americas before Columbus arrived. I spent several weeks teaching about the cultures of Native America —which included the Aztec, Maya, and Inca civilizations. No longer. It's not in the fifth-grade standards. Now it's in the fourth grade, but the standards concentrate on the Spanish Conquistadors and talk mainly of Cortez, Coronado, and Pizarro. In short, I no longer teach a curriculum. I teach test-preparation. (Beam-Conroy, 2001)

Teaching to the test creates a narrow, bromidic curriculum that is rife in schools labeled at-risk — disproportionately attended by Black and Brown students—and scarcely seen in wealthier, Whiter districts (Nichols & Berliner, 2007), resulting in a profoundly different, and higher quality, educational experience for White students (Au, 2016).

Testing Traumatizes Teachers Too

In addition to narrowed curricula, teachers shift pedagogical approaches as a response to top-down testing pressures. Teachers find their creativity suppressed and their autonomy diminished (Crocco & Costigan, 2007). At the classroom level, time that might otherwise be spent on extension of learning or development of new skills is instead dedicated to test preparation and content standards review (Au, 2011; Neill, 2003; Nelson, 2013). As such, miracles in test score improvement on these assessments, while framed as successes, may be attributed to teachers' focusing on predictably tested content rather than overall student learning or mastery of standards (Amrein & Berliner, 2002; Jennings & Bearak, 2014; Koretz, 2005; Nichols et al., 2012). Teachers find themselves wrestling with the tension between their desire for delivering engaging instruction and their desires for their

students to do well on the high-pressure tests (Vogler, 2008). Even when teachers reported high curricular rigor in their classrooms, ethnographic observations in one high poverty district revealed a marked increase in test-prep behaviors because of high-stakes assessments (DeBray, 2005). Additionally, high-stakes tests negatively impact emergent bilingual students who are forced into sheltered English classes and tested in a language that they barely know (Bach, 2020). Teachers turn to test strategies and tricks to support these learners in meeting graduation requirements; these practices diminish opportunities for real learning (Bach, 2020) and alienate educators (Palmer & Rangel, 2011). Pedagogy becomes teacher-centered and the structure of knowledge becomes more fragmented (Au, 2007). For example, students train for tests by completing worksheets with discrete tasks and chunks of information that they memorize, rather than learning holistically about a phenomenon, concept, or theory (Au, 2007; Nelson, 2013). Overall, teaching quality is lower when top-down pressures lead to narrowed curricula and test-friendly pedagogy (Blazer & Pollard, 2017; Valli et al., 2012), which has the additionally harmful result of driving teachers to leave the teaching profession (Thibodeaux et al., 2015; Walker, 2014).

But most importantly, the joy, creativity, and possibility of education are lost as worksheets and practice tests proliferate (Longo-Schmid, 2016). One educator, who is the closest learning partner with students, astutely observed, "All children are left behind because we are so test-driven in schools today that we do not give children the opportunity to explore their minds or to think outside the box" (Cole, 2009, p. 6). This is more likely to happen in under-resourced schools, which are disproportionately attended by Black and Brown youth (Au, 2016; Darling-Hammond, 2007). High-stakes assessments and their attendant neoliberal policies have foreclosed educational opportunities by narrowing curriculum and offering test strategies in place of pedagogy. The testing industrial complex cemented a system by which we "value what we measure" (because we can measure it) and fail to ask whether we should instead "measure what we value" (Biesta, 2014, p. 46). Continuing to uphold this system will further harm students, with particularly harsh impacts on students of color, low income students, and emergent bilingual students (Au, 2016; Bach, 2020). Shifting assessment practices is a critical step in building anti-racist and just schools (Au, 2016; Fine & Pryiomka, 2020).

The Time for Transformative Change

For nearly two decades, the deficit-based narrative that low test scores signal failing schools justified sanctioning the schools most in need of support (Gagnon & Schneider, 2019; Urrieta, 2004). Over twenty years of investment in high stakes assessment has not "closed the gaps" or paid off the educational debt (Ladson-Billings, 2006). High-stakes testing is flawed, harmful, and impedes opportunities for transformative and liberatory education (Au, 2106; Urrieta, 2004). In this section, we present research and examples of alternative, authentic projects of assessment which have proven successful around the United States. Authentic assessments are ways of evaluating students' ability to apply knowledge and skills to real world problems that exist outside of the classroom (Wiggins, 1998). It is time for a more holistic and locally derived accountability system that builds capacity as it measures—one that invests resources into and draws strength from the journey of collective knowledge production within the classroom (Urrieta, 2004).

In Texas and across the nation, students are subjected to a "fundamental misalignment between the nation's aspirations for its students and the assessments used to measure whether they are achieving those goals" (Darling-Hammond & Adamson, 2010, p. 2). There exists a compelling interest for school districts in Texas, alongside their rapidly growing and increasingly passionate anti-testing constituency, to research new tools that can simultaneously foster and gauge student learning. Texas

should invest in assessments that are student-led and teacher guided and that provide immediate, meaningful feedback far superior to that offered by end-of-year summative, standardized testing (Bland & Gareis, 2018; Darling-Hammond & Adamson, 2010; Guisbond, 2014). Fortunately, sufficient research and real-life models exist to guide our state in this direction (e.g. Bland & Gareis, 2018; Fine & Pryiomka, 2020; Fontana, 1995).

What Makes an Assessment Authentic?

Authentic assessments demand higher order thinking and measure students' knowledge and skills at deeper levels than traditional assessments (Koh et al., 2011). For an assessment to be considered authentic, it must have value beyond the actual score or grade, indicating that the assessment task itself is meaningful to the learner (Frey et al., 2012). Authentic assessments are attentive to both the *product* and the *process* of student engagement with the assignment. In this way, authentic assessments are *assessments for learning*, not assessments of learning (Bland & Gareis, 2018). Authentic assessment involves the tailoring of assessment to students' needs and involves evaluations by teachers, within a school or as part of a state- or district-wide assessment team, through collectively composed rubrics (Archbald, 1991). Appendix A contains a chart that provides more clarity on the ways that authentic assessments, also called alternative or performance assessments (Bland & Gareis, 2018), differ from traditional assessments.

Authentic assessments can take many forms, including constructed-response simulations (see Appendix B), performance assessments, portfolios, essays, debates, and projects. What defines them is not their specific form or content, but the powerful way they link instruction with assessment and meet individual students' needs (Bland & Gareis, 2018; Darling-Hammond & Snyder, 2000). Classrooms with authentic assessment are student-centered. Educators emphasize cooperative work; focus more on writing, problem solving, and real-world, hands-on activities; and deemphasize rote learning and teaching (Koretz et al., 1996). Appendix B includes a sample authentic assessment which asks students to respond to a real-life scenario analyzing and synthesizing select materials to draw conclusions and present recommendations. The sample assessment in Appendix B demonstrates many features of authentic assessment.

- It is in-depth, grounded in the real-world, and makes expectations visible to students, allowing them to design and evaluate their own work.
- Demands students employ critical thinking to solve a problem of an interdisciplinary nature.
- Students must access and synthesize prior knowledge rather than relying on the most recent lesson to construct their response.
- Students are afforded multiple opportunities and avenues to provide an evidence-based response.
- Students are measured by what they present, not what is easy to grade. Authentic assessments like this one allow educators to construct nuanced assessments of students' reading, writing, thinking, and multi-modal communication skills which can then immediately inform the direction of future instruction.

Here in Texas, the Texas Performance Standards Project has developed a number of project-based assessments already aligned to the TEKS (Texas Performance Standards Project, n.d.).

Authentic assessments are lauded by both researchers and practitioners for their ability to work collaboratively with students to identify engaging topics and design meaningful learning (e.g. Bland & Gareis, 2018; Cook et al., 2020; Fine & Pryiomka, 2020; Koh et al., 2011; New York Performance Schools Consortium, 2018). According to researchers (e.g. Bland & Gareis, 2018; Fine & Pryiomka, 2020), authentic assessments "propel the education system in a direction that corresponds with *how* individuals actually learn" (Stanford School Redesign Network, 2008, n.p., emphasis added). And, teachers and students confirm that authentic assessment systems align with how individuals *want to learn* (Cook et al., 2020; Gisi, 2020). Schools that redesign around project- and portfolio-based learning build strong ties with their communities, and feel, in the words of one Manor New Tech High School student, "like family" (Lynch et al., 2013, p. 35). They invite artists and performers into their schools and support their students in exploring careers and projects outside the classrooms. Time commitments shift; Fridays sometimes include presentations by university scholars, a lab-based epidemiologist, or a local artist. While students engage with career professionals, teachers can collaborate on curriculum or work on assessment-design. Learning crosses borders.

New York Performance Standards Consortium Success

Developed over twenty years ago and currently thriving in 38 New York public schools, the New York Performance Standards Consortium (NYPSC) has achieved remarkable results with all learners, but particularly with vulnerabilized students. Liberated from oppressive state and federal testing, assessments of learning in NYPSC schools measure student fluency in skills like problem-solving, communication, research, expository writing, and public speaking (Cook et al., 2020; Fine & Pryiomka, 2020; New York Performance Standards Consortium, 2018). Educators and students in NYPSC schools employ "practitioner-developed, student-focused, and externally reviewed projects, papers, performances, experiments, and experiences called performance-based assessment tasks (PBATs)" to evaluate student learning (Fine & Pryiomka, 2020, p. v). PBATs acknowledge a different purpose to schools and a different theory of equity (Cook et al., 2020). Rather than standards and accountability—everyone taking the same test—as the measure of equity (Lipman, 2003), NYSPC schools focus on equity in terms of access, success, and holistic learning (Fine & Pryiomka, 2020).

NYPSC students have achieved at levels far above their peers who attend traditional schools based on high quality metrics (Fine & Pryiomka, 2020):

- 77 % of NYPSC students who began high school in the fall of 2010 graduated in four years versus 68 % for all New York City students.
- In 2015, 71 % of English learners at consortium schools graduated on time, versus 37 % of English learners citywide.
- 2018 data (New York Performance Standards Consortium, 2018) shows that:
    - Latinx students and students with disabilities are twice as likely to graduate if they attend a NYPSC school.
    - 5.3% of NYPSC students drop-out compared to 11.3% statewide.
    - Male graduates of Consortium schools identified as members of minoritized populations enroll in college at twice the rate of similar students nationwide.
    - Graduation rates for English learner students was nearly 30% higher at consortium schools compared to non-consortium schools.

A major reason for the success of the NYPSC is its ability to build "internal accountability at the school level in the service of teaching deeply, rigorously, and fairly" (Knecht, 2007, p. 63). Authentic assessment is oriented towards the learner, and the school and community become accountable to the student rather than accountable to external stakeholders and external measures as is the case with high-stakes assessment systems. Educators and students are challenged to be their best and to work together to achieve collectively defined goals on which the students are then assessed.

NYPSC students aren't just graduating and matriculating to college at higher rates, they are also doing better in college and persisting beyond the first year at higher levels than non-consortium students. Research released this summer found that NYPSC students achieved higher first-semester college GPAs, earned more initial credits, and were more likely to persist in college after the first year than peers from NYC schools, despite those students having higher SAT scores. Black males who attended NYPSC schools had particularly improved higher education outcomes compared to Black males who did not attend consortium schools (Fine & Pryiomka, 2020). Overall, these findings suggest:

- The performance-based assessment tasks (PBATs) that students complete in NYPSC schools enhance academic progress for students.
- Performance-based assessments may be better indicators of postsecondary success than standardized test scores.
- Authentic assessments are powerful tools to interrupt the persistent inequity maintained by standardized testing.

For the past 20 years, high-stakes accountability systems have offered quantitative extrapolations of student experience (Au, 2016; Bach, 2020; Conn, 2019; Darling-Hammond, 2007). One student's qualitative reflection after graduating from a NYPSC school reveals the power of authentic assessments through shifts in pedagogy and practice. Emphasizing this point she stated, "Being educated at a consortium school had a profound effect on my life. Every student is entitled to an educational community as enriching and inspiring as mine" (NYPSC, 2018).

Additional Authentic Assessment Projects

Kentucky is another state with a history of using research-based, authentic assessments including project- and portfolio-based evaluations (Fontana, 1995). In Danville, Kentucky, 98% of teachers at Bate Middle school voted for reorganizing and introducing project-based learning (PBL) and setting expectations that included "social and emotional skills, ethics, technological literacy, and career readiness" that resulted in the school being designated an "exemplar school" by the Partnership for 21st Century Skills (Kamenetz, 2014, n.p.).

Texas already has examples of PBL or portfolio-based work integrated with authentic assessments. Twenty-eight Texas secondary campuses are members of the New Tech Network, an education collective providing a model for innovation that combines PBL and technology. During its 86th legislative session, Texas lawmakers passed HB3, which provided an additional per pupil allotment of $50 for schools collaborating with the New Tech Network. These campuses buzz with energy, charged by engaged students being challenged across disciplines alongside teachers who are relishing in the curiosity of their emerging leaders. "Teachers are happier... Every day you realize why you wanted to be a teacher. It's exciting again" explains UT Austin Professor Jennifer Adair, who has helped a

number of campuses develop authentic learning and assessments in these schools (Gisi, 2020). With House Bill 1867 and the work already underway, Texas is well positioned to develop the architecture for a system of next-generation assessment.

House Bill 1867

In this section, we outline the parameters of House Bill 1867 (HB 1867), as well as provide our recommendation for an amendment to the original version which draws on the transformative power of authentic assessment.

HB 1867 would establish a Texas Commission of Assessment and Accountability, tasked with uncovering issues with the current testing regime and making recommendations for a high-quality statewide system of assessment. The nineteen-person commission would be comprised of members appointed by the Governor (four), Lieutenant Governor (seven), and the Speaker of the House (seven), as well as one member of the State Board of Education. Commission members would be demographically representative of the state of Texas and would include at least one teacher, parent, member of the business community, member of the civic community, superintendent, school trustee, and district staff. In addition, three members of the legislative chamber would serve on the commission. The commission is tasked with recommending a system of accountability and assessment that is valid, fair, timely, informative, fiscally responsible, curriculum-aligned, and just. A report would be provided to the governor and legislature no later than December 31, 2022. HB 1867 creates an opportunity to make plans for the next generation of assessment tools which would be more precise and better reflect student growth and teacher productivity.

Piloting Authentic Assessments in Texas

Following the research on authentic assessments, we recommend an amendment to HB 1867 that would establish a pilot program for select high schools in a number of public school districts to devise assessment alternatives to high-stakes, standardized testing. Guided by a local task force or "learning community" at the school district level, pilot schools will be empowered to study, create, and pilot research-based, TEKS aligned, curricularly-embedded assessments that serve as the launching point for the state's move away from flawed, one-size-fits-all, criterion-based, high-stakes exams.

Points of Consideration and Recommendations for Implementation

The ten pilot districts will begin the process leading to a successful implementation of the Commission's forthcoming recommendations. Pilot districts should meet the criteria (listed below) and develop a timeline for district status updates and outcomes to be delivered to both the state legislature and state education agency. Student learning and growth data should be obtained at various points throughout the school year. The Commission approves the school district's deliverables and timelines.

Pilot Cohort

For next-generation authentic assessments to be comprehensive, precise, reflective of student growth, and respectful of community resources and values they must be entrusted to local task forces. A number of districts in Texas have a demonstrated track record and commitment to developing authentic, comprehensive accountability measures that make them ideally suited to participate in HB 1867's pilot program. For example, districts like Humble ISD and Austin ISD, who joined the TEA-sponsored Local Accountability Systems program, as well as Frisco ISD and Northside ISD, are all long-time members of the legislatively endorsed Community-Based Accountability System and all fit the criteria for district and school inclusion in this pilot study.

Participating districts and high schools must reflect the geographic and demographic diversity of the state and include representation from rural, urban, and suburban school districts. Additionally, some pilot schools should have more bilingual students and more students with special needs. Geographic and demographic diversity—together with parent and community involvement—helps ensure student learning is recognized in multiple and varied forms that reflect the diversity of the state and of complexity of student dreams. Moreover, for authentic assessment to be grounded in the local realities, resources, and opportunity structures of communities, parent and community representation and participation are critical. Below are two possible frameworks for thinking about district participation that would best inform statewide scaling:

1. Select pilot districts around the state with variation in their location (e.g., rural, urban, suburban) and demographic representation (e.g., student and educator racial identification and socioeconomic status) to strengthen the likelihood that successful pilot models will be replicable and sustainable statewide.
2. Focus work on a few regional partnerships, while still aiming for demographic diversity among selected districts and regions. The strength of this approach is that it allows for piloting locally developed authentic assessments, with the added benefit of being able to begin the collaborative and iterative work of statewide scoring alignment.

School districts should also be exempted from district-, state-, and federally mandated, standardized testing requirements pending approval from the U.S. Department of Education and at the behest of the state's Commissioner of Education. This will provide school districts with the latitude needed to explore authentic, alternative forms of assessment without the interference and distraction of standardized testing.

Timeline

School districts should be given sufficient time to explore the efficacy of authentic assessments. For this, the state-level committee should require an initial planning year, followed by four years of implementation, the equivalent of a high school cycle from 9th grade through graduation. We recommend that schools introduce authentic assessments with a 9th grade cohort of students and add a grade each year as they build capacity. Upon completion of the plan of study, the statewide committee should submit a report to the legislature and to the state education agency detailing the school district's findings, results, conclusions, and recommendations.

Trusting Teachers

In the wake of the unprecedented shuttering of school buildings due to COVID-19 and the ongoing oscillations between in-person, online, and hybrid instructional approaches, the value of teachers statewide has been demonstrated time and again, even as lawmakers and critics continue to undervalue and undercompensate them for their labor and love (Bradford, 2021; Hadavi, 2021; Mays, 2021). Educators have shown their intimate knowledge of what students need, their extraordinary skill in designing curriculum that attends to those needs, and their creativity and flexibility even in the most unpredictable and dangerous of contexts. Texas teachers deserve our respect and our trust as we innovate Texas assessments and accountability.

Texans trust and have faith in their local schools. Over 50% of respondents in a 2019 Raise your Hand Texas (2020) poll gave their local schools a grade of A, and nearly 70% gave local teachers an A or B. Local people trust local teachers; teachers should be taking the lead on assessment. In the same poll, a majority of respondents demanded divestment from high-stakes testing and investment in teacher salaries. Even staunch STAAR supporter and TEA Commissioner Mike Morath confessed: "Not to say I wouldn't spend money on accountability, but investing in educators will give you huge capacity long term" (Swartz, 2019). HB 1867 provides an opportunity to build a cutting edge, integrated system of authentic assessment, with the added benefit of investing in teachers. Teachers are best equipped to inform the development of a system of authentic assessments that measure the actual knowledge and skills taught within a TEKS-aligned classroom. Authentic assessments represent an investment in our educators and school leaders, not consultants. School leaders and educators are professionals that know students, families, and communities *and* understand pedagogy, curriculum, and learning theory.

Conclusion

HB 1867 provides an exciting opportunity to usher in a research-informed, school-centered, and socially just system of accountability and assessment to Texas. With the pilot-program amendment, it also provides an opportunity for immediate investment in students and educators through new forms of student assessment that are a departure from the crude and racist measures of multiple-choice bubble testing. They incorporate local input, promote professional collaboration, and deepen student learning—creating more powerful futures. Authentic assessments developed by pilot districts will establish communication between students, educators and policy makers and drive continual and timely improvement of student learning, teacher instruction, and state well-being. At the end of the pilot period, we will be on our way to implementation, and ultimately institutionalization, of authentic assessment methods and measures embedded within a larger accountability framework that is informational rather than punitive. Texas has been a national leader on school accountability for decades, and now is the time for Texas to take the lead in redesigning assessment and accountability.

Alejandro (Trae) Madrigal III is a master's student in the UT Austin Education Policy and Planning program. He studies accountability pressures impacting students and schools, with a focus on the impacts of high-stakes testing on student schooling experiences. Additionally, he studies school stakeholder experiences in marketized schooling environments and purposes of education. He is a former middle and high school social studies educator, as well as a former powerlifting and basketball coach.

Eliza Epstein is an educator, scholar, activist, and student. She is a doctoral candidate in the UT Austin Educational Policy and Planning program studying liberatory, abolitionist, and anti-colonial policies and pedagogy, with a focus on the expansion of Ethnic Studies policy at the state and local level. Additionally, she studies community college transfer and purposes of education. She is an organizer with Undoing White Supremacy Austin, District 5 for Black Lives, and the Ethnic Studies Network of Texas. She is a former high school English educator and cross country coach and worked for over a decade as a film editor in California.

References

Advancement Project. (2010). *Test, punish, and push out: How "zero tolerance" and high-stakes testing funnel youth into the school-to-prison pipeline.* https://www.urbancollaborative.org/sites/default/files/resource_files/testpunishpushout.pdf

Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives, 10*(0), 18. https://doi.org/10.14507/epaa.v10n18.2002

Amrein-Beardsley, A., Berliner, D. C., & Rideau, S. (2010). Cheating in the first, second, and third degree: Educators' responses to high-stakes testing. *Education Policy Analysis Archives, 18*(14). https://doi.org/10.14507/epaa.v18n14.2010

Archbald, D. A. (1991). Authentic assessment: Principles, practices, and issues. *School Psychology Quarterly, 6*(4), 279-293. https://doi.org/10.1037/h0088821

Apple, M. W. (2006). *Educating the "right" way: Markets, standards, god, and inequality* (2nd ed.). Routledge.

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher, 36*(5), 258-267. https://doi.org/10.3102/0013189X07306523

Au, W. (2016). Meritocracy 2.0: High-stakes, standardized testing as a racial project of neoliberal multiculturalism. *Educational Policy, 30*(1), 39-62. https://doi.org/10.1177/0895904815614916

Au, W., & Gourd, K. (2013). Asinine assessment: Why high-stakes testing is bad for everyone, including English teachers. *English Journal*, *103*(1), 14-19.

Ayala, E. (2016, April 27). STAAR test question with no right answer among dozens of complaints as Texas school chiefs say results can't be trusted. *The Dallas Morning News*. https://www.dallasnews.com/news/2016/04/28/staar-test-question-with-no-right-answer-among-dozens-of-complaints-as-texas-school-chiefs-say-results-can-t-be-trusted/

Bach, A. J. (2020). High-stakes, standardized testing and emergent bilingual students in Texas: A call for action. *Texas Journal of Literacy Education, 8*, 18-37.

Baker, O., &Lang, K. (2013). *The effect of high school exit exams on graduation, employment, wages and incarceration* (Working Paper No. 19182). National Bureau of Economic Research.

Beam-Conroy, T. (2001). Texas is the model for President Bush's education agenda. As this Texas teacher warns, Watch out! Your classroom may never be the same. *Rethinking Schools, 16*(1). https://rethinkingschools.org/articles/bamboozled-by-the-texas-miracle/

Biesta, G. (2014). Measuring what we value or valuing what we measure? Globalization, accountability and the question of educational purpose. *Pensamiento Educativo: Revista de Investigación Educacional Latinoamericana, 51*(1), 46-57. https://doi.org/10.7764/PEL.51.1.2014.5

Blakeslee, N. (2013, May). Crash test. *Texas Monthly*. https://www.texasmonthly.com/news-politics/crash-test/

Bland, L. M., & Gareis, C. R. (2018). Performance assessments: A review of definitions, quality characteristics, and outcomes associated with their use in K-12 schools. *Teacher Educators' Journal*, *11*, 52-69.

Blazer, D., & Pollard, C. (2017). Does test preparation mean low-quality instruction. *Educational Researcher, 46*(8), 420-433.

Boggs, A., Meyerhoff, E., Mitchell, N., & Schwartz-Weinstein, Z. (n.d.). Abolitionist university studies: An invitation. *Abolition University*. https://abolition.university/invitation/

Bradford, D. (2021). A rolling national teacher strike is why schools are closed. *Education Next*, *21*(3). https://www.educationnext.org/rolling-national-teacher-strike-is-why-schools-are-closed/

Braginsky, N. (2020). *Not an "achievement gap", a racial capitalist chasm.* The Law and Political Economy Project. https://lpeproject.org/blog/not-an-achievement-gap-a-racial-capitalist-chasm/

Carpenter, J. (2019, March 4). Lawmakers take up STAAR over wording of test questions. *Houston Chronicle.* https://www.houstonchronicle.com/newshouston-texas/ education/article/Lawmakers-take-up-STARR-over-wording-of-test-13662711.php

Carpenter, J. (2020, November 18). Bipartisan group of Texas lawmakers call on state to move toward canceling next year's STAAR tests. *Houston Chronicle.* https://www.houstonchronicle.com/news/houston-texas/education/article/Bipartisan-group-of-Texas-lawmakers-call-on-state-15736944.php

Chang, J. (2016, September 3). Texas districts document dozens of problems with 2016 STAAR testing. *Austin American-Statesman.* https://www.statesman.com/NEWS/20160903/Texas-districts-document-dozens-of-problems-with-2016-STAAR-testing

Cheek, D. W. (1993). Plain talk about alternative assessment. *Middle School Journal, 25*(2), 6-10. https://doi.org/10.1080/00940771.1993.11495197

Cole, M. (2009). *Equality in the secondary school: Promoting good practice across the curriculum*. A&C Black.

Conn, D. R., & Tenam-Zemach, M. (2019). Confronting the Assessment Industrial Complex: A call for a shift from testing rhetoric. *Journal of Curriculum Theorizing, 34*(3). https://journal.jctonline.org/index.php/jct/article/view/835

Counsell, S. L., & Wright, B. L. (2018). High-stakes accountability systems: Creating cultures of fear. *Global Education Review, 5*(2), 189-202.

Crocco, M. S. (2014, November 26). Green & Write: Texas, textbooks, and the politics of history standards. *Michigan State University School of Education*. https://education.msu.edu/green-and-write/2014/texas-textbooks-and-the-politics-of-history-standards/

Crocco, M. S., & Costigan, A. T. (2007). The narrowing of curriculum and pedagogy in the age of accountability urban educators speak out. *Urban Education, 42*(6), 512-535.

Curtis, G. (2019, May 29). El Paso families opting out of STAAR. *KFOX 14*. https://kfoxtv.com/news/kfox14-investigates/kfox14-investigates-el-paso-families-opting-out-of-staar

Darling-Hammond, L. (2007). Race, inequality and educational accountability: The irony of 'No Child Left Behind.' *Race, Ethnicity, and Education, 10*(3), 245-260. https://doi.org/10.1080/13613320701503207

Darling-Hammond, L. & Adamson, F. (2010). *Beyond basic skills: The role of performance assessment in achieving 21st century standards of learning*. Stanford Center for Opportunity Policy in Education. https://edpolicy.stanford.edu/sites/default/files/beyond-basic-skills-role-performance-assessment-achieving-21st-century-standards-learning-executive-summary_0.pdf

Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education*, *16*(5), 523-545. https://doi.org/10.1016/S0742-051X(00)00015-9

Davies, D.M. (Host). (1999-present). *Texas Matters* [Audio podcast]. Texas Public Radio. https://www.tpr.org/show/texas-matters/2020-02-14/texas-matters-the-great- texas-textbook-war

Davis, A. (2005). *Abolition democracy*. Seven Stories Press.

Dee, T. S., & Jacob, B. A. (2006). *Do high school exit exams influence educational attainment or labor market performance?* (No. w12199). National Bureau of Economic Research. https://doi.org/10.3386/w12199

DeMatthews, D. E., & Knight, D.S. (2019). The Texas special education cap: Exploration into the statewide delay and denial of support to students with disabilities. *Education Policy Analysis Archives*, *27*(2). https://eric.ed.gov/?id=EJ1202394

Deming, D. J., Cohodes, S., Jennings, J., & Jencks, C. (2016). When does accountability work? Texas system had mixed effects on college graduation rates and future earnings. *Education Next, 16*(1), 71-76.

Desai, K., & Sanya, B. N. (2016). Towards decolonial praxis: Reconfiguring the human and the curriculum. *Gender & Education*, *28*(6), 710-724. Educational Administration Abstracts.

Elmore, R. (2002). Unwarranted intrusion. *Education Next, (2)*1.

Fabricant, M., & Fine, M. (2013). *The changing politics of education: Privatization and the dispossessed lives left behind*. Paradigm Publishers.

Farber, S.A. (2008). U.S. scientists' role in the eugenics movement (1907-1939). *Zebrafish*, *5*(4), 243-245. https://doi.org/10.1089/zeb.2008.0576

Fine, M., & Pryiomka, K. (2020). *Assessing college readiness through authentic student work: How the City University of New York and the New York Performance Standards Consortium are collaborating toward equity*. Learning Policy Institute. https://eric.ed.gov/?id=ED606677

Fontana, J. (1995). Portfolio assessment: Its beginnings in Vermont and Kentucky. *NASSP Bulletin, 79*(573), 25-30. https://doi.org/10.1177/019263659507957305

Frey, B. B., Schmitt, V. L., & Allen, J. P. (2012). Defining authentic classroom assessment. *Practical Assessment, Research & Evaluation, 17*(2), 1-18.

Gagnon, D. J., & Schneider, J. (2019). Holistic school quality measurement and the future of accountability: Pilot-test results. *Educational Policy*, 33(5), 734-760. https://doi.org/10.1177/0895904817736631

Gisi, K. (2020, January 22). More Texas schools implementing project-based learning. *Spectrum News 1*. https://spectrumlocalnews.com/tx/san-antonio/news/2020/01/23/more-texas-schools-implementing-project-based-learning

Glennie, E., Bonneau, K., Vandellen, M., & Dodge, K. A. (2012). Addition by subtraction: The relation between dropout rates and school-level academic achievement. *Teachers College Record (1970), 114*(8), 1-26.

Grady, J., Marquez, R., & McLaren, P. (2012). A critique of neoliberalism with fierceness: Queer youth of color creating dialogues of resistance. *Journal of Homosexuality, 59*(7), 982-1004. https://doi.org/10.1080/00918369.2012.699839

Guisbond, L. (2014). *Testing reform victories: The first wave*. The National Center for Fair & Open Testing. https://www.fairtest.org/sites/default/files/TestingReformVictoriesReport.pdf

Guthrie, R. V. (1998). *Even the rat was white: A historical view of psychology* (2nd edition). Allyn & Bacon.

Hadavi, T. (2020, December 11). 2020 has shone a light on the importance of good teachers, but many are paid less than a living wage in the U.S. *CNBC*. https://www.cnbc.com/2020/12/11/why-teachers-salaries-are-so-low-in-the-us.html

Hagopian, J. (Ed.). (2015). *More than a score: The new uprising against high-stakes testing*. Haymarket Books.

Haney, W. (2000). The myth of the Texas miracle in education. *Education Policy Analysis Archives, 8*(0), 41. https://doi.org/10.14507/epaa.v8n41.2000

Heilig, J., Darling-Hammond, L. (2008). Accountability Texas-style: The progress and learning of urban minority students in a high-stakes testing context. *Educational Evaluation and Policy Analysis*, *30*, 75-110. https://doi.org/10.3102/ 0162373708317689

Holme, J. J., Richards, M. P., Jimerson, J. B., & Cohen, R. W. (2010). Assessing the effects of high school exit examinations. *Review of Educational Research*, *80*(4), 476-526. https://doi.org/10.3102/0034654310383147

Horelica, S. (2019, March 13). Texas mayor bans STAAR testing in his city. *KHOU News*. https://www.khou.com/article/news/education/texas-mayor-bans-staar-testing-in-his-city/285-001a733f-bb7e-4e15-bb8c-ba06658313c6

Jennings, J. L., & Bearak, J. M. (2014). "Teaching to the test" in the NCLB era: How test predictability affects our understanding of student performance. *Educational Researcher, 43*(8), 381-389. https://doi.org/10.3102/0013189X14554449

Jones, B.D. (2007). The unintended outcomes of high-stakes testing. *Journal of Applied School Psychology, 23*(2), 65-86. https://doi.org/10.1300/J370v23n02_05

Kamenetz, A. (2014, June 2). In Kentucky, moving beyond dependence on tests. *WUTC, The University of Tennessee at Chattanooga Public Radio NPR ED*. http://wutc.org/post/kentucky-students-succeed-without-tests

Karier, C. J. (1972). Testing for order and control in the corporate liberal state. *Educational Theory, 22,* 159-180.

Kendi, I.X. (2016a). Why the academic achievement gap is a racist idea. *Black Perspectives*. https://www.aaihs.org/why-the-academic-achievement-gap-is-a-racist-idea/

Kendi, I. X. (2016b). *Stamped from the beginning: The definitive history of racist ideas in America*. Nation Books.

Knecht, D. (2007). The consortium and the commissioner: A grass roots tale of fighting high stakes graduation testing in New York. *The Urban Review, 39*(1), 45-65. https://doi.org/10.1007/s11256-007-0043-0

Kevles D. J. (1999). Eugenics and human rights. *BMJ (Clinical research ed.)*, *319*(7207), 435-438. https://doi.org/10.1136/bmj.319.7207.435

Koh, K. H. (2017, February 27). Authentic assessment. *Oxford Research Encyclopedia of Education.* https://doi.org/10.1093/acrefore/9780190264093.013.22

Koretz, D. M. (1996). *Perceived effects of the Kentucky Instructional Results Information System (KIRIS)*. Rand Institute on Education & Training.

Koretz, D. (2005). Alignment, high stakes, and the inflation of test scores (CSE Report 655). *University of California Los Angeles.* https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.567.7619&rep=rep1&type=pdf

Kuhn, J. (2013). *Test-and-punish: How the Texas education model gave American accountability without equity.* Park Place Publications.

Kuhn, J. (2015). The word that made me an activist. In J. Hagopian (Ed.), *More than a Score* (245-252). Haymarket Books.

Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in US schools. *Educational Researcher*, *35*(7), 3-12.

Lemann, N. (2000). *The big test: The secret history of the American meritocracy*. Macmillan.

Lipman, P. (2003). *High stakes education: Inequality, globalization, and urban school reform*. Routledge. https://doi.org/10.4324/9780203465509

Lobman, C. (2014). "I feel nervous. . . very nervous" Addressing test anxiety in inner city schools through play and performance. *Urban Education, 49*(3), 329-359. https://doi.org/10.1177/0042085913478621

Longo-Schmid, J. (2016). Teachers' voices: Where policy meets practice. In K. Kappler Hewitt & A. Amrein-Beardsley (Eds.), *Student growth measures in policy and practice: Intended and unintended consequences of high-stakes teacher evaluations* (pp. 49-71). Palgrave Macmillan.

Love, B. (2019). *We want to do more than survive* (1st ed.). Beacon Press.

Lynch, S. J., Spillane, N. K., Peters Burton, E., Behrend, T. S., Ross, K. M., House, A., & Han, E. M. (2013). *Manor New Tech High School: A case study of an inclusive STEM-focused high school in Manor, Texas* (OSPrI Report 2013-01). George Washington University, Opportunity Structures for Preparation and Inspiration in STEM. https://ospri.research.gwu.edu/sites/g/files/zaxdzs2456/f/downloads/OSPrI_Report_2013-01.pdf

Mays, M. (2021, February 3). "I know this is a drastic step but ...": San Francisco sues its schools over closures. Politico, n.p. https://www.politico.com/states/california/story/2021/02/03/san-francisco-sues-school-district-over-closed-campuses-with-no-end-in-sight-1361744

McKay, C., Regunberg, A., & Shea, T. (2015). Testing assumptions: Zombies, flunkies, and the Providence Student Union. In J. Hagopian (Ed.), *More than a score* (135-140). Haymarket Books.

McNeil, L. M. (2005). Faking equity: High stakes testing and the education of Latino youth. In Angela Valenzuela (Ed.), *Leaving children behind: How "Texas-style" accountability fails Latino youth*. State University of New York Press.

McNeil, L. M., Coppola, E., Radigan, J., & Heilig, J. V. (2008). Avoidable losses: High-stakes accountability and the dropout crisis. *Education Policy Analysis Archives, 16*(3), 1-48. https://eric.ed.gov/?id=EJ800872

Menendez, J. (2019, April 13). The STAAR test is hurting our kids and we need to act now. *Texas Monthly*. https://www.texasmonthly.com/opinion/state-senator-jose-menendez-staar-test-hurting-kids-need-act-now/

Michels, Patrick (2014, March 27). Updated: Texas' testing opt-out movement grows to a 'trickle'. *The Texas Observer*. https://www.texasobserver.org/texas-testing-opt-movement-grows-trickle/

National Education Association (2020, June 25). History of standardized testing. https://www.nea.org/professional-excellence/student-engagement/tools-tips/history-standardized-testing-united-states

National Association of Secondary School Principals (2006). Tracking and ability grouping in middle level and high schools. https://www.nassp.org/tracking-and- ability-grouping-in-middle-level-and-high-schools/

Neill, M. (2003) Leaving children behind: how No Child Left Behind will fail our children, *Phi Delta Kappan, 85*(3), 225-228.

New York Performance Standards Consortium. (2018). Redefining assessment: Data report on the New York Performance Standards Consortium. https://drive.google.com/file/d/1VEb5XiDStR7nzFZxLc6WMMywOMAnCnSq/view

Nichols, S. L., Glass, G. V., & Berliner, D. C. (2006). High-stakes testing and student achievement: does accountability pressure increase student learning? *Education Policy Analysis Archives, 14*(1), 1-172.

Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Harvard Education Press.

Nichols, S., Glass, G., & Berliner, D. (2012). High-stakes testing and student achievement: Updated analyses with NAEP data. *Education Policy Analysis Archives, 20*(20), 1-35. https://doi.org/10.14507/epaa.v20n20.2012

Oakes, J. (1987). Tracking in secondary schools: A contextual perspective. *Educational Psychologist, 22*(2), 129-153. https://doi.org/10.1207/s15326985ep2202_3

Oakes, J. (2005). *Keeping track: How schools structure inequality*. Yale University Press.

Palmer, D., & Rangel, V. S. (2011). High stakes accountability and policy implementation: Teacher decision making in bilingual classrooms in Texas. *Educational Policy*, *25*(4), 614-647.

Public School Review. (n.d.). Success Academy Charter School-h.arlem 2. https://www.publicschoolreview.com/success-academy-charter-school-harlem-2-profile

Putwain, D., & Remedios, R. (2014). The scare tactic: Do fear appeals predict motivation and exam scores? *School Psychology Quarterly, 29*(4), 503-516. https://doi.org/10.1037/spq0000048

Putwain, D. W., & Symes, W. (2011). Teachers' use of fear appeals in the mathematics classroom: Worrying or motivating students? *British Journal of Educational Psychology, 81*(3), 456-474.

Raise Your Hand Texas. (2020). 2020: *Texans trust teachers.* https://www.raiseyourhandtexas.org/2020poll

Rosales, J. & Walker, T. (2021, March 20). The racist beginnings of standardized testing. *NEA Today*. https://www.nea.org/advocating-for-change/new-from-nea/racist-beginnings-standardized-testing

Rothstein, R. (2000, September 13). How tests can drop the ball. *The New York Times.* https://www.nytimes.com/2000/09/13/us/lessons-how-tests-can-drop-the-ball.html

Scott, Laurel L. (2012, December 12). Area civic leaders, school districts sign resolution opposing high-stakes testing: STAAR was to require students to take 15 end-of-course exams to graduate from high school. *Lubbock Avalanche Journal.* https://www.lubbockonline.com/article/20121210/NEWS/312109776

Shalaby, C. (2020, 2 June). *Repurposing our pedagogies* [Video]. Education for Liberation Network Facebook. https://www.facebook.com/watch/live/?v=260890484987341&ref=watch_permalink

Stanford School Redesign Network. (2008). *What is performance-based assessment?*https://edpolicy.stanford.edu/sites/default/files/events/materials/2011-06-linked-learning-performance-based-assessment.pdf

Steele, C. M. (1999, August 1). Thin ice: Stereotype threat and black college students. *The Atlantic.* https://www.theatlantic.com/magazine/archive/1999/08/thin-ice-stereotype-threat-and-black-college-students/304663/

Stoskopf, A. (2002). Echoes of a forgotten past: Eugenics, testing, and education reform. *The Educational Forum,* 66(2), 126-133.

Stoskopf, A. (2012). Racism in the history of standardized testing: Legacies for today. In Au, W. & Tempel, M.B. (Eds.), *Pencils down: Rethinking high-stakes testing and accountability in public schools* (1st ed., pp. 34-39). Rethinking Schools.

Sunderman, G. & Kim, J. (2004) *Inspiring vision, disappointing results: four studies on implementing the No Child Left Behind Act* (Cambridge, MA, Harvard Civil Rights Project).

Swartz, M. (2019a, March 13). Testing the test: New legislation filed to audit STAAR. *Texas Monthly*. https://www.texasmonthly.com/news/testing-the-test-new-legislation-filed-to-audit-staar/

Swartz, M. (2019b). Are Texas kids failing? Or are the tests rigged? *Texas Monthly*. https://www.texasmonthly.com/news-politics/texas-kids-failing-staar-tests-rigged/

Szabo, S., & Sinclair, B. (2012). STAAR reading passages: The readability is too high. *Schooling, 3*(1), 14.

Szabo, S., & Sinclair, B. (2019). Readability of the STAAR test is still misaligned. *Schooling, 10*(1). http://www.nationalforum.com/Electronic%20Journal%20Volumes/Szabo,%20Susan%20Readability%20of%20STARR%20%20is%20%20Misaligned%20Schooling%20V10%20N1,2019.pdf

Taylor, K. (2015, April 6). At Success Academy Charter Schools, high scores and polarizing tactics. *The New York Times.* http://www.nytimes.com/2015/04/07/nyregion/atsuccess-academy-charter-schools-polarizing-methods-and-superior-results.html

Terman, L. M. (1916). *The measurement of intelligence: An explanation of and a complete guide for the use of the Stanford Revision and Extension of the Binet-Stanford Intelligence Scale.* Houghton Mifflin.

Texas Parents Opt Out of State Tests (n.d.). *Home* [Facebook page]. Facebook. Retrieved July 1st, 2021, from https://www.facebook.com/optouttexas/

Texas Performance Standards Project. (n.d.). https://www.texaspsp.org/

Texas State Teachers Association. (16 March, 2020). *TSTA urges state to cancel STAAR testing this year.* https://tsta.org/wp-content/uploads/2020/03/PR0316cancelSTAAR.pdf

Thibodeaux, A. K., Labat, M. B., Lee, D. E., & Labat, C. A. (2015). The effects of leadership and high-stakes testing on teacher retention. *Academy of Educational Leadership Journal, 19*(1), 227-249.

Tienken, C. H., Colella, A., Angelillo, C., Fox, M., McCahill, K. R., & Wolfe, A. (2017). Predicting middle level state standardized test results using family and community demographic data. *RMLE Online*, *40*(1), 1-13. https://doi.org/10.1080/19404476.2016.1252304

Troy, G. (2016, June 26). The racist origins of the SAT. *The Daily Beast.* http://giltroy.com/the-daily-beast/the-racist-origins-of-the-sat/

Tuck, E. (2009). Suspending damage: A letter to communities. *Harvard Educational Review*, *79*(3), 409-428.

Tugend, A. (2019, October 9). Questioning their fairness, a record number of colleges stop requiring the SAT and ACT. *The Hechinger Report.* https://hechingerreport.org/questioning-their-fairness-a-record-number-of-colleges-stop-requiring-the-sat-and-act/

Urrieta, L. J. (2004). Assistencialism and the politics of high-stakes testing. *The Urban Review, 36*(3), 211-226.

Valli, L., Croninger, R., & Buese, D. (2012). Studying high-quality teaching in a highly charged policy environment. *Teachers College Record, 114*(4), 1-33.

Vogler, K. E. (2008). Comparing the impact of accountability examinations on Mississippi and Tennessee social studies teachers' instructional practices. *Educational Assessment, 13*, 1-32.

Walker, T. (2014). NEA Survey: Nearly half of teachers consider leaving profession due to standardized testing. http://neatoday.org/2014/11/02/nea-survey-nearly-half-of-teachersconsider-leaving-profession-due-to-standardized-testing-2/

Weiner, E. J. (2014). *Deschooling the imagination: Critical thought as social practice*. Routledge.

Wiggins, Grant. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. Jossey-Bass.

Appendix A

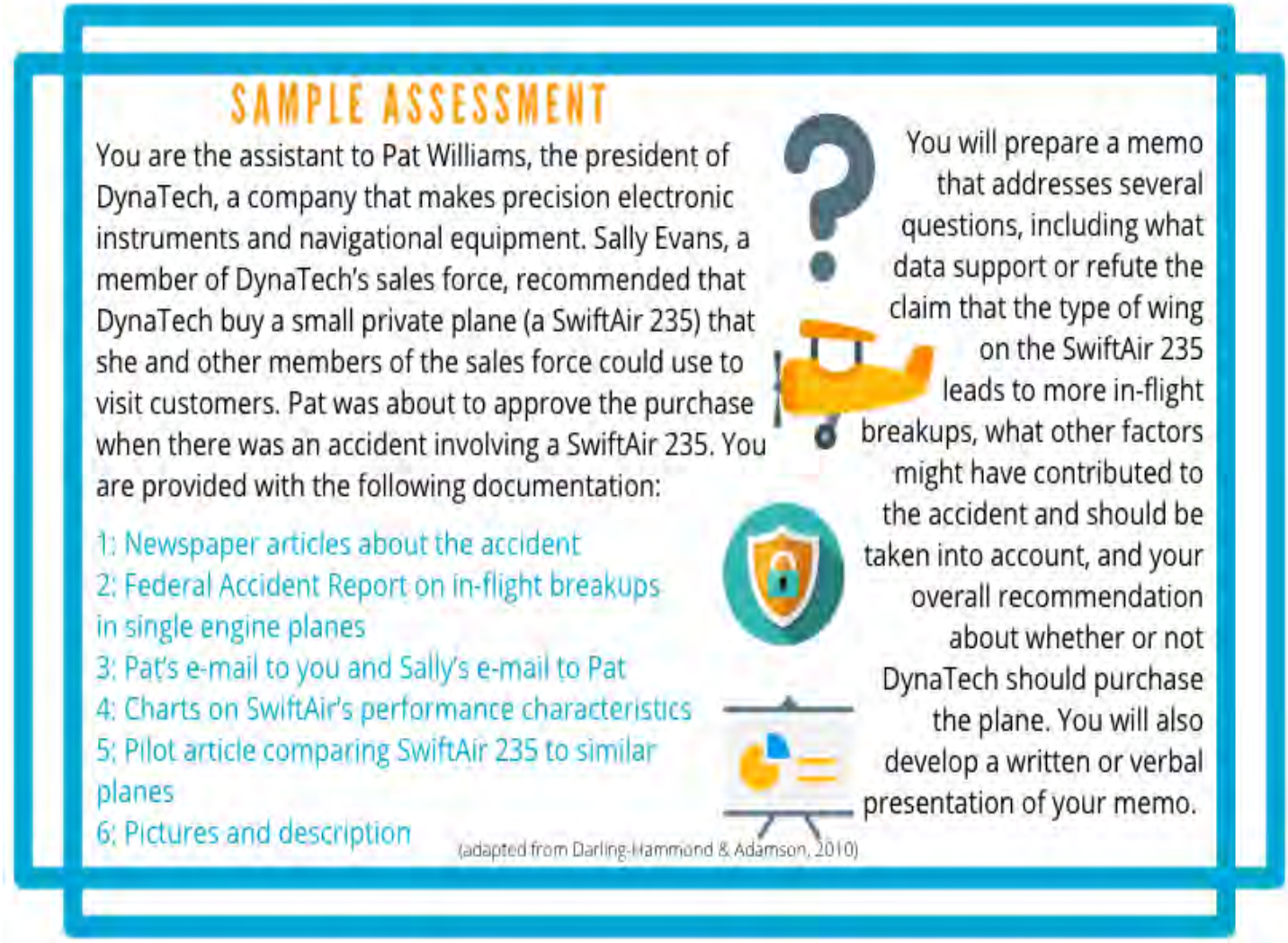
Table comparing authentic and traditional assessments

Typical tests	Authentic tasks	Indicators of authenticity
Require correct responses	Require a high-quality product or performance, and a justification of the solutions to problems encountered	Correctness is not the only criterion; students must be able to justify their answers.
Must be unknown to the student in advance to be valid	Should be known in advance to students as much as possible	The tasks and standards for judgment should be known or predictable.
Are disconnected from real-world contexts and constraints	Are tied to real-world contexts and constraints; require the student to "do" the subject.	The context and constraints of the task are like those encountered by practitioners in the discipline.
Contain items that isolate particular skills or facts	Are integrated challenges in which a range of skills and knowledge must be used in coordination	The task is multifaceted and complex, even if there is a right answer.
Include easily scored items	Involve complex tasks that for which there may be no right answer, and that may not be easily scored	The validity of the assessment is not sacrificed in favor of reliable scoring.
Are "one shot"; students get one chance to show their learning	Are iterative; contain recurring tasks	Students may use particular knowledge or skills in several different ways or contexts.
Provide a score	Provide usable diagnostic information about students' skills and knowledge	The assessment is designed to improve future performance, and students are important "consumers" of such information.

Note. Adapted from Wiggins, Grant. (1998). *Educative assessment: Designing assessments to inform and improve student performance* (pp. 21-42). Jossey-Bass.

Appendix B

Sample Authentic Assessment



Note. Adapted from Darling-Hammond, L. & Adamson, F. (2010). *Beyond basic skills: The role of performance assessment in achieving 21st century standards of learning* (p. 2). Stanford Center for Opportunity Policy in Education.