

# TEXAS EDUCATION REVIEW

Spring 2021 Volume 9, Issue 2

# **TABLE OF CONTENTS**

# Spring 2021: Volume 9, Issue 2

# Articles

Out of the Classroom and Less Likely to Graduate: The Relationship Between Exclusionary
Discipline and Four-Year Graduation Rates in Texas
Lenderman & Hawkins

# Special Issue: Texas 87th Legislature

Establishing Equity: Aligning Dual Language Bilingual Education to HB3 Sec. 11.1185 Texas Early Childhood Literacy & Mathematics Proficiency Plans
The Testing Industrial Complex: Texas and Beyond
Shooting for the STAAR: An Authentic Assessment Pilot Proposal to Replace Inequitable High- Stakes Accountability

# TEXAS EDUCATION REVIEW

# **SPRING 2021 EDITORIAL BOARD**

# **Managing Editors:**

Alex J. Armonda & Lebon D. James III

# **Editors:**

Will Davies Eliza Epstein Madeline Haynes Jase Kugiya Danielle Lindo Adam J. Martinez Hadiza Mohammed Heath Robinson Laura D. Turner Jaekyung B. Willows

# Welcome to Volume 9, Issue 2 (Spring 2021) of the Texas Education Review (TxEd)

Coinciding with the current Texas 87<sup>th</sup> Legislative Session, Issue 9 (2) of the Texas Education Review contains a Special Issue on the Texas 87<sup>th</sup>. The Special Issue is edited by William J. Davies, a Doctoral student in the Educational Leadership and Policy Studies Program. The issue contains one open-call article: a quantitative analysis of racialized trends in disciplinary practices in Texas schools (Lenderman & Hawkins). It also contains three articles as part of the Special Issue on the Texas 87<sup>th</sup>: a critical policy analysis on aligning dual language practices with Texas early childhood literacy and mathematics proficiency plans (Hernández & Núñez Porras); a policy analysis centering the troubling relationship between standardized testing in Texas and the school to prison pipeline (Del Carmen Unda & Lizárraga-Dueñas); and a critical policy review that links neoliberal highstakes accountability regimes to the historical context of the eugenics movement (Madrigal & Epstein).

# Information for Contributors

The Texas Education Review is an independent, peer reviewed, student-run scholarly publication based at the College of Education at The University of Texas at Austin. The Texas Education Review was founded and is operated by doctoral students at The University of Texas at Austin's College of Education, which consistently ranks as one of the best public university graduate education programs in the United States. The Texas Education Review aims to advance scholarship by publishing an academic journal of the highest quality including works by graduate students, professors, and practitioners, focusing on education policy and related issues. This journal features articles, essays, notes, and reviews relevant to a national and international audience of scholars and practitioners.

The Texas Education Review focuses on analysis of education policy and related issues, with nonexclusive preference given to issues affecting the State of Texas. Each issue shall display unparalleled excellence in content and style. Further, The Texas Education Review fosters the academic and professional development of its members through participation in the editorial process and each member displays the highest standards of integrity and professional excellence in every endeavor. From Sweatt v. Painter and No Child Left Behind, to charter schools, curriculum policy, and textbook adoption, the State of Texas has played and will continue to play a critical role in shaping education policy in the United States. The Texas Education Review is located directly on The University of Texas's campus in the heart of downtown Austin. Its close proximity to the Texas Capitol, Texas Education Agency, and State Board of Education offers unparalleled access to the thought leaders, policy makers, and academics who are driving education policy in Texas.



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



This work is licensed under a Creative Commons Attribution 4.0 International License. Permissions beyond the scope of this license may be available at <u>www.review.education.texas.edu</u>

# Out of the Classroom and Less Likely to Graduate: The Relationship Between Exclusionary Discipline and Four-Year Graduation Rates in Texas

KRISTIAN LENDERMAN

University of Houston

JACQUELINE HAWKINS University of Houston

**To cite this article:** Lenderman, K. & Hawkins, J. (2021). Out of the classroom and less likely to graduate: The relationship between exclusionary discipline and four-year graduation rates in Texas. *Texas Education Review*, *9*(2), 6-20. http://dx.doi.org/10.26153/tsw/13913

#### Out of the Classroom and Less Likely to Graduate: The Relationship Between Exclusionary Discipline and Four-Year Graduation Rates in Texas

#### KRISTIAN LENDERMAN University of Houston

#### JACQUELINE HAWKINS University of Houston

In the summer of 2019, Texas legislators passed SB 2432. The bill made it easier for schools to punish students using Disciplinary Alternative Education Programs (DAEPs), a form of exclusionary discipline that removes students from their educational environment (Swaby, 2019). In the 87<sup>th</sup> legislative session, new bills like Texas H.B. 1201 (2021), which died in committee, were attempting to curb the use, prevalence, and severity of various forms of exclusionary discipline placements for students through preventative and restorative practices. Many district leaders want guidance on how they should use exclusionary discipline, but the answer is not simple. This study builds upon previous research showing the negative impacts of exclusionary discipline on students and its impact on graduation.

The Texas Higher Education Coordinating Board has set the goal of having 60% of Texans earn a post-secondary credential by 2030 (Texas Higher Education Coordinating Board, 2017). Raising high school graduation rates by understanding the factors that affect them is vital to meeting this critical benchmark. The following paper examines data collected from the Texas Education Agency (TEA) that shows the relationship between single and multiple assignments to In-School Suspension (ISS), Out of School Suspension (OSS), and DAEPs and four-year high school graduation rates for students who were in ninth grade between 2011 and 2014 in Texas. This analysis shows both the disproportional use of exclusionary discipline for Black and Latinx students, and that students assigned to exclusionary discipline placements had lower rates of graduation. The study provides a deeper understanding of the connections between these issues to help inform policymakers in revising discipline protocols and implementation in schools.

#### Background

Many districts are rethinking their approach to discipline amid school safety concerns (Colombi & Osher, 2015). After the Columbine shooting in Colorado 20 years ago, many schools took a zero-tolerance approach that removed students from schools for offenses related to drugs, weapons, or signs of violence. Proponents of this strategy argue that removing disruptive students will help prevent further disruption and improve classroom climate. However, researchers on the American Psychological Association's Zero-Tolerance Task Force found that zero-tolerance approaches did not make schools safer (Reynolds et al., 2008). They found that instead of deterring future misbehavior, exclusionary discipline can increase rates of future misbehavior. Concerns about the negative impacts on students' mental health and increased exposure to the juvenile justice system have called into question the efficacy of zero-tolerance techniques. The following section reviews previous research that has shown the connection between the school-to-prison pipeline, exclusionary discipline practices, and the implications of exclusionary discipline on graduation rates.

#### School-to-Prison Pipeline

The school-to-prison pipeline describes the connection between exclusionary discipline used in schools and the justice system (Potter et al., 2017). Students of color – particularly Latinx and Black students – are more likely to be punished through exclusionary discipline actions, and they are more likely to experience involvement with the justice system. In a statewide study of Texas students, Fabelo et al. (2011) found that students who are expelled or suspended are three times more likely to be involved in the justice system. Furthermore, students with more than one disciplinary action were more likely to have juvenile justice contact (Fabelo et al., 2011).

Researchers have also established that the impact of the school to prison pipeline disproportionally falls on students of color. In a study of racial threat in schools, Payne and Welch (2010) found that schools with higher percentages of Black students have harsher discipline protocols. In another study by Skiba et al. (2011), researchers found that students of color were more often and more harshly disciplined than were White students for similar behaviors. Students in the 2015–2016 school year lost an estimated 11 million days of instruction across the country, with Black students losing the highest proportion of instructional days, 66 days of instruction per 100 students (Losen & Whitaker, 2018). In Texas, Fabelo et al. (2011) showed that Black students were more likely to receive a discretionary suspension than their White peers.

# **Discipline Use in Texas**

Exclusionary discipline takes many forms in the current educational landscape, including a range of tools from informal removals from classrooms and formal juvenile justice placements. The three most common forms of exclusionary discipline reported across Texas are In-School Suspensions (ISS), Out-of-School Suspensions (OSS), and Disciplinary Alternative Education Programs (DAEPs). Between the 2011 and 2014 school years, there were almost 8 million assignments to ISS, OSS, and DAEP, impacting 3.5 million Texas students (TEA, 2020). Nearly one in five students experienced exclusionary discipline across all grade levels (TEA, 2020). ISS, the least restrictive form of discipline tracked by Texas schools, allows students to remain on campus, but in a supervised area removed from other students. When assigned to OSS, students are removed from the school environment for a maximum of three consecutive days.

Students assigned to DAEPs are removed from their home campus and sent to another campus for a set number of days decided by administrators (Tex. Educ. Code Ann., 2017). While students at DAEPs receive instruction when they are away from their home classrooms, the time they are off campus is often longer than a typical in-school or out-of-school suspension. In 2008, the average DAEP stay for students across the state of Texas was 27 school days (Fabelo et al., 2011). After completing the assigned number of days, students return to their home campus. Texas law under Sec. 37.008 requires DAEPs to meet both the educational and behavioral needs of students but leaves the design mostly up to districts and school boards (Tex. Educ. Code Ann., 2017). The Texas Education Agency (TEA) has further clarified these rules by providing guidance on accountability measures, clarifying teacher training requirements, and requiring minimal transition procedures for students return to their home campuses with the skills needed to succeed, but exclusionary discipline assignments can reinforce and promote negative behaviors (McIntosh et al., 2008). Students pingpong back and forth from these alternative campuses to their home campus, disrupting academic and social ties. A previous statewide study found that students who were assigned to DAEPs once

returned to DAEPs about 20% of the time (Blackmon, 2016). Furthermore, in a study of two districts in Texas, students of color were more likely to return to a DAEP setting than were White students (Booker & Mitchell, 2011). Additionally, students who experience multiple forms of exclusionary discipline are even more likely to have negative outcomes in life, such as dropping out of school and having contact with the criminal justice system, than those who are assigned just once (Skiba et al., 2014).

#### Focus on Graduation

Amid heightened focus on student learning and progress, policy makers, parents, educators, and communities use graduation rates to measure school systems' success or failure. The Texas Higher Education Coordinating Board aims to have 60% of Texans ages 25-34 earn a certificate or degree by 2030 to increase the percentage of students gaining marketable skills for 21<sup>st</sup> century jobs (Texas Higher Education Coordinating Board, 2017). Texas's goals emphasize the importance of secondary high school completion because of the costs of high school dropouts. High school graduation has become critical for workforce participation, choice of future opportunities, and broader economic development. It is even more important now as we consider the impacts and implications of learning loss on students' attainment and potential.

Researchers have used ninth-grade outcomes as early warning signs to predict high school graduation rates. Allensworth (2013) found that ninth-grade performance predicted graduation rates correctly 80% of the time. Students who are retained in ninth grade have increased odds of dropping out of high school (Bornsheur et al., 2011). Students who do not graduate from high school make on average \$10,000 less per year than a high school graduate (Breslow, 2012). Before the pandemic, unemployment for high school dropouts was 12%, a third higher than for those with a high school diploma, and at the beginning of 2021 the employment gap had increased further (U.S. Department of Labor, 2021). Furthermore, incarceration rates were 63 times higher for young people who dropped out of high school compared to those with a college degree (Breslow, 2012).

There is an established relationship in the literature between exclusionary discipline and high school dropout rates. In their meta-analysis, Noltemeyer et al. (2015) found a statistically significant inverse relationship between suspensions and academic achievement. Furthermore, they found a statistically significant positive relationship between OSS and dropout rates (Noltemeyer et al., 2015). In another study of ninth graders, researchers found that after the first suspension in a student's ninth-grade year, an additional suspension increased their odds of dropping out by 20% (Balfanz et al., 2014). A study of the short and long-term outcomes of suspended students in New York City Public Schools showed that students who were suspended were less likely to graduate in four, five or six years compared to non-suspended peers (Chu & Ready, 2018). Not only are the impacts for individual students important to understand, but the effects on larger society are important to consider. Marchbanks et al. (2014) estimated the cost of dropouts associated with discipline in Texas between \$5.0 billion and \$9.0 billion in lost wages per cohort of students over their total lifetimes. The unintended consequences of exclusionary discipline should be considered when deciding how to shift the use of discipline for students and teachers.

#### Methods

Because of the connection between the ninth-grade school year and graduation outcomes (Allensworth, 2013; Balfanz et al., 2014), this study focused on students' experiences with exclusionary discipline in ninth grade. The research questions explored in this paper are as follows:

- 1. What relationship do single and multiple assignments to ISS, OSS, or DAEP in ninth grade have with students' four-year graduation rates? Do the resulting relationships between four-year graduation rates differ between types of exclusionary discipline?
- 2. Are certain student groups considering ethnicity, gender, special education services, and socioeconomic status more likely to experience exclusionary discipline through ISS, OSS, or DAEP once or more than once in their ninth-grade year?

This quantitative study used a descriptive design to determine the relationship between multiple placements in exclusionary discipline and four-year graduation rates. Data for students who experienced ISS, OSS, or DAEP assignments in ninth grade were analyzed according to ethnicity, gender, special education services, and economic status to evaluate whether these factors change their proportional representation in discipline assignments. Four-year graduation rates of students who were not disciplined once, and disciplined more than once in their ninth-grade year were compared using a descriptive analysis in Excel. The researchers focused on four-year graduation to measure the ninth-grade cohorts' on-time graduation rate. All graphs and figures were created using Excel. The analysis consisted of cleaning, compiling, and organizing the data to address the questions generated from literature review.

#### Sample

The data set tracked graduation outcomes for 1,669,391 first-time ninth graders who were enrolled in ninth grade in Texas public high schools between 2011 and 2014. The researchers used this time period to allow for enough years for each cohort to reach the four-year graduation mark. During this time period, there were on average 400,000 ninth-grade students each school year in the state of Texas. The researchers used data drawn from the records of students who began ninth grade in 2011, 2012, 2013, and 2014 and were expected to graduate in the spring of 2015, 2016, 2017, and 2018. This study accessed archival demographic data from TEA for ninth-grade students during this time. The researchers asked TEA for graduation outcome data for students who experienced ISS, OSS, or DAEP placements once or more than once through a Texas Public Information Act request submitted on the agency's online portal found on the TEA website. The researchers analyzed student graduation outcomes and demographic information for different groups based on disciplinary assignments. These data included information on the total number of ninth graders according to ethnicity, gender, economic status, special education, and at-risk status. The requested data included summaries of four-year graduation outcomes for students who were:

- Not disciplined
- Single placement in ISS
- Single placement in OSS
- Single placement in DAEP
- Multiple placements in ISS

- Multiple placements in OSS
- Multiple placements in DAEP

Students who left the state before their four-year graduation date were not included in the data set. Students who experienced the same discipline category more than once were not counted in the singular category. Because students who experienced different types of exclusionary discipline were not removed from the data set, the discipline categories were not mutually exclusive. To compare across all three types of discipline, only Black, Latinx, and White student populations had large enough sample sizes to examine. Due to the smaller number of Asian, Pacific Islander, Native Alaskan, and Native Hawaiian students placed in DAEP settings, the sample size was obscured because of FERPA requirements in the data set. The experiences of these students are extremely important and should be further examined using differentiated data in the future.

The analysis explored how multiple placements in a DAEP in ninth grade related to four-year graduation rates. The same relationships for students assigned to ISS and OSS were analyzed to demonstrate the relationship between suspensions and graduation rates. For the second analysis, data were compared by race and/or ethnicity, at-risk status, socioeconomic status, disability, and gender. Archival demographic data were compared with data from all ninth-grade students in Texas in order to show the proportional representation of students who experienced multiple placements at ISS, OSS, and DAEPs. This data was used to compare against the general student population instead of the more narrow confinements of the data requested.

#### Findings

The findings are organized by the two foci of the study: exploring the relationship between ninth grade discipline assignments and four-year graduation, and the relationship between demographics and exclusionary discipline assignments.

#### Graduation Rates

**Graduation rates by students disciplined or not disciplined.** Students who were not disciplined graduated at much higher rates than students who were. On average, 48% of students who were disciplined in the ninth-grade year at least once graduated in four years, compared to 84% of students who were not disciplined in their ninth-grade year. In addition, graduation rates for students who were only disciplined once were higher than for those disciplined more than once. Figure 1 shows the differing graduation rates for students who experienced a discipline action once and more than once. On average, the graduation rate decreased by 20 percentage points for students who experienced multiple disciplinary assignments.

Lenderman & Hawkins

# Figure 1





100%

*Note.* Graduation Rates by Disciplined Once and More, Texas Public Schools, 2011-2015. Adapted from Graduation Outcomes for Ninth Graders Placed in ISS, OSS, or DAEP Once or More received from TEA on January 24, 2020 in response to Texas Public Information Act.

**Graduation rates by type and amount of discipline.** More restrictive types of exclusionary discipline were associated with lower graduation rates. As shown in Figure 1, students assigned to ISS once had a 71% graduation rate, the highest rate within the discipline categories. That rate drops to 52% for students who were assigned more than once to ISS. About 57% of students who were assigned to OSS once graduated in four years, and 36% of students who were assigned multiple times did so. Compared to students assigned to ISS or OSS once, students assigned to DAEP one time were less likely to graduate. Only 44% of students assigned to DAEPs once graduated in four years. Students who were assigned to DAEPs multiple times had the lowest graduation rates (25%) out of the three disciplinary actions. These data highlight the potential negative impact that DAEPs can have on students' graduation outcomes and call into question the long-term effects on students of this tool of discipline.

# Student Demographics

**Discipline Rates by Student Demographics.** The second question explored in this paper focused on the relationship between ninth-grade student demographics and proportional representation of students assigned to ISS, OSS, and DAEP once or more than once. As shown in Figures 3 and 4, students who were disciplined were more likely to be Black and Latinx, compared to the proportions of students who were not disciplined. White students were less likely to be disciplined, compared to

their proportion in the general population. Furthermore, students who were disciplined were more likely to be economically disadvantaged, receiving special-education services, and male.

**Discipline rates by socioeconomic, at-risk, and special education status.** Students with disabilities, students considered at-risk, and economically disadvantaged students were overrepresented among disciplined students compared to the general student population. Students are *at-risk* if they fall into a variety of categories. For example, students can qualify as at-risk for not meeting standards on early standardized testing or experiencing homelessness or dropping out of school. As shown in Figure 2, students who were economically disadvantaged represented 60% of Texas ninth graders, but made up 77% of students disciplined. Although 9% of ninth grade students across Texas are students with disabilities, almost 16% of those who were disciplined had a disability. Similarly, half of the ninth graders in Texas are considered at-risk, but almost 70% of those disciplined were at-risk. Furthermore, for economically disadvantaged students, at-risk students, and students with disabilities, as the restrictiveness of the placements increased so did the disproportionate representation of these subgroups. In the student groups with more than one DAEP and OSS assignment, students receiving special education services were overrepresented by a factor of two compared to their proportion of the ninth-grade students.

# Figure 2



Texas Rates of Exclusionary Discipline for Economically Disadvantaged, Special Education and At-Risk Students

*Note.* Adapted from Graduation Outcomes for Ninth Graders Placed in ISS, OSS, or DAEP Once or More received from TEA on January 24, 2020 in response to Texas Public Information Act.

**Discipline rates by race/ethnicity.** Black and Latinx students, compared to the general student population shown in Figure 3, were overrepresented in receiving ISS, OSS and DAEP assignments both once and more than once in their ninth-grade years. Conversely, as shown in Figure 4, White students were underrepresented in every discipline category. Only Black, Latinx, and White students are included in the data analysis below because the sample size was obscured because of FERPA requirements in the data set

# Figure 3



#### 2011-2015 Texas Ninth Grade Population by Race/Ethnicity

Note. Adapted from Texas Education Agency. (2016). Enrollment in Texas public schools, 2015–16 (GE17 601 04). Retrieved from <u>https://tea.texas.gov/sites/default/files/enroll\_2015-16.pdf</u>

# Figure 4



Total Disciplinary Assignments by Ethnicity

*Note.* Adapted from Graduation Outcomes for Ninth Graders Placed in ISS, OSS, or DAEP Once or More received from TEA on January 24, 2020 in response to Texas Public Information Act.

White students made up about 30% of Texas ninth-grade students, but 23% or less of students disciplined. Black and Latinx students made up higher percentages of students with more than one discipline action in their ninth-grade year. Out of all ninth-grade students, Latinx students made up 51% of students during the time of this study, yet 54% of students assigned to ISS once and 57% of students assigned more than once to ISS were Latinx. Black students made up 11% of the Texas ninth grade population but made up 26% of students assigned to OSS once and 31% of the students assigned to OSS more than once. Conversely, lower percentages of White students had multiple disciplinary assignments. White students made up 30% of the ninth-grade population yet were only 21% of students assigned to DAEP once and 17% of students assigned more than once.

Not only did Black and Latinx students represent a higher proportion of students who were disciplined more than once, but as the restrictiveness of discipline increased from ISS to OSS to DAEP, so did the disproportionality of Black and Latinx students. Latinx students made up 54% of students assigned to ISS, but 58% of students assigned to DAEP. Similarly, Black students made up 19% of students assigned to ISS, but 22% of students assigned to DAEP. White students made up 23% of students assigned to ISS and only 21% of students assigned to DAEP, showing a proportional decrease in their assignments as the type of discipline becomes more exclusionary. Out of all discipline categories, OSS had the most disproportionate assignments for Black and Latinx students.

**Discipline Rates by Gender.** On average, 48.6% of Texas students in each cohort studied were assigned female, and 51.4% were identified as male. This balance drastically changed when looking at gender among students assigned to exclusionary discipline. Across all grade levels, the gender division that the data showed in those who were assigned to ISS, OSS, or DAEP in Texas was 30% female and 70% male (TEA, 2020). In the graduating cohort data, there is a higher proportion of male students disciplined as shown in Figure 5. Male students are overrepresented in every category of discipline, but male overrepresentation is more dramatic in multiple assignments than in single assignments. Furthermore, female students are least present in multiple OSS and multiple DAEP assignments.

# Figure 5



Texas Discipline Rates - Ninth Graders by Gender

*Note.* Discipline assignments by gender, Texas Public Schools, 2011–2015. Adapted from Graduation Outcomes for Ninth Graders Placed in ISS, OSS, or DAEP Once or More received from TEA on January 24, 2020 in response to Texas Public Information Act request.

**Summary of findings.** Using student demographic information in analyzing assignments to ISS, OSS, and DAEP, this study showed that groups of students who have been historically marginalized in the education system were more likely to experience exclusionary discipline than other groups of students. Students who were disciplined with exclusionary action were more likely to be male, Black, Latinx, at-risk, low income, or enrolled in special education programming. The more restrictive the form of exclusionary discipline category, the more disproportionality was present in each cohort year.

#### Limitations

The data set used in this study had several limitations. The first is that exclusionary discipline is one of many factors and variables that impact students' graduation rates, so we cannot determine the extent to which discipline alone is associated with graduation. Second, the data set used did not account for students who were assigned to different types of exclusionary discipline in the same year or students who were not first time ninth graders. This means some students may have been counted more than once in the data and others may have been included in the total population, but not the disciplined populations. Third, the data were shared as a report of averages of students, not individual data points. This data limited the types of analysis available to understand the information.

#### Discussion

Ninth grade has been established as a critical year for high school success in previous research, and this research adds to the literature showing the negative relationship between exclusionary discipline in ninth grade and graduation rates (Allensworth, 2013). Across Texas, students who were not disciplined had higher four-year graduation rates than did students who were disciplined. In all discipline categories, students who were disciplined graduated at lower rates than the state four-year average. Students in their ninth-grade year who had only one disciplinary assignment had higher four-year graduation rates than peers who experienced more than one assignment. The percentage of students graduating in four years dropped 20 percentage points from one assignment to more than one assignment across ISS, OSS, and DAEP. The decrease in graduation rates emphasizes the need to question the use of exclusionary discipline more than once in the same year as a form of behavior intervention. When looking at graduation rate by type of exclusionary discipline and demographic, students across all demographic groups assigned multiple times to DAEPs graduated at the lowest rates, aligning with previous research around the impact of exclusionary discipline used for ninth graders (Balfanz et al., 2014).

As found in other research, ninth-grade students who were disciplined were more likely to be Black, Latinx, and multiracial than were the sample who were not disciplined (Booker & Mitchell, 2011). This reflects the statewide and nationwide patterns of overrepresentation for Black and Latinx students in disciplinary assignments (Fabelo et al., 2011). In this study, Black and Latinx students, compared to the general population, were overrepresented in receiving ISS, OSS and DAEP assignments both once and more than once in their ninth-grade years across Texas. Lower percentages of white students had multiple disciplinary assignments. This reflects previous research showing that Black and Latinx students are disciplined more harshly and more often than their White counterparts (Fabelo et al., 2011).

Students with disabilities, students at-risk, and students who qualified as economically disadvantaged were overrepresented in the discipline analyses when compared to the general student population. Furthermore, for these students, the more restrictive the discipline assignment, the more disproportionality existed. This highlights the intersectional experience of students who are already at the margins of the educational experience being pushed further from opportunity through these exclusionary methods. These data emphasize the overrepresentation of vulnerable students in ninth-grade discipline rates, pointing to a bigger question about why and how exclusionary discipline is used in schools.

Finally, male students were overrepresented in every discipline category. The disproportionality follows the same pattern as other student characteristics discussed above. Male students made up 73% of students assigned to DAEP more than once, the most disproportional representation of all disciplinary actions. Previous research echoes these patterns in gender differences, showing the disparate impacts that exclusionary discipline has on Black male students (Losen & Whitaker, 2018). The data show that some students are more likely to be removed from their classroom and school than others.

# **Implications for Policy Makers**

Students who are disciplined with exclusionary methods in ninth grade have lower rates of four-year graduation. Considering the state's goal of having 60% of students with a post-secondary credential by 2030, the use of exclusionary discipline could be a barrier to meeting this goal, specifically for the groups of students who are more likely to experience exclusionary discipline based on their identities. These findings suggest that students in Texas who are disciplined via ISS, OSS, or DAEP in their ninth-grade year should be flagged for extra interventions and drop out preventions to support them in reaching the goal of four-year graduation. Policymakers may consider the disparate impact that the use of exclusionary discipline can have on specific groups of students who are more likely to experience discipline as they create guidelines, policy, and laws that either promote or hinder the use of these tools in schools because of their potential impact on graduation rates for students. The legislation proposed in the 87th session of the Texas legislature focused on implementing restorative programming that prompts students to talk about the underlying issue that caused the behavior, and to find ways to make things right with the harmed party. These restorative conversations can look like peer mediation, student conferences, or restitution circles, and have been shown in emerging research to reduce recidivism and assignments to exclusionary discipline (Rodriguez, 2007; Wilson et al., 2017; Anyon et al., 2016). Connecting the dots between the use of exclusionary discipline and accountability measures is an important step in considering the overall impact of discipline systems on the whole child.

#### Conclusion

As the COVID-19 pandemic has emphasized the importance of in-class learning and a new wave of bills addressing exclusionary discipline entered the 87th Texas legislature, critically examining the use of discipline in schools is extremely timely. Students assigned to DAEPs in their ninth-grade year had lower rates of graduation than students experiencing other forms of discipline. Furthermore, the data shows the overrepresentation of Latinx and Black students when it comes to the more exclusionary discipline assignments. Healthy graduation rates are a foundation of a healthy economy in Texas, as the state faces growing challenges. As leaders create new rules and expectations to meet health guidelines, educators and practitioners should examine their use of exclusionary discipline to understand the long-term consequences of its implementation.

**Dr. Kristian Lenderman** is a former educator who supports teachers and school leaders in building inclusive and just learning environments. Her research focuses broadly on school climate, equity, and teacher training. She received her Ed.D. from the University of Houston in 2020.

**Dr. Jacqueline Hawkins** currently serves as the Special Populations Ed.D. Program Lead in the Department of Educational Leadership and Policy Studies at the University of Houston. Dr. Hawkins' research focuses on prevention and intervention work for educators and parents who engage with students in applied contexts across the education pipeline. She engages a systems change approach to research professional development, transition support, and early interventions for students who strive for better outcomes.

#### References

- Allensworth, E. (2013). The use of ninth-grade early warning indicators to improve Chicago schools. *Journal of Education for Students Placed at Risk (JESPAR)*, 18(1), 68-83. https://doi.org/10.1080/10824669.2013.745181
- Anyon, Y., Gregory, A., Stone, S., Farrar, J., Jenson, J., McQueen, J., Downing, B., Greer, E., & Simmons, J. (2016). Restorative interventions and school discipline sanctions in a large urban school district. *American Educational Research Journal*, 53(6), 1663-1697. <u>http://dx.doi.org/10.3102/0002831216675719</u>
- Balfanz, R., Byrnes, V., & Fox, J. (2014). Sent home and put off-track: The antecedents, disproportionalities, and consequences of being suspended in the ninth grade. *Journal of Applied Research* on Children: Informing Policy for Children at Risk, 5(2). <u>https://digitalcommons.library.tmc.edu/childrenatrisk/vol5/iss2/13/</u>
- Blackmon, R. (2016). Transitioning students from a disciplinary alternative education program to a regular campus. Legislative Budget Board Staff. <u>https://www.lbb.state.tx.us/Documents/Publications/Issue\_Briefs/3091\_Transitioning\_Students\_from.pdf</u>
- Booker, K., & Mitchell, A. (2011). Patterns in recidivism and discretionary placement in disciplinary alternative education: The impact of gender, ethnicity, age, and special education status. *Education & Treatment of Children*, 34(2), 193-208. <u>https://doi.org/10.1353/etc.2011.00166</u>
- Bornsheuer, J., Polonyi, M., Andrews, M., Fore, B., Onwuegbuzie, A. (2011). The relationship between ninth-grade retention and on-time graduation in southeast Texas high school. *Journal of At-Risk Issues, 16*(2), 9-16. <u>https://files.eric.ed.gov/fulltext/EJ960072.pdf</u>
- Breslow, J. (2012). By the numbers: Dropping out of high school. *PBS: Frontline*. <u>https://www.pbs.org/wgbh/frontline/article/by-the-numbers-dropping-out-of-high-school/</u>
- Chu, E.M., & Ready, D.D. (2018). Exclusion and Urban Public High Schools: Short- and Long-Term Consequences of School Suspensions. *American Journal of Education, 124*, 479-509.
- Colombi, G. & Osher, D. (2015). Education Leaders Report: Advancing School Discipline Reform. National Assocation of State Boards of Education. <u>https://www.air.org/sites/default/files/downloads/report/Advancing-School-Discipline-Reform-Sept-2015.pdf</u>
- Disciplinary Alternative Education Programs, Tex. Educ. Code Ann., § 37.008 (2017). <u>https://stat-utes.capitol.texas.gov/Docs/ED/htm/ED.37.htm</u>
- Fabelo, T., Thompson, M., Plotkin, J. D., Carmichael, D., Marchbanks, M., & Booth, E. (2011). Breaking schools' rules: A statewide study of how school discipline relates to students' success and juvenile justice system involvement. Council of State Governments Justice Center. <u>https://csgjusticecenter.org/wpcontent/uploads/2012/08/Breaking\_Schools\_Rules\_Report\_Final.pdf</u>
- H.B. No. 1201, 87th Texas Legislature, 2021 Reg. Sess. (Tex. 2021). <u>https://capitol.texas.gov/tlo-docs/87R/billtext/pdf/HB012011.pdf#navpanes=0</u>
- Kuhfeld, M., Tarasawa, B., Johnson, A., Ruzek, E., & Lewis, K. (2020). Learning during COVID-19: Initial findings on students' reading and math achievement and growth. NWEA. <u>https://www.nwea.org/content/uploads/2020/11/Collaborative-brief-Learning-during-COVID-19.NOV2020.pdf</u>
- Losen, D., & Whitaker, A. (2018). *11 million days lost: race, discipline, and safety at U.S. public schools—Part 1.* American Civil Liberties Union of Southern California. <u>https://www.aclu.org/sites/de-fault/files/field\_document/final\_11-million-days\_ucla\_aclu.pdf</u>

- Marchbanks, M., Blake, J., Smith, D., Seibert, A., Carmichael, D., Booth, E., & Fabelo, T. (2014). More than a drop in the bucket: The social and economic costs of dropouts and grade retentions associated with exclusionary discipline. *Journal of Applied Research on Children: Informing Policy for Children at Risk, 5*(2). <u>http://digitalcommons.library.tmc.edu/childrenatrisk/vol5/iss2/17</u>
- McIntosh, K., Horner, R. H., Chard, D. J., Dickey, C. R., & Braun, D. H. (2008). Reading skills and function of problem behavior in typical school settings. *Journal of Special Education*, 42, 131-147. doi: <u>10.1177/0022466907313253</u>
- Noltemeyer, A. L., Ward, R. M., & Mcloughlin, C. (2015). Relationship between school suspension and student outcomes: A meta-analysis. *School Psychology Review*, 44(2), 224-240. <u>https://doi.org/10.17105/spr-14-0008.1</u>
- Potter, H., Boggs, B., & Dunbar, C. (2017). "Discipline and punishment: How schools are building the school-to-prison pipeline." *The School to Prison Pipeline: The Role of Culture and Discipline in School* (pp. 65-90). Emerald Publishing Limited. <u>https://doi.org/10.1108/S2051-231720160000004005</u>
- Reynolds, C., Graham, S., Skiba, R., Sheras, P., Conoley, J., & Garcia-Vazquez, E. (2008). Are zero tolerance policies effective in the schools? *American Psychologist*, 63, 852-862. <u>dx.doi.org/10.1037/0003-066X.63.9.852</u>
- Rodriguez, N. (2007). Restorative justice at work: Examining the impact of restorative justice resolutions on juvenile recidivism. *Crime & Delinquency*, *53*(3), 355-379. <u>https://doi.org//10.1177/0011128705285983</u>
- Skiba, R., Arrendondo, M., & Williams, N. (2014). More than a metaphor: The contribution of exclusionary discipline to a school-to-prison pipeline. *Equity & Excellence in Education*, 47(4), 546-564. <u>https://doi.org/10.1080/10665684.2014.958965</u>
- Swaby, A. (2019, July 24). Texas just made it easier to punish students who harass teachers. Will the law be misused? *Texas Tribune*. <u>https://www.texastribune.org/2019/07/24/texas-made-it-easier-punish-students-who-harass-teachers/</u>
- Texas Education Agency. (2016). Enrollment in Texas public schools, 2015–16 (GE17 601 04). https://tea.texas.gov/sites/default/files/enroll\_2015-16.pdf

Texas Education Agency. (2020). Counts of students and discipline actions by discipline action groups—PEIMS 2018-2019 [Dataset]. Texas Education Agency, Division of Research and Analysis. https://rptsvr1.tea.texas.gov/adhocrpt/Disciplinary\_

Data\_Products/DAG\_Summaries/Download\_State\_DAG\_Summaries.html

- Tex. Educ. Code Ann. § 37.008 (2017)
- Texas Higher Education Coordinating Board. (2017). 60 x 30 TX. http://www.60x30tx.com/
- U.S. Department of Labor. (2021). The Employment Situation. *Bureau of Labor Statistics*. <u>https://www.bls.gov/news.release/pdf/empsit.pdf</u>
- Wilson, D., Olaghere, A., & Kimbrell, C. S. (2017). Effectiveness of restorative justice principles in juvenile justice: A meta-analysis. U.S. Department of Justice. https://www.ncjrs.gov/pdffiles1/ojjdp/grants/250872.pdf



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





This work is licensed under a Creative Commons Attribution 4.0 International License. Permissions beyond the scope of this license may be available at <u>www.review.education.texas.edu</u>

# Establishing Equity: Aligning Dual Language Bilingual Education to HB3 Sec. 11.1185 Texas Early Childhood Literacy & Mathematics Proficiency Plans

PATRICIA NÚÑEZ PORRAS The University of Texas at Austin

JULIA HERNÁNDEZ

University of Texas at Austin

**To cite this article:** Núñez Porras, P. & Hernández, J. (2021). Establishing Equity: Aligning Dual Language Bilingual Education to HB3 Sec. 11.185 Texas Early Childhood Literacy & Mathematics Proficiency Plans. *Texas Education Review*, *9*(2), 22-29. http://dx.doi.org/10.26153/tsw/13912

#### Establishing Equity: Aligning Dual Language Bilingual Education to HB3 Sec. 11.1185 Texas Early Childhood Literacy & Mathematics Proficiency Plans

PATRICIA NÚÑEZ PORRAS The University of Texas at Austin

#### JULIA HERNÁNDEZ The University of Texas at Austin

#### **Executive Summary**

To the chair of the Education Committee, this policy brief provides recommendations that address dual language bilingual education and the House Bill 3 (HB3) Texas Early Childhood Literacy & Mathematics Proficiency Plans requirement. How can the HB3 Texas Early Childhood Literacy & Mathematics Proficiency Plan requirement set appropriate reading goals for dual language students designated as English Learners? These recommendations are attentive to the unique trajectory of developing literacy in two languages (biliteracy) and accurately report students' biliteracy growth. Adopting the stated recommendations will ensure valid measures of students' literacies, establish a fairer and more effective accountability system, and drive an accurate appropriation of funds to support robust dual language bilingual education implementation. It is noteworthy to state that the recommendations are consistent with the Education Finance Committee's core principle that students meet the state's educational expectation of 60 percent of students meeting the third grade reading standard by 2030 (TEA, 2018).

#### Inequitable Literacy Metrics & Bilingual Learners

Currently, Texas serves over one million students designated as English Language learners, emergent bilinguals. The population of emergent bilingual students has increased significantly from 800,554 in the 2008- 09 school year (Latham Sikes, C., & Villanueva, C., 2021). Nationally, Texas has also held onto its years-long lead as the state that has experienced the largest annual numeric increase of Hispanic residents since 2010 (Ura, A., Ahmed, N., 2018). "As of the 2019-2020 school year, EL students in pre-K through third grade comprised 44% of all EL students in the public education system. English learners in the elementary grades (preK-5th) make up 62% of all identified ELs in Texas schools" (Ura, A., Ahmed, N., 2018). Following a failure to reform the state's school finance formula in a special session during 2017, the Commission on Public School Finance was instated by and submitted a report with recommendations to the legislature before the start of the 86<sup>th</sup> (Texas Education Agency, n.d.). Among these recommendations were proposed new allotments and programs to improve early literacy including one for Dual Language education programming (TEA, 2017).

Historically, reporting literacy development for students in bilingual education programs, including One and Two-Way dual language programs<sup>1</sup>, has been problematic (Arteagoitia, I., & Yen, S. J.

<sup>&</sup>lt;sup>1</sup> Dual Language education refers to the bilingual program that provides grade level content and literacy in two languages, English and another program language. One-Way Dual Language programs are configured by participating students that are designated as English Learners. Two-Way Dual Language programs serve two student groups, students designated as English learners and students who are monolingual or dominant in English at the time of enrollment (Howard, 2018).

2020; Valdés, G., & Figueroa, R.A. 1994). Arteagoitia and Yen (2020) brought attention to the issue when they explained that "a mere monolingual lens...completely disregards the competencies children who speak a language other than English may have in that language or languages." Districts require and track literacy growth in only one of the two program languages, traditionally English. The screeners, diagnostic evaluations, and assessments administered in these programs use metrics that are monolingually-normed for English. The use of monolingual metrics produces systematic measurement error and seriously affects emergent bilinguals' outcomes by limiting bilingual students' ability to demonstrate their full literacies (Abedi & Linquanti 2012, Solano-Flores, 2016). The misalignment occurs when monolingual metrics are used to measure the literacies of emergent bilingual students. To ensure equity in dual language bilingual education, both program languages must be recognized in assessment practices to capture the wide range of knowledge and skills held across both program languages, support dual language bilingual education goals, and adequately address students' academic needs (Arteagoitia, I., & Yen, S. J., 2020).

The conventional narrative during the COVID-19 pandemic is that students, especially students of color, will experience a considerable amount of learning loss (Taboada, M. B. 2020; Dorn, E., Hancock, B., Sarakatsannis, J., & Viruleg, E. 2020). Monolingual assessment practices and the use of monolingual screeners, diagnostic evaluations and assessments will only exacerbate the erroneous perception of emergent bilinguals as underperforming. This underlines the need to change the current system of assessment practices, tools, and reporting that fail to both accurately capture the existing literacy-related aptitudes that emergent bilinguals already possess and allow emergent bilinguals to fully demonstrate their biliteracy growth.

Inequitable monolingual assessment practices reinforce and are reinforced by deficit frameworks. These frameworks, with a long history and still prominent today, have been institutionalized in our systems of education and often drive educational thought and practice (Valencia, R., 1997). Schooling within these systems has pathologized the language and cultural practices of minoritized groups. Influential and problematic studies like Hart and Risley (1995), whose "word gap" research concluded that the linguistic "deficiencies" were the cause of academic failures have provided the modern foundations of these longstanding deficit frameworks (Dudley- Marling, C. & Lucas, K., 2009). These studies have resulted in and continue to shape the educational structures and systems that reflect monolingualism and its norms as the only standard for all students.

# Early Childhood Literacy & Mathematics Proficiency Plan: "Trajectories Toward Biliteracy"

The misalignment of assessment metrics for emergent bilinguals has long had serious detrimental consequences (Valdés, G., & Figueroa, R.A. 1994). One of these, known as *consequential validity*, refers to specific contexts where results are used to falsely identify emergent bilinguals as "struggling." The "struggling" label has exacerbated inequitable education design for emergent bilinguals which denies access to high-quality learning experiences. This perpetuates the "increase in rote learning," which has led to *a "decline* of teaching methods that encourage higher-order thinking skills, the writing of essays, and conducting research" (García, 2009, p.369; see also Glaser, 1990).

Given such inequitable effects on the learning experiences of emergent bilingual students, the current assessment practices for measuring emergent bilinguals' literacy growth must change. Students in dual language bilingual education are developing literacy in two languages simultaneously (biliteracy) and will have skills that display themselves in the use of both of their languages. Therefore, equitable assessment practices for emergent bilinguals would capture literacy growth in both program languages. A biliteracy approach for measuring literacy growth will allow students to demonstrate what they know across their two languages, therefore providing teachers, policy makers and students themselves a more holistic and accurate picture of their literacy capabilities.

Acquisition of literacy in multiple languages requires reading goals that differ from monolingual (only-target language or only-English) literacy development, not because they are unattainable but because the progression of literacy development is *different*. One promising approach, "Trajectories Toward Biliteracy," includes reading goals in both program languages established as "targeted zones." Hopewell and Escamilla's (2010) findings reposition emergent bilinguals (designated as ELs), as meeting literacy goals where before monolingually-normed assessments deemed them to be "struggling." The table below shows the results of a literacy study conducted with a population of third-grade emergent bilinguals. When data in only one program language was analyzed, 84% of 3rd grade emergent bilinguals were flagged as underperforming and requiring individual support. However, when *both* the English and Spanish literacy scores were collected to measure literacy development, only 40% of the students met criteria for needing individual support. This meant that an overwhelming 60% percent of students actually met the third grade literacy standard. The sharp contrast in this analysis reveals the importance of considering students' full linguistic resources to determine literacy proficiency.

# Table 1

	English Language scores only— DRA2	Spanish language read- ing scores only —EDL2	Trajectory toward biliteracy- EDL2 + DRA2
% requiring ILP	83.6% (n= 224)	55.3% (n=148)	39.6% (n=106)
% not requiring ILP	16.4% (n=44)	44.7% (n=120)	60.4% (n=162)

Students Requiring Individual Literacy Plan (Hopewell and Escamilla, 2013)

*Note.* ILP = Individual Literacy Plan; EDL = *Evaluación del Desarrollo de Lecto-escritura*; DRA = Developmental Reading Assessment.

The dual-language education goal of attaining high proficiency levels of bilingualism and biliteracy calls for an appropriate way to measure progress (Howard, 2018). Starting from the idea that emergent bilingual students draw on all their linguistic resources to develop literacy, Escamilla and Hopewell (2010; 2014) provide an important model for assessment that normalizes metrics of literacy development in and between students' two languages. In dual-language programs, languages and literacies develop cohesively and in reciprocal ways— put another way, they develop bidirectionally, expanding on each other.

# **Existing Policy**

According to HB3 (2019), Texas Early Childhood Literacy & Mathematics Proficiency Plan, (Sec. 11.185) districts are required to design a plan to report data collection on students' literacy growth by identifying specific annual reading goals in kindergarten through third grade. Current policy allows a district's literacy proficiency plan to set separate goals for students in a bilingual education or special language program. While the policy clearly acknowledges the need to set and align goals specific to a students' program enrollment, the policy would be strengthened by *requiring* separate goals for one- and two-way dual language bilingual education programs that include both English and the target language.

Setting reading goals in *both* program languages must be prioritized to improve the schooling opportunities and experiences of emergent bilinguals. The need to differ from a monolingual literacy trajectory is evident, not because the monolingual goals are unattainable, but because the trajectory of biliteracy is *different*. For educators and policymakers, recognizing and leveraging key differences between monolingualism and bilingualism shifts the pervasive deficit view of emergent bilinguals' academic potentials to a more asset-based frame. This shift impacts curriculum and pedagogy in ways that have profound effects on teaching and learning.

# Accurate Assessment Leads to Effective Spending

In the last session, the 86th Texas legislature placed heavy emphasis on resourcing the education of students identified as economically disadvantaged and English learners (TEA 2019). House Bill 3 was passed and sets various allotments, including a dual language education funding allotment. Robust dual language programs are costly (Lara-Alecio et al., 2005) and the dual language allotment requires attaching compliance and monitoring measures to support dual language implementation (TECELI, 2021). These measures also serve as criteria to allocate funds appropriately for emergent bilinguals. Well-implemented monitoring and compliance measures give emergent bilinguals access to high-quality learning environments that appreciate and foster high proficiency levels of bilingual-ism and biliteracy.

Hopewell and Escamilla (2014) explain that in an era of monetary shortfalls in public education, resources could be better used to develop appropriate assessments as opposed to using flawed ones that systematically misdiagnose youth and the unnecessary interventions that are developed as a response. Furthermore, monolingual literacy assessments that are currently used to measure the literacy development of bilingual students are costly and further burden districts with unnecessary expenses (Hopewell & Escamilla 2014).

# Figure 1

Cost of falsely identifying developing bilingual students as underperforming:
<ul> <li>purchasing unnecessary remedial reading programs</li> <li>allocating funds for unnecessary intervention staff</li> <li>providing remedial after school tutoring programs vs. enrichment</li> <li>offering remedial summer school programing vs. enrichment</li> </ul>

#### **Policy Recommendations**

In order for Texas to adequately serve emergent bilinguals, it must write policies that systematically safeguard equity. The following recommendations ensure equity and strengthen current policy by guiding districts to use metrics that allow developing bilinguals to fully access and demonstrate their literacy skills across multiple languages. These recommendations are based on a robust body of research and the known benefits of bilingualism.

#### Mandate Annual Biliteracy Goals

Mandate district's literacy proficiency plan to develop separate goals for dual language programs and establish a bilingual trajectory. Districts must identify specific annual reading goals in both program languages (Kindergarten-third grade) based on instruments being used (Hopewell & Escamilla 2010, 2014).

#### Require Accountability Measures in Both Program Languages

Require one- and two-way dual language programs to report annual student reading growth in both of the program languages. This mandate ensures accountability of dual language implementation (Hopewell & Escamilla 2010, 2014; TECELI, 2021).

These recommendations serve as safeguards to the overidentification of emergent bilinguals as underperforming and ensure that districts capitalize on the students' full linguistic abilities with allocated funds used as intended to provide an equitable education for all students.

#### Conclusion

In summary, the current 87<sup>th</sup> Texas legislative session can amend existing policy through revisions that increase the accuracy of literacy growth measurement, reduce costs, and strengthen the fidelity of dual language bilingual education program implementation.

The recommendations place an additive lens on policy. Such a policy would ensure students are able to showcase the academic advantages that their bilingualism and biliteracy afford them (Bialystok, E., Craik, F. I., & Luk, G., 2012). Ultimately, these policy recommendations reframe the current deficit narrative students designated as English Learners have historically endured due to the misalignment of assessment practices. The Texas Commission on Public School Finance set forth a goal of 60 percent of the student population meeting reading proficiency by 2030. The recommended policy amendments promise to support such an initiative while highlighting the benefits of robust and faithfully implemented dual language bilingual education which benefits all Texans.

**Julia Hernández, M.Ed.,** is a doctoral student in the Educational Leadership and Policy program in the Department of Curriculum and Instruction at the University of Texas at Austin. She is a Latina educator in the field of bilingual education and part of various community dual language bilingual education advocacy organizations. Her research focuses on equitable education of emergent bilingual children in the US that includes policy for social justice, culturally sustaining pedagogies and the view of bilingualism and biliteracy development.

**Patricia Núñez, M.Ed.**, is a doctoral student in Cultural Studies in the Department of Curriculum and Instruction at the University of Texas at Austin. She is a long time Latina educator working in México and the U.S. and continues to work in U.S. public schools in the field of bilingual education. Her research focuses on highlighting the strengths of bilingual communities and students and the revitalization of present and heritage learning practices of the home. She is an active community advocate for equitable access to quality educational programs for traditionally underserved students.

#### References

- Abedi, J., & Linquanti, R. (2012). Issues and opportunities in improving the quality of large scale assessment systems for English Language Learners. Stanford University. <u>https://ell.stanford.edu/sites/de-fault/files/pdf/academic-papers/07-Abedi%20Linquanti%20Issues%20and%20Opportuni-ties%20FINAL.pdf</u>
- Arteagoitia, I., & Yen, S. J. (2020). Equity in representing literacy growth in dual language bilingual education for emerging bilingual students. *TESOL Quarterly*, *54*(3), 719-742. <u>https://doi.org/10.1002/tesq.588</u>
- Bialystok, E., Craik, F. I., & Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends in Cognitive Sciences*, 16(4), 240-250. <u>https://doi.org/10.1016/j.tics.2012.03.001</u>
- Dorn, E., Hancock, B., Sarakatsannis, J., & Viruleg, E. (2020). COVID-19 and student learning in the United States: The hurt could last a lifetime. McKinsey & Company. <u>https://www.apucis.com/frontend-assets/porto/initial-reports/COVID-19-and-student-learning-in-the-United-States-FINAL.pdf.pagespeed.ce.VHbS948yF4.pdf</u>
- Dudley-Marling, C., & Lucas, K. (2009). Pathologizing the language and culture of poor children. Language Arts, 86(5), 362-370.
- Escamilla, K., Hopewell, S., Butvilofsky, S., Sparrow, W., Soltero-González, L., Ruiz Figueroa, O., & Escamilla, M. (2014). *Biliteracy from the start: Literacy squared in action*. Caslon Publishing.
- García, Ofelia. (2009). Bilingual education in the 21st century: A global perspective. Blackwell Publishing.
- Glaser, R. (1990). *Testing and assessment: O tempora! O mores.* University of Pittsburg, Learning Research and Development Center.
- Hopewell, S., & Escamilla, K. (2014). Struggling reader or emerging biliterate student? Reevaluating the criteria for labeling emerging bilingual students as low achieving. *Journal of Literacy Research*, 46(1), 68-89. <u>https://doi.org/10.1177/1086296X13504869</u>
- Hopewell, S. & Escamilla, K. (2014). Biliteracy development in immersion contexts. *Journal of Immersion and Content-based Language Education*, 2(2), 181-195. doi: <u>10.1075/jicb.2.2.02hop</u>
- Howard, E. R., Lindholm-Leary, K. J., Rogers, D., Olague, N., Medina, J., Kennedy, B., Sugarman, J., & Christian, D. (2018) *Guiding principles for Dual Language Education* (3<sup>rd</sup> ed.). Center for Applied Linguistics.
- Intercultural Development Research Association. (April 2019). Policy brief: Most English Learners would be excluded from the proposed Dual Language weight. <u>https://www.idra.org/wp-content/uploads/2019/04/Most-English-Learners-Would-Be-Excluded-IDRA-Policy-Brief-2019.pdf</u>
- Kennedy B., & Medina J. (2017). Practitioner brief: Dual Language Education: Answers to questions from the field. Center for Applied Linguistics. <u>https://www.cal.org/resource-cen-</u> ter/briefs/dual-language-education-answers-to-questions-from-the-field
- Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in schools. *Educational Researcher*, *35*(7), 3-12. https://doi.org/10.3102/0013189X035007003
- Latham Sikes, C., & Villanueva, C. (2021). Creating a more bilingual Texas: A closer look at bilingual education in the Lone Star state. IDRA & Every Texan. <u>http://idra.news/BilingualTx</u>
- Marian, V., & Shook, A. (2012). The cognitive benefits of being bilingual. *Cerebrum: The Dana Forum* on Brain Science, 2012, 1-12.
- Solano Flores, G. (2016). Assessing English Language Learners: Theory and practice (1st ed.). Routledge.
- Texas Early Childhood English Learner Initiative, (2021). Texas Early Childhood English Learner Initiative Policy Roadmap, <u>https://bilingualtexas.org/roadmap</u>

- Texas Education Agency. (2018). *Texas commission on public school finance report.* <u>https://tea.texas.gov/sites/default/files/Public%20School%20Finance%20Commission%20Report%20v8%20-%20clean.pdf</u>
- Taboada, M. B. (2020, November 13). Austin-area students failing at higher rates amid coronavirus pandemic. *Austin American-Statesman*. <u>https://www.statesman.com/news/20201113/austin-area-students-failing-at-higher-rates-amid-coronavirus-pandemic</u>
- Ura, A., & Ahmed, N., (2018, June 21) Hispanic Texans on pace to become largest population group in state by 2022. *Texas Tribune*. <u>https://www.texastribune.org/2018/06/21/hispanic-texans-pace-become-biggest-population-group-state-2022/</u>
- Valdés, G., & Figueroa, R.A. (1994). Bilingualism and testing: A special case of bias. Ablex Publishing. Valencia, R. R. (1997). The evolution of deficit thinking: Educational thought and practice. Falmer Press.



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





This work is licensed under a Creative Commons Attribution 4.0 International License. Permissions beyond the scope of this license may be available at <u>www.review.education.texas.edu</u>

# The Testing Industrial Complex: Texas and Beyond

MARIA DEL CARMEN UNDA

The University of Texas at Austin

LIZETH LIZÁRRAGA-DUEÑAS

The University of Texas at Austin

**To cite this article:** Del Carmen Unda, M. and Lizárraga-Dueñas, L. (2021). The Testing Industrial Complex: Texas and Beyond. *Texas Education Review*, 9(2), 31-42. http://dx.doi.org/10.26153/tsw/13911

#### The Testing Industrial Complex: Texas and Beyond

MARIA DEL CARMEN UNDA The University of Texas at Austin

#### LIZETH LIZARRAGA-DUEÑAS The University of Texas at Austin

#### Introduction

The Testing Industrial Complex (TIC) is a system (and at the same time a cycle) in which high stakes standardized testing fuels neoliberal education reforms and vice versa (Roberts, 2015; Croft et al., 2015). These "reforms" and cycles have monetized for profit the public education system in which curriculum, students, and teachers have been packaged and sold for corporate profit. The Prison Industrial Complex (PIC) is a system in which inmates, which are disproportionately Black, Indigenous, people of color (BIPOC), are packaged then sold to private corporations for profit (Alexander, 2010; Davis & Shaylor, 2001). This policy brief examines two systems - the Testing Industrial Complex and the Prison Industrial Complex - and how they directly impact students in the state of Texas and the U.S. (see Figure 1). In detail below, we examine two alternatives particularly worthy of consideration: a) multiple measures accountability and b) evidence-based interventions. We close with policy recommendations for state-level policy makers and school leaders.

#### The Prison Industrial Complex

The Prison Industrial Complex (PIC) was created in the aftermath of slavery in the United States (Gilmore, 2000; Heiner, 2007; Stevenson, 2019; Wacqant, 2002). In the words of Dr. Angela Y. Davis, the U.S. went "from the prison of slavery to the slavery of prison" (Heiner, 2007, p. 221). The PIC has historical origins that date back to a post-civil war era that replaced slave labor with inmate labor. Within this model, slave plantations were replaced by contemporary prison farms (and thus cheap prison labor) that disproportionately consisted of African American inmates, and in current times continues to consist mainly of BIPOC (Roberts, 2015). Roberts (2015) states that the PIC is a system that resembles commerce in that it involves buying and selling cheap prisoner labor (p. 155). In fact, in 1979 the United States government passed the Justice System Improvement Act which allowed U.S. corporations to pay prisoners far less than minimum wage nor do they have to follow workers' rights (Roberts, 2015). Many large corporations rely on prison labor to perform duties that were once solely handled by the government for financial gain (Gordon, 1999). According to data by Market research firm IBISWorld, private correctional facilities are a \$4.8 billion dollar industry, with profits of \$629 million (White, 2015). GEO Inc, which holds roughly 37-28 percent of the industries market, reported that they also operate correctional facilities in a number of countries overseas however, more specifically, two-thirds of its profits comes from correction and detention facilities in the U.S. (White, 2015). An analysis written by Bryan Stevenson in The New York Times confirmed that the United States has the highest rate of incarceration of any nation on Earth (2019). Stevenson (2019) estimates that the United States represents "4 percent of the planet's population but 22 percent of its imprisoned." He goes on to state that in the early 1970s prisons held fewer than 300,000 people and that number has grown to more than 2.2 million in prison with 4.5 million on probation or parole.

Figure 1



# How High-Stakes Testing feeds the School-to-Prison pipeline in Texas

*Note*: How high-stakes testing feeds into the school-to-prison pipeline. This model is an aggregate from the works of Alexander (2010). Davis & Shaylor (2001), Roberts (2015) Croft, Roberts, & Stenhouse (2015)

#### The Testing Industrial Complex

The Testing Industrial Complex (TIC) is not only a system, but it also involves cyclical patterns where high stakes standardized testing sustains neoliberal education reforms<sup>1</sup>. This neoliberal logic monetizes the public education system where the many elements of schooling like curriculum, assessments, and even students and teachers are bargained for corporate profit (Roberts, 2015). The United States education system allows large corporations to profit off of students and teachers by sustaining an entire testing industry that forces teachers to teach a narrow curriculum where students learn to memorize and fill in circles and learn little to no critical skills. According to Croft, Roberts, and Stenhouse (2015), the TIC mirrors the PIC, such as both incorporate the: (a) use of surveillance and unwarranted policing to feed punitive reform measures used to solve what are in reality economic, social, and political problems, (b) confluence of bureaucratic, political, economic and racialized interests with the underlying purpose of diverting profits from public entities to private corporations; (c) increases in high stakes outcomes; and (d) a perception that the complex is practically impossible to dismantle (p. 73)

#### **High-Stakes Testing**

In the last twenty years, the United States federal government has dramatically escalated the demand for centralized accountability in the United States public education system. The concept of highstakes testing was birthed as a direct result of this demand (Altshuler and Schmautz, 2006; Kamenetz, 2015). From the passage of No Child Left Behind (NCLB) Act in 2001 passed by the Bush administration to Every Student Succeeds Act (ESSA) in 2015 passed by the Obama administration both required states to test students in reading and math once a year in grades 3 through 8, test once in high school, and additionally tests students once in science in grade school, middle school and high school (U.S. Department of Education, ESSA; U.S. Department of Education NCLB). Both NCLB and ESSA as federal K-12 general education policies were intended to reform the education system and improve student achievement however, they mostly demanded strict accountability via high stakes testing for results of student achievement without the necessary infrastructure nor resources to accomplish this goal. These tests are labeled high-stakes due to the fact that individual student scores and overall school scores are tied to individual promotion, graduation, or momentary allotments to schools or systems (Altshuler and Schmautz, 2006; Kamenetz, 2015; Ritt, 2016). Before the practice of high stakes, consequence-based testing becomes further embedded and normalized in our schools we need to consider the specific effects of such testing on students.

Studies indicate that there is no consistent evidence to suggest that high-stakes testing leads to increases in student learning and achievement (Cannell, 1988; Camilli, 2000; Haney, 2000; Jacob, 2001; Linn et. al., 1990; Marchant and Paulson, 2005; Shepard, 1990). In fact, a review of the literature outlines that high stakes testing has negative impacts on learning environments (Ritt, 2016; Rushton and Juola-Rushton, 2008) and student learning/achievement (Amrein et al., 2002; Amrein and Berliner, 2003; Nichols, et al., 2006; Nichols, et. al., 2012). Empirical evidence suggest that increased high

<sup>&</sup>lt;sup>1</sup> Neoliberalism refers to the market-oriented reform laws and policies that "eliminate price controls, deregulating capital markets, lowering trade barriers" and drastically decreasing the governments influence of the economy and public services (Boas & Gans-Morse, 2009). In other words, neoliberalism is an effort to privatize public services such as hospitals, education, transportation, social security. The application of neoliberal values to education reform can be quite problematic considering that by privatizing education it shifts the responsibility for high quality education from the state to the individual (Brathwaite, 2016). Neoliberalism in education thus ignores the systematic and structural inequalities that persist in public schools.

stakes test scores do not equate to increased learning (Cannell, 1989; Kortez, et al., 1996). Additionally, other studies have found that high stakes testing have colossal negative impacts for low income and students of color (Au, 2016; Horn, 2010; McNeil, 2000; Pierre, 2016, Zabala, 2007). Ample research has demonstrated that both Black (Lee, 1998; Madaus and Clarke, 2001; Roth et al., 2001) and Latin(o/a/x) (Altshuler and Schmautz, 2006; Valenzuela, 2005) students experience bias from standardized testing. The fact that Black and Latinx students are more likely to have negative impacts from standardized testing is particularly concerning since students who fail such exams are more likely to drop out of high school and have a statistically higher rate of ending up in prison (Au, 2016; Darling-Hammond, 2007, Rios, 2011). Thus, standardized testing can lead to traumatic consequences for Black and Brown students and their families and communities.

#### High-Stakes State Testing and Texas

In 1979, the state of Texas implemented a statewide testing program that changes periodically to comply with state/federal mandates and rulemaking from the state's primary oversight agency, the Texas Education Agency (TEA). Since its inception, the Texas statewide testing system has steadily grown in size, scope, and rigor. When it was first implemented in 1979 the Texas assessment program required that students take basic skills competencies in mathematics, reading, and writing for grades three, five, and nine (Texas Education Agency, 2008). Presently, the statewide testing program is titled the State of Texas Assessment of Academic Readiness (STARR) and was first implemented in spring 2012. As of today (due to COVID-19 these requirements will most likely change for the 2021-22 school year), it includes annual assessment for:

- reading and mathematics, grades 3-8
- writing, grades 4 and 7
- science, grades 5 and 8
- social science, grade 8
- end-of-course assessment for English I, English II, Algebra I, Biology, and U.S. History.

Therefore, the state and TEA require that students take STARR exams a total of 15 times between third and ninth grade (Texas Education Agency, 2007-2020). Additionally, TEA requires students in fifth and eighth grade to pass the STAAR exam to be able to advance into the next grade level<sup>2</sup>.

# The Intersectionality of High-Stakes Testing, Texas, and Capitalism

As previously mentioned, standardized testing companies are siphoning millions of dollars from students, teachers, and communities across the country, with little evidence that these systems are improving student performance, closing achievement gaps, or motivating teacher improvement. In the year 2000, PBS reported that, "while test sales in 1955 were \$7 million (in 1998 dollars), that figure was \$263 million in 1997, an increase of more than 3,000 percent." Thirteen years later, the Londonbased Pearson Company secured a five-year contract with TEA for \$468 million dollars to provide state assessments (Smith, 2013). In 2015, TEA announced that it would be switching over to the

<sup>&</sup>lt;sup>2</sup> It is important to note that because of the COVID-19 pandemic, Texas Governor Greg Abbot announced that grade promotion's dependence on passing the STAAR exam is waived for the 2020-2021 school year (Office of the Texas Governor-Greg Abbott, 2020). However, for the 2020-2021 school year Texas education officials decided to administer the STARR exam in person during COVID-19 (Agnew & Bohra, 2021).

Education Testing Services (ETS) to develop and administer the state-required exams. TEA paid ETS a total of \$468 million dollars for the five-year contract (Smith, 2015; Texas Education Agency ETS Contract, 2016). In line with this, *The Washington Post's* Valerie Strauss (2015) revealed that collectively, Pearson, ETS, Houghton Mifflin Harcourt, and McGraw-Hill have spent more than \$20 million dollars lobbying in states and on Capitol Hill from 2009 to 2014. Strauss also reported that ETS's outgoing president Kurt Landgraf received more than \$1.3 million dollars in total compensation in 2013. There is no evidence that adding more standardized tests increases student learning and achievement (Cannell, 1988; Camilli, 2000; Haney, 2000; Jacob, 2001; Linn et. al., 1990; Marchant and Paulson, 2005, Shepard, 1990). On the other hand, we have plenty of evidence that increasing testing is very profitable for those who sell the tests and supply the infrastructure (Alexander, 2010; Davis & Shaylor, 2001; Roberts, 2015). Testing fever will end only when the greed of the standardized-testing-industrial complex is satisfied--in other words, never. In the next section, we will offer alternatives to standardized testing supported by a large body of education research.

# Beyond High-Stakes Standardized Testing

# **Multiple Measures Accountability**

A number of educational scholars have argued that states need to evaluate all students beyond test scores and should implement the use of multiple measures for accountability (Cook-Harvey et al., 2016; Darling-Hammond et al., 2016; Egalite et al., 2017; Mathis, 2015; Mathis & Trujillo, 2016; Punuel et al., 2016). One of the main criticisms regarding a test-based model is that standardized testing does not measure all the important aspects of a successful school and student learning (Gipps, 1999; Hartman et al., 2017; Mathis, 2015; Mathis & Trujillo, 2016). This claim, combined with the backlash and testing fatigue from students and parents against what they consider to be excessive testing, has led to the organic development of demands for "multiple measures" state accountability systems (Mathis, 2015; Segool et al., 2013).

Mathis (2015) defined multiple measures as "a more comprehensive set of measures [that] will more validly capture the broader set of cognitive and affective learning goals for schooling" (p. 2). Advocates of multiple measures speak of a "dashboard" composed of data on elements such as truancy, graduation rates, and disciplinary referrals (Mathis & Trujillo, 2016), while other scholars have called for aggregation of data on chronic absenteeism, student safety, risky behaviors, and belonging (Penuel et al., 2016). In a report in collaboration with the Learning Policy Institute, Cook-Harvey et al. (2016) thoroughly outlined potential indicators for a multiple measures system, including but not limited to the following (see Table 1):

# Table 1

Potential Indicators for a Multiple Measures System

Graduation/school progress				
<ul> <li>Fourth, fifth, and sixth grade cohort graduation rates</li> <li>Proportion of 8th graders that progress into the 9th grade</li> <li>Drop out rates</li> </ul>				
Career and College Readiness				
<ul> <li>Proportion of students completing college preparatory coursework and/or improved technical education (CTE) sequence or both.</li> </ul>				
•Proportion of students meeting standard and graduation portfolios, industry- approved certificates, licenses, or badges recognized by post-secondary institutions.				
Access to Resources				
<ul> <li>Ratios of students to counselors and specialists</li> </ul>				
•Teacher qualifications				
•Safe and adequate facilities				
School Climate				
•Evidence from students and staff surveys about school offerings, instruction, supports, trust, and belonging.				
Student Participation				
<ul> <li>Average daily attendence/chronic absenteeism rates</li> </ul>				
•Suspension and expulsion rates				

Other educational scholars argue that the state of Texas should implement authentic assessments designed to meet the needs of all students, which include project and portfolio-based assessments, and that schools and school districts should create Individual Graduation Committees (IGCs) which can also serve as a way to increase high school graduation rates (Hartman et al, 2017).

# Multiple Measures Accountability and Federal ESSA

Multiple measures accountability, authentic assessments, and project-based and portfolio-based assessments are in full compliance with the Federal Every Student Succeeds Act (ESSA) of 2015. ESSA outlines that states must "involve multiple up-to-date measures of student academic achievement, including measures that assess higher-order thinking skills and understanding, which may include measures of student academic growth and may particularly be delivered in the form of portfolios projects or extended performance tasks" (§ 1177-25). In other words, ESSA requires multiple measures for accountability, giving states the option of evaluating students using more than singlemeasure test score gains (Cook-Harvey et al., 2016, p. 1; Egalite et al., 2017, p. 767). ESSA (2015) explicitly allows states and school districts to go beyond standardized testing and allows the use of portfolios, projects, or extended-performance tasks as well as adaptive assessments as part of state systems (§ 1177-25). However, the state of Texas does not currently employ this approach. Instead, the TEA continues to test students using single measure, standardized, high stakes testing programs despite the clear recommendations outlined by scholars and advocates rooted in significant concerns about the inefficiency, ineffectiveness, and inequity of the current testing system. Education scholars have advised school leaders, via research scholarship, to diversity accountability indicators for students in order to create an equitable education system (Cook-Harvey et al., 2016; Darling-Hammond et al., 2016; Egalite et al., 2017; Mathis, 2015; Mathis & Trujillo, 2016; Punuel et al., 2016).

#### **Evidence Based Interventions**

Per federal regulations outlined by ESSA, policy makers and school leaders must use research-based practices to improve the education system for students (Callahan & Hopkins, 2017, p. 762; Dynnarski, 2015, p. 1; Egalite et al., 2017). If implemented well with the sufficient allocation of resources this can improve student performance, reduce educational disparities, and increase graduation rates for all students. ESSA states that local education agencies must utilize "evidence-based interventions" in order to receive federal funding (Callahan & Hopkins, 2017, p. 762; Cook-Harvey, Darling-Hammond et al., 2016). As such, ESSA (2015) defines "evidence-based interventions" as programs "that demonstrate a rationale based on high quality research findings or positive evaluation that [shows they are] likely to improve student outcomes... and... includes ongoing efforts to examine... effects" (Every Student Succeeds Act, 2015). Penuel et al., (2016) have urged policymakers, school administrators, and teachers to identify multiple evidence-based studies and resources to make sure that new accountability policies measure what they are intended to measure. Equally important, they strongly recommended that school leaders and administrators gather the evidence and studies ahead of time to correctly implement such practices.

Callahan and Hopkins (2017) argue that ESSA's definition of "evidence-based interventions" aligns with the requirements that emerged from the *Castaneda vs. Pickard* (1981) decision that was tried in the United States District Court for Southern District in Texas. Although *Castaneda vs. Pickard* (1981) focused primarily on English Learner students, it established a three-part assessment for determining if education programs are 1) based on sound educational research and theory, 2) well-implemented with sufficient resources and personnel, and 3) evaluated regularly to ensure progress towards linguistic and academic goals. These criteria define "evidence-based interventions" that meet the requirements established henceforth by the Equal Education Opportunities Act of 1974. Using *Castaneda vs. Pickard* (1981) evidence-based framework, below we outline policy recommendations derived from the review of educational research presented above.

# Policy Recommendations for Texas State-level Policymakers

- End assessment contracts with for-profit corporations that produce and administer standardized tests
- Involve multiple stakeholders (students, teachers, the community, families, parents, policymakers, and educational scholars) in the design and implementation of a state evaluation program.
- Texas policymakers, TEA, and school districts apply a stringent criterion when adopting interventions. Employ high quality peer-reviewed research findings moving forward.
- Per the Federal Every Student Succeeds Act (2015), implement multiple measures accountability that goes beyond single-measure, high stakes standardized testing.
- Texas policymakers, TEA, and school leaders/administrators should establish, develop, and train school teams that collect and analyze both quantitative and qualitative data. Prioritize schools with the most need and least resources.

**Lizeth Lizárraga-Dueñas** is a certified special education teacher in the Austin, Texas area. She likes to volunteer her time in community programs like Academia Cuauhtli and First Lego League robotics after-school programs. Her research interests focus on educational policy making and the complex relationship between academic literacy in school and informal educational settings.

**María Del Carmen Unda** is a doctoral student at the Department Educational Leadership & Policy at the University of Texas, Austin, and a fellow with the Texas Center for Education Policy. Her research focuses on educational policymaking, policy implementation, and educational assess at the K-12 level particularly related to students of color. She currently works at Academia Cuauhtli, a culturally revitalization community-based school in Austin, Texas and as a community organizer with Nuestro Grupo

#### References

- Agnew, D., & Bohra, N. (2021, April 6). First day of STAAR testing canceled for thousands of students experiencing technical issues across Texas. *Texas Tribune*. <u>https://www.texastrib-</u> <u>une.org/2021/04/06/texas-staar-testing/</u>
- Alexander, M. (2010). The new Jim Crow: Mass incarceration in the age of colorblindness. The New Press.
- Altshuler, S., & Schmautz, T. (2006). No Hispanic student left behind: The consequences of "high stakes" testing. *Children & Schools, 28*(1), 5–14. <u>https://doi.org/10.1093/cs/28.1.5</u>
- Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing & student learning. Education Policy Analysis Archives, 10(18). <u>https://doi.org/10.14507/epaa.v10n18.2002</u>
- Amrein, A. L., & Berliner, D. C. (2003). The effects of high-stakes testing on student motivation and learning. *Educational leadership*, 60(5), 32-38.
- Au, W. (2015). Meritocracy 2.0: High-Stakes, standardized testing as a racial project of neoliberal multiculturalism. *Education Policy*, 30(1), 39-61. https://doi.org/10.1177%2F0895904815614916
- 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
- Boas, T. C., & Gans-Morse, J. (2009). Neoliberalism: From new liberal philosophy to anti-liberal slogan. *Studies in Comparative International Development*, 44(2), 137-161.
- Brathwaite, J. (2017). Neoliberal education reform and the perpetuation of inequality. *Critical Sociology*, *43*(3), 429-448.
- Callahan, R. M., & Hopkins, M. (2017). Policy brief: Using ESSA to improve secondary English Language Learners opportunities through course taking. *Journal of School Leadership, 27*, 755-766. <u>https://doi.org/10.1177%2F105268461702700507</u>
- Camilli, G. (2000). Texas gains on NAEP: Points of light? *Education Policy Analysis Archives, 8*(42). https://doi.org/10.14507/epaa.v8n42.2000
- Cannell, J. J. (1988). Nationally normed elementary achievement testing in America's public schools: How all 50 states are above the national average. *Educational Measurement: Issues and Practice*, 7(2), 5-9.
- Cannell, J. (1989). The "Lake Wobegon" report: How public educators cheat on standardized achievement tests. Friends for Education.
- Cook-Harvey, C.M., Darling-Hamming, L., Lam, L., Mercer, C., & Roc, M. (2016). *Equity and ESSA:* Leveraging educational opportunity through the Every Student Succeeds Act. Learning Policy Institute. <u>https://learningpolicyinstitute.org/sites/default/files/product-files/Equity\_ESSA\_RE-PORT.pdf</u>
- Croft, S. J., Roberts, M. A., & Stenhouse, V. L. (2015). The perfect storm of education reform: High-Stakes testing and teacher evaluation. *Social Justice*, 42(1), 70-92. https://www.jstor.org/stable/24871313#metadata\_info\_tab\_contents
- Darling-Hammond, L. (2007). Race, inequality and educational accountability: The irony of "No Child Left Behind." Race, Ethnicity, and Education, 10(3), 245–260.
- Darling-Hammond, L., Bae, S., Cook-Harvey, C. M., Lam, L., Mercer, C., Podolsky, A. & Stosich, L. (2016). Pathways to New Accountability Through the Every Student Succeeds Act. Learning Policy Institute. <u>https://learningpolicyinstitute.org/product/pathways-new-accountability-throughevery-student-succeeds-act</u>
- Davis, A. Y., & Shaylor, C. (2001). Race, gender, and the prison industrial complex: California and beyond. *Meridians*, 2(1), 1–25. <u>https://doi.org/10.1215/15366936-2.1.1</u>

- Dynarski, M. (2015). Using research to improve education under the Every Student Succeeds Act. *Economic Studies*, 1(8), 1-5. <u>https://www.brookings.edu/wp content/up-loads/2016/07/Download-the-paper-3.pdf</u>
- Egalite, A. J., Fusarelli, L. D., & Fusarelli, B. C. (2017). Will decentralization affect educational inequality? The Every Student Succeeds Act. *Educational Administration Quarterly 2017, 53*(5), 757-781.
- Every Student Succeeds Act, 20 U.S.C § 6301 (2015). <u>https://www.ed.gov/essa?src=ft</u>
- Fulcher, P. A. (2012). Hustle and flow: Prison privatization fueling the prison industrial complex. *Washburn Law Journal, 51*(3), 583-618.
- Gibbs, C. (1999). Socio-cultural aspects of assessment. Review of the Research in Education. 24(1), 355-392. <u>https://doi.org/10.3102%2F0091732X024001355</u>
- Gilmore, K. (2000). Slavery and prison-understanding the connections. *Social Justice*, 27(3), 195-205. https://www.jstor.org/stable/29767242
- Gordon, A. (1999). Globalism and the prison industrial complex: An interview with Angela Davis. Race & Class, 40(2-3), 145–157. <u>https://doi.org/10.1177/030639689904000210</u>
- Haney, W. (2000). The myth of the Texas miracle in education. *Education Policy Analysis Archives*, 8(41). <u>https://doi.org/10.14507/epaa.v8n41.2000</u>
- Hartman, C., Pulte, G., Gutierrez, K., & Davies, W. (2017). High-stakes testing in Texas high schools: The case for individual graduation committees and authentic assessment [Policy Brief]. Texas Center for Education Policy. <u>https://drive.google.com/file/d/0B-CexrNlohBdMidINiNubEdDLUk/view</u>
- Heiner, B.T. (2007). "From the prison of slavery to the slavery of prison" Angela Y. Davis's Abolition Democracy. Radical Philosophy Today, 5, 219-227. <u>https://www.pdcnet.org/radphiltoday/content/radphiltoday\_2007\_0005\_0000\_0219\_0227</u>
- Horn, C. (2003). High-stakes testing and students: Stopping or perpetuating a cycle of failure? *Theory* Into Practice, 42(1), 30-41. <u>https://doi.org/10.1207/s15430421tip4201\_5</u>
- Jacob, B. (2001). Getting tough? The impact of high school graduation exams. *Educational Evaluation* and Policy Analysis, 23(2), 99-121.
- Kamenetz, A. (2015). The Test: why our schools are obsessed with standardized testing but you don't have to be. PublicAffairs.
- Koretz, D., Barron, S., Mitchell, K., & Stecher, B. (1996). Perceived effects of the Kentucky Instructional Results Information System (KIRIS). RAND. <u>https://www.rand.org/pubs/monograph\_reports/MR792.html</u>
- Lee, C. (1998). Culturally Responsive Pedagogy and Performance-Based Assessment. *The Journal of Negro Education, 67*(3), 268–279. <u>https://doi.org/10.2307/2668195</u>
- Linn, R. L., Graue, M. E., & Sanders, N. M. (1990). Comparing state and district test results to national norms: The validity of claims that "everyone is above average." *Educational Measurement: Issues and Practice*, 9(3), 5-14.
- Madaus, G., & Clarke, M. (2001). The adverse impact of high stakes testing on minority students: Evidence from 100 years of test data. In G. Orfield & M. Kornhaber (Eds.), Raising standards or raising barriers? Inequality and high stakes testing in public education, 51–85. http://files.eric.ed.gov/fulltext/ED450183.pdf
- Mathis, W. M. (2015). Research-based options for educational policymaking: School accountability, multiple measures and inspectorates in a post-NCLB world. National Education Policy Center. https://nepc.colorado.edu/sites/default/files/publications/Mathis%20RBOPM-1\_0.pdf
- Mathis, W. J. & Trujillo, T. M. (2016). Lessons from NCLB for the Every Students Succeeds Act. National Education Policy Center. <u>https://nepc.colorado.edu/publication/lessons-from-NCLB</u>

- Marchant, G. J., & Paulson, S. E. (2005). The relationship of high school graduation exams to graduation rates and SAT scores. *Education Policy Analysis Archives*, 13(6). <u>http://epaa.asu.edu/epaa/v13n6/</u>
- McNeil, L. M. (2000). Contradictions of school reform: Education cost of standardized testing. Routledge.
- 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). <u>https://doi.org/10.14507/epaa.v14n1.2006</u>
- Nichols, S., Glass, G., & Berliner, D. (2012). High-stakes testing and student achievement: Updated analyses with NAEP data. *Education Policy Analysis Archives, 20*(20). https://doi.org/10.14507/epaa.v20n20.2012
- No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 101, Stat. 1425 (2002). https://www2.ed.gov/nclb/landing.jhtml
- Office of the Texas Governor-Greg Abbott. (2020, July 27). Governor Abbott Waives Grade Promotion Requirements for 2020-2021 STAAR testing. <u>https://gov.texas.gov/news/post/governor-abbott-waives-grade-promotion-requirementsfor-2020-2021-staar-testing</u>
- Penuel, W., Meyer, E., & Valladares, M. R. (2016). Making the most of the Every Student Succeeds Act (ESSA)- Helping states focus on school equity, quality, and climate. National Education Policy Center. <u>https://files.eric.ed.gov/fulltext/ED578780.pdf</u>
- Pierre, T. (2016). The impact of high stakes testing on low income students and students of color. [Masters Dissertation, State University of New York, Empire State College]. ProQuest Dissertations & Theses Global.
- PBS. (2000). The testing industry big four. <u>https://www.pbs.org/wgbh/pages/front-line/shows/schools/testing/companies.html</u>
- Ravitch, D. (2010). The death and life of the great American school system: How testing and choice are undermining education. Basic Books.
- Rios, V. (2011). Published: Policing the lives of Black and Latino boys. New York University Press.
- Ritt, M. (2016). The Impact of High-stakes Testing on the Learning Environment. [Masters Dissertation, St. Catherine University/University of St. Thomas St. Paul, Minnesota] Sophia the St. Catherine University Repository. <u>https://sophia.stkate.edu/msw\_papers/658</u>
- Roberts, M. A. (2015). The Testing Industrial Complex: Incarcerating education since 2001. In M. Abrendroth, & B.J. Porfilio (Eds.) Understanding neoliberal rule in k-12 schools. 153-180. Information Age Publishing.
- Roth, P., Bevier, C., Bobko, P., Switzer, F., III, & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A metanalysis. *Personnel Psychology*, 54(2), 297-330. <u>https://doi.org/10.1111/j.1744-6570.2001.tb00094.x</u>
- Rushton, S., & Juola-Rushton, A. (2008). Classroom learning environment, brain research and the no child left behind initiative: 6 years later. *Early Childhood Education Journal, 36*(1), 87-92.
- Segool, N., Carlson, J. S., Goforth, A. N., Von Der Embse, N., & Barterian, J. A. (2013). Heightened test anxiety among young children: Elementary school students' anxious responses to highstakes testing. *Wiley Periodicals: Psychology in the Schools*, 50(5), 489-499.
- Shepard, L. A. (1990). Inflated test scores gains: Is the problem old norms or teaching the test? Educational Measurement: Issues and Practice, 9(3), 15-22.
- Smith, M. (2013, April 21). Seeking to pare state exams, lawmakers take aim at testing firms. *Texas Tribune*. <u>https://www.texastribune.org/2013/04/21/taking-aim-testing-firm-quest-pare-state-exams/</u>

- Smith, M. (2015, May 18). Pearson loses bulk of Texas testing contract. *Texas Tribune*. <u>https://www.texastribune.org/2015/05/18/pearson-loses-bulk-texas-student-testing-con-tract/</u>
- Stevenson, B. (2019, August 14). Slavery gave America a fear of black people and a taste for violence punishment. Both still define our criminal-justice system. *The New York Times*.
- https://www.nytimes.com/interactive/2019/08/14/magazine/prison-industrial-complex-slaveryracism.html
- Strauss, V. (2015, March 30). Report: Big education firms spend millions lobbying for pro-testing policies. Washington Post. <u>https://www.washingtonpost.com/news/answersheet/wp/2015/03/30/report-big-education-firms-spend-millions-lobbying-for-pro-testingpolicies/</u>
- Texas Education Agency. (2008). Chapter 1: Historical overview of assessment in Texas. https://tea.texas.gov/sites/default/files/digest09-chap01.pdf
- Texas Education Agency (2007-2020). STAAR resources. <u>https://tea.texas.gov/student-assess-ment/testing/staar/staar-resources</u>
- Texas Education Agency ETS Contract (2016, October 14). Amendment to standard contract between Texas Education Agency and Educational Testing Service (ETS). <u>https://tea.texas.gov/sites/de-fault/files/3317\_ETS\_Amend\_1.pdf</u>
- Valenzuela, A. (2005). Leaving children behind: how Texas-style accountability fails Latino youth. State University of New York Press.
- Wacquant, L. (2002). Slavery to mass incarceration. New Left Review, 13, 41-60.
- White, M. (2015). Locked-in profits: the U.S. prison industry by the numbers. *NBC News*. <u>https://www.nbcnews.com/business/business-news/locked-in-profits-u-s-prison-industry-numbers-n455976</u>
- Zabala, D. (2007). State high school exit exams: Gaps persist in high school exit exam pass rates—Policy brief 3. Center on Education Policy.



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





This work is licensed under a Creative Commons Attribution 4.0 International License. Permissions beyond the scope of this license may be available at <u>www.review.education.texas.edu</u>

# 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.<sup>1</sup> 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 profitdriven 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

<sup>&</sup>lt;sup>1</sup> 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<sup>2</sup> 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).<sup>3</sup> 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

 $<sup>^{2}</sup>$  By the time of publication, Harden had been traded to the Brooklyn Nets, though the analysis of his contract remains relevant.

<sup>&</sup>lt;sup>3</sup> 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).<sup>4</sup> 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.<sup>5</sup> 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

<sup>&</sup>lt;sup>4</sup> 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.

<sup>&</sup>lt;sup>5</sup> 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.<sup>6</sup> 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.

<sup>&</sup>lt;sup>6</sup> 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,<sup>7</sup> 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

<sup>&</sup>lt;sup>7</sup> 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.

<sup>&</sup>lt;sup>8</sup> 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.<sup>9</sup>

# 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

<sup>&</sup>lt;sup>9</sup> 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 topdown 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 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). Class-rooms with authentic assessment are student-centered. Educators emphasize cooperative work; fo-cus 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.
  - o 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 read-iness" that resulted in the school being designated an "exemplar school" by the Partnership for 21<sup>st</sup> 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 9<sup>th</sup> grade through graduation. We recommend that schools introduce authentic assessments with a 9<sup>th</sup> 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 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 multiplechoice 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. <u>https://www.urbancollaborative.org/sites/de-fault/files/resource\_files/testpunishpushout.pdf</u>
- Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. Education Policy Analysis Archives, 10(0), 18. <u>https://doi.org/10.14507/epaa.v10n18.2002</u>
- 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). <u>https://doi.org/10.14507/epaa.v18n14.2010</u>
- Archbald, D. A. (1991). Authentic assessment: Principles, practices, and issues. School Psychology Quarterly, 6(4), 279-293. <u>https://doi.org/10.1037/h0088821</u>
- 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. <u>https://doi.org/10.3102/0013189X07306523</u>
- Au, W. (2016). Meritocracy 2.0: High-stakes, standardized testing as a racial project of neoliberal multiculturalism. *Educational Policy*, 30(1), 39-62. <u>https://doi.org/10.1177/0895904815614916</u>
- 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*. <u>https://www.dallasnews.com/news/2016/04/28/staar-test-question-with-no-right-answer-among-dozens-</u>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). <u>https://rethinkingschools.org/articles/bamboozled-by-the-texas-miracle/</u>
- 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. <u>https://doi.org/10.7764/PEL.51.1.2014.5</u>
- Blakeslee, N. (2013, May). Crash test. *Texas Monthly*. <u>https://www.texasmonthly.com/news-poli-tics/crash-test/</u>
- 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*. <u>https://abolition.university/invitation/</u>
- Bradford, D. (2021). A rolling national teacher strike is why schools are closed. *Education Next*, 21(3). <u>https://www.educationnext.org/rolling-national-teacher-strike-is-why-schools-are-closed/</u>
- Braginsky, N. (2020). Not an "achievement gap", a racial capitalist chasm. The Law and Political Economy Project. <u>https://lpeproject.org/blog/not-an-achievement-gap-a-racial-capitalist-chasm/</u>

- Carpenter, J. (2019, March 4). Lawmakers take up STAAR over wording of test questions. *Houston Chronicle*. <u>https://www.houstonchronicle.com/newshouston-texas/education/article/Law-makers-take-up-STARR-over-wording-of-test-13662711.php</u>
- Carpenter, J. (2020, November 18). Bipartisan group of Texas lawmakers call on state to move toward canceling next year's STAAR tests. *Houston Chronicle*. <u>https://www.houstonchronicle.com/news/houston-texas/education/article/Bipartisan-group-of-Texas-lawmakers-callon-state-15736944.php</u>
- Chang, J. (2016, September 3). Texas districts document dozens of problems with 2016 STAAR testing. *Austin American-Statesman*. <u>https://www.statesman.com/NEWS/20160903/Texas-districts-document-dozens-of-problems-with-2016-STAAR-testing</u>
- 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). <u>https://jour-nal.jctonline.org/index.php/jct/article/view/835</u>
- 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*. <u>https://education.msu.edu/green-and-write/2014/texas-textbooks-and-the-politics-of-history-standards/</u>
- 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-optingout-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. <u>https://edpolicy.stanford.edu/sites/default/files/beyond-basic-skills-role-performance-assessment-achieving-21st-century-standards-learning-executive-summary\_0.pdf</u>
- Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education*, 16(5), 523-545. <u>https://doi.org/10.1016/S0742-051X(00)00015-9</u>
- 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. <u>https://doi.org/10.1089/zeb.2008.0576</u>
- 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. <u>https://eric.ed.gov/?id=ED606677</u>
- Fontana, J. (1995). Portfolio assessment: Its beginnings in Vermont and Kentucky. NASSP Bulletin, 79(573), 25-30. <u>https://doi.org/10.1177/019263659507957305</u>
- 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. <u>https://doi.org/10.1177/0895904817736631</u>
- Gisi, K. (2020, January 22). More Texas schools implementing project-based learning. *Spectrum News* 1. <u>https://spectrumlocalnews.com/tx/san-antonio/news/2020/01/23/more-texas-schools-implementing-project-based-learning</u>
- 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. <u>https://www.fairtest.org/sites/default/files/TestingReformVictoriesReport.pdf</u>
- 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. <u>https://doi.org/10.14507/epaa.v8n41.2000</u>
- 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 Anal*ysis, 30, 75-110. <u>https://doi.org/10.3102/0162373708317689</u>
- 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. <u>https://doi.org/10.3102/0034654310383147</u>
- 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-hiscity/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. <u>https://doi.org/10.1300/J370v23n02\_05</u>
- Kamenetz, A. (2014, June 2). In Kentucky, moving beyond dependence on tests. *WUTC, The Univer*sity of Tennessee at Chattanooga Public Radio NPR ED. <u>http://wutc.org/post/kentucky-students-</u> <u>succeed-without-tests</u>
- 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*. <u>https://www.aaihs.org/why-the-academic-achievement-gap-is-a-racist-idea/</u>
- 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. <u>https://citeseerx.ist.psu.edu/viewdoc/down-load?doi=10.1.1.567.7619&rep=rep1&type=pdf</u>
- 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. <u>https://ospri.re-search.gwu.edu/sites/g/files/zaxdzs2456/f/downloads/OSPrI Report 2013-01.pdf</u>

- Mays, M. (2021, February 3). "I know this is a drastic step but ...": San Francisco sues its schools over closures. Politico, n.p. <u>https://www.politico.com/states/califor-nia/story/2021/02/03/san-francisco-sues-school-district-over-closed-campuses-with-no-end-in-sight-1361744</u>
- 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. <u>https://eric.ed.gov/?id=EJ800872</u>
- Menendez, J. (2019, April 13). The STAAR test is hurting our kids and we need to act now. *Texas Monthly*. <u>https://www.texasmonthly.com/opinion/state-senator-jose-menendez-staar-test-hurting-kids-need-act-now/</u>
- Michels, Patrick (2014, March 27). Updated: Texas' testing opt-out movement grows to a 'trickle'. *The Texas Observer*. <u>https://www.texasobserver.org/texas-testing-opt-movement-grows-</u> <u>trickle/</u>
- National Education Association (2020, June 25). History of standardized testing. <u>https://www.nea.org/professional-excellence/student-engagement/tools-tips/history-stand-ardized-testing-united-states</u>
- National Association of Secondary School Principals (2006). Tracking and ability grouping in middle level and high schools. <u>https://www.nassp.org/tracking-and-ability-grouping-in-middle-level-and-high-schools/</u>
- 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. <u>https://drive.google.com/file/d/1VEb5XiDStR7nzFZxLc6WMMvwOMAnCnSq/view</u>
- 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. <u>https://doi.org/10.1207/s15326985ep2202\_3</u>
- 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. <u>https://www.pub-licschoolreview.com/success-academy-charter-school-harlem-2-profile</u>
- Putwain, D., & Remedios, R. (2014). The scare tactic: Do fear appeals predict motivation and exam scores? *School Psychology Quarterly*, 29(4), 503-516. <u>https://doi.org/10.1037/spq0000048</u>

- 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. <u>https://www.nea.org/advocating-for-change/new-from-nea/racist-beginnings-stand-ardized-testing</u>
- Rothstein, R. (2000, September 13). How tests can drop the ball. *The New York Times*. <u>https://www.nytimes.com/2000/09/13/us/lessons-how-tests-can-drop-the-ball.html</u>
- 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. <u>https://www.lubbockonline.com/article/20121210/NEWS/312109776</u>
- Shalaby, C. (2020, 2 June). Repurposing our pedagogies [Video]. Education for Liberation Network Facebook. <u>https://www.facebook.com/watch/live/?v=260890484987341&ref=watch\_permalink</u>
- Stanford School Redesign Network. (2008). What is performance-based assessment?<u>https://edpolicy.stan-ford.edu/sites/default/files/events/materials/2011-06-linked-learning-performance-based-assessment.pdf</u>
- Steele, C. M. (1999, August 1). Thin ice: Stereotype threat and black college students. *The Atlantic.* <u>https://www.theatlantic.com/magazine/archive/1999/08/thin-ice-stereotype-threat-and-black-college-students/304663/</u>
- 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*. <u>https://www.texasmonthly.com/news/testing-the-test-new-legislation-filed-to-audit-staar/</u>
- Swartz, M. (2019b). Are Texas kids failing? Or are the tests rigged? *Texas Monthly*. <u>https://www.tex-asmonthly.com/news-politics/texas-kids-failing-staar-tests-rigged/</u>
- 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.* <u>http://www.nytimes.com/2015/04/07/nyregion/atsuccess-academy-</u> <u>charter-schools-polarizing-methods-and-superior-results.html</u>
- 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 <u>https://www.facebook.com/optouttexas/</u>
- 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*. <u>https://tsta.org/wp-content/uploads/2020/03/PR0316cancelSTAAR.pdf</u>
- 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. <u>https://doi.org/10.1080/19404476.2016.1252304</u>
- Troy, G. (2016, June 26). The racist origins of the SAT. *The Daily Beast*. <u>http://giltroy.com/the-daily-beast/the-racist-origins-of-the-sat/</u>
- 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*. <u>https://hechingerreport.org/questioning-their-fairness-a-record-number-of-colleges-stop-requiring-the-sat-and-act/</u>
- 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. <u>http://neatoday.org/2014/11/02/nea-survey-nearly-half-of-teachersconsiderleaving-profession-due-to-standardized-testing-2/</u>
- 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

# SAMPLE ASSESSMENT

You are the assistant to Pat Williams, the president of DynaTech, a company that makes precision electronic instruments and navigational equipment. Sally Evans, a member of DynaTech's sales force, recommended that DynaTech buy a small private plane (a SwiftAir 235) that she and other members of the sales force could use to visit customers. Pat was about to approve the purchase when there was an accident involving a SwiftAir 235. You are provided with the following documentation:

 Newspaper articles about the accident
 Federal Accident Report on in-flight breakups in single engine planes

3: Pat's e-mail to you and Sally's e-mail to Pat

4: Charts on SwiftAir's performance characteristics 5: Pilot article comparing SwiftAir 235 to similar planes

6: Pictures and description

(adapted from Darling-Hammond & Adamson, 2010)

You will prepare a memo that addresses several questions, including what data support or refute the claim that the type of wing on the SwiftAir 235 leads to more in-flight breakups, what other factors might have contributed to the accident and should be taken into account, and your overall recommendation about whether or not DynaTech should purchase the plane. You will also develop a written or verbal presentation of your memo.

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.