Texas Education Review

**Avoiding the Gaze of the Test:
High Stakes Literacy Policy Implementation**

Thea Williamson, MEd
The University of Texas at Austin

Avoiding the Gaze of the Test: High-Stakes Literacy Policy Implementation

Thea Williamson, MEd
The University of Texas at Austin

Abstract

This qualitative embedded case study documents the policy implementation of literacy assessment in a Texas urban high school, using Foucault's theory of the panopticon to understand how teaching and learning were shaped by the state high-stakes exit exam. In addition to the strong influence of the State of Texas Assessments of Academic Readiness (STAAR) test, data also indicated that despite an atmosphere of surveillance, teachers also worked strategically to make room for other types of literacy instruction. This was most visible in English I and II teachers' commitment to process-based writing instruction focused on real-world text genres.

Keywords: policy implementation, high-stakes testing, English instruction, urban education

Research has documented the intractable nature of the grammar of schooling (Tyack & Tobin, 1994) that infuses educational institutions and structures the nature of teaching and learning; educational theorists continue to investigate the reasons why, despite nearly a century of secondary school reform, curriculum and instruction remains very traditional (Cuban, 2013). As the young people in U.S. public schools are increasingly multilingual and identify as students of color (Kena et al., 2016), the way students are taught about reading and writing often fails to acknowledge a plurality of voices, languages, and ways of creating and understanding texts. Instead, standardized conceptualizations of literacy dominate schools (Cochran-Smith, 1991, 2001). Given that policy research has indicated local practices in schools can remain stagnant even in the face of large-scale policy changes (McLaughlin, 1998), the lack of movement towards equity in educational achievement and overall improvements in the nation's literacy education is not shocking.

However, in addition to understanding the ways in which practices in schools are resistant to change, it is equally as important to attend to the effects policy does have in schools. The state of Texas claims its current "accountability system uses safeguards to minimize unintended consequences" (Texas Education Agency, 2016), yet research has documented alarming effects on the nature of language teaching and learning (Au, 2007, 2011). It does not appear that state or federal agencies are concerned with the qualitative changes in education as a result of high-stakes standardized testing: the lived effects of an accountability regime.

This study then provides a detailed, empirical examination of those unintended consequences—how they influence the nature of teaching and learning English in a Texas public high school, as well as the structures in schooling that serve to maintain the status quo in education. In this investigation I pursue the following research questions:

1) How does an urban school implement high-stakes state and local policies regarding literacy assessment?

2) Where and to what extent are teachers able to create space for individualized instruction and non-tested curricular content, particularly in the context of an urban school under institutional surveillance?

Study Context: Why Texas? Why Urban Schools?

An urban school in Texas provides an important context for understanding literacy assessment, both its consequences and how teachers make sense of accountability demands. As I will address below, Texas has a long history of standardized state-wide accountability measures, thus schools in Texas offer windows into larger, longitudinal trends regarding the nature of high-stakes testing. Understanding the consequences of accountability measures is critical considering the general trend towards standardized measures of learning seen in federal policies like No Child Left Behind (NCLB), the Common Core State Standards (CCSS) movement, and the Every Student Succeeds Act (ESSA). In addition to showing what schools are like under long-term surveillance of state tests, the student demographics in Texas are representative of larger trends nation-wide showing that the students in U.S. public schools are becoming more linguistically diverse and predominantly students of color.

In addition to documenting how testing shapes educational opportunities in schools, this study also seeks to illuminate the ways that teachers are able to make space for teaching and learning not defined by the narrow parameters of the test. Representations of what is possible in schools beyond testing and state-defined metrics of success are needed to show that it is indeed possible to imagine school as something different from the traditional model of education that has not served all students well. I come to this work as a former high school English teacher who also struggled to protect the curriculum from the effects of a rigid accountability system.[i] Some of the teachers at Midgard High, where this study took place, were far more adept at this shielding work than I was as a young teacher; it is my pleasure to share their work with a larger audience through this writing.

Schools like Midgard High already have a high percentage of multilingual students (nearly 40% of enrolled students were designated English Language Learners in 2016-17) and a large percentage of the population who identify as students of color (97% in 2016-17). This is the demographic future of schooling, and so by focusing attention on Midgard, researchers and policymakers can make better-informed decisions about the future of education. Midgard is one of seven high schools in an urban emergent (Milner, 2012) district that served approximately 84,000 students at the time of writing. The distinction of urban emergent is important because University City, the metropolitan area served by the district, is a small but quickly-growing city that shared concerns of schools in densely populated areas: a complicated district bureaucracy, a scarcity of resources, diverse students, and historic inequities in educational achievement among racial, economic, and linguistic groups (Ladson-Billings, 2006; Tyack, 1974).

Literature Review

Policy Implementation: Moving from Legislative Theory to Situated Practice

As policymakers seek to implement large-scale reforms in education, research has documented the challenges of changing teachers' classroom practice. Despite nearly a century of secondary school reform, instructional practices related to reading and writing closely resemble

the classrooms of the early 20th century (Applebee & Langer, 2011; Gillespie, Graham, Kiuhara, & Hebert, 2014; Kiuhara, Graham, & Hawken, 2009). One explanation of this phenomenon is, as researchers first documented in the late 1970s, that teachers tend to treat policy decisions, including curriculum reform, as peripheral to their daily concerns related to teaching and learning (McLaughlin, 1998).

        Much of the research related to policy implementation frames large-scale institutional reform efforts as positive, associating changes in teachers' practice with improved conditions for students in schools as well as improved quality of instruction (Spillane, Reiser, & Reimer, 2002). However, there is also evidence suggesting top-down reform efforts fail to serve students equitably, and policy changes increasing the quantity and high-stakes nature of standardized assessments have been harmful to specific student populations, including Latinx students (McNeil, 2005) and bilingual students (Palmer & Rangel, 2011). Early education policy studies revealed the importance of local contextual factors in mediating and shaping large-scale reform efforts (McLaughlin, 1998), interpreting and making sense of what they have been asked to do by administrative authority figures (Spillane, Reiser, & Reimer, 2002). These studies in educational policy are largely predicated on the assumption of the state's beneficence, yet the history of education in the United States suggests that state institutions have rarely acted in the best interest of students of color, working-class students, and multilingual students (Ladson-Billings, 2006; Paris & Alim, 2014; Tyack, 1974). According to Spillane et al. (2002), as well as foundational research on policy implementation (McLaughlin, 1998), teachers tend to be framed as an obstacle to reform. The underlying assumption in this literature is that administrative bureaucracies, not classroom teachers, know what is in students' best interests.

        In the context of the current movement towards standardization of both curriculum and assessment, research examining literacy policy and implementation has documented how teachers are able to work within the context of rigid curriculum policies (Maniates & Mahiri, 2011; Yoon, 2013) and accountability structures (Achinstein & Ogawa, 2006; Palmer & Rangel, 2011; Zoch, 2015) to advocate for and enact literacy pedagogy that they understand to better support their students' needs and learning goals. This resistance work is difficult when teachers are alone, ideologically isolated in a testing-focused school (Agee, 2004; Smagorinsky, Lakly, & Johnson, 2002). The current study seeks to bridge research on policy implementation with research about classroom implementation of literacy policy, as "historically there has been a divide between those who study policy and those who study subject matter" (Dutro & Valencia, 2004, p. 4). As I will discuss in this paper, policy can have a strong influence on how a subject is taught in schools.

        The notion that what is actually learned in the classroom can be different than what official policy dictates is not a new idea in educational research; Anyon's (1981) hallmark study documented the changes in curriculum-in-use across class lines. In teaching, there is a folk saying that regardless of what noise and demands exist in the outside world, a teacher can always "shut the door" and become the all-powerful decision maker in their own classroom. Whether teachers are framed as obstacles in the way of policies or local agents protecting their students from policies, there is a consensus across both policy and classroom-based literacy studies that teachers' interpretation of and orientations towards policies are critical for understanding the lived effects of policies.

High-stakes Literacy Assessment in Texas

There is a growing body of research concerned with the consequences of high-stakes testing broadly across the country (Au, 2007, 2011) and also in the contexts of Texas and California, which have experienced state-level accountability reform for several decades centered around curriculum standards and standardized assessment. Several of the literacy policy studies referenced in the previous section took place in California, so here I examine research and policy related to Texas teaching contexts to give a sense of teachers' historical experience with accountability. The testing system is so rooted in Texas history that current teachers educated in public schools lived through this system as students. Thus, an analysis of the Texas case shows possible outcomes of long-term standardized testing in literacy, outcomes that teachers and policymakers in states around the country might consider as they adopt accountability measures.

In the current Texas accountability system, English Language Arts (ELA) is the only content area with two high-stakes standardized assessments in high school. Math, Science, and Social Studies have just one exam; Algebra I, Biology, and U.S. History respectively (House Bill 5, 2013). Table 1 presents a comprehensive list of ELA assessments for high school students, along with their corresponding stakes. Note that standardized assessments have been tied to students obtaining a high school diploma since 1986.

Table 1

Texas Standardized Literacy Assessments

Name of Test	Legislation Year	High School ELA Component	Consequences
TABS *(Texas Assessment of Basic Skills)*	1979	9th grade reading basic skills	none (Cruse & Twing, 2000)
	1983	9th grade reading basic skills	re-testing (Cruse & Twing, 2000)
TEAMS *(Texas Educational Assessment of Minimum Skills)*	1986	9th and 11th grade reading basic skills	graduation requirement (11th only)
TAAS *(Texas Assessment of Educational Skills)*	1990	10th grade academic reading, writing, English II	graduation requirement (10th) (TEA Accountability Manual)
TAKS *(Texas Assessment of Knowledge)*	2003-2011	9th, 10th, 11th grade English Language Arts	graduation requirement (11th only)
STAAR *(State of Texas Assessments of Academic Readiness)*	2009-present	9th and 10th grade postsecondary readiness in English	both End of Course (EOC) exams required for graduation (TEC §39.025)

Recent research in the state of Texas examining the effects of the most recent iteration of testing (STAAR) on elementary and middle school literacy indicates this trend towards test-centered instruction has a narrowing and limiting effect on how literacy is defined in schools. Davis & Willson (2015) documented the "managerial partitioning" (p. 372) of literacy practices into discrete tasks isolated from one another and from the contexts in which they occur, rather than teaching literacy practices as mutually influential and interdependent as well as influenced by social context. Understanding reading and writing as contextualized and mutually interdependent, rather than technical skills that function outside of sociopolitical and cultural context (Street, 1985), is an orientation to literacy supported by a long tradition in sociocultural and anthropological research on literacy (Collins & Blot, 2003; Street, 2013). Technical views of literacy like those described in Davis & Willson's (2015) study disguise the ideologically loaded nature of standardized literacy assessments, which privilege a White, middle-class, monolingual variety of the English language (Behizadeh, 2014). This is particularly problematic for linguistically diverse students in Texas who may speak and write non-dominant language varieties, particularly Spanish and Black English (Kinloch, 2010). Additionally, Davis and Willson (2015) found that Texas schools tested with mock benchmarks and subsequently used standardized test data make "consequential inferences that are likely invalid" (p. 372) with respect to students' reading and writing abilities. This narrowing and compartmentalizing of literacy according to tests' definitions of skills and tasks confirm findings from earlier studies of the TAAS testing system (Hoffman, Assaf, & Paris, 2001; Palmer & Rangel, 2011).

There are two reasons why it is important to note that all of these Texas studies relate primarily to elementary literacy contexts. First of all, students in these studies have existed within the testing regime their entire schooling lives, internalizing messages about their ability from the likely misguided inferences teachers and administrators have made based on test scores. Secondarily, there is a dearth of studies that examine literacy and the effects of testing in a high school context, and the limited studies that do exist are limited to the experiences of one early-career teacher (Agee, 2004; Smagorinsky et al., 2002). This study addresses the lack of research on high-stakes testing in secondary schools, filling out the broader picture of the effects of standardized assessment across K-12 education.

Theoretical Lenses

In this paper, I draw on Foucault's (1977/1995) theory of panopticism as well as Greene's (2000, 2010) theory of social imagination and vision to understand how state policy regarding high-stakes literacy assessment was implemented in the local context of Midgard High. Foucault presents the panopticon as a way of understanding how an institution is able to exert complete but indirect control over a population through spatial partitioning, isolation, and constant inspection. "Coercive assignment" (p. 199) or institutional sorting and labeling plays a central role in establishing control, as does the diffusion of authority throughout the system, rather than locating it in a single identifiable figure or event: "the surveillance is permanent in its effects, even if it is discontinuous in its action" (p. 201). For example, tracking students into leveled English classes through standardized assessment is a system "for measuring, supervising and correcting the abnormal" (p. 199), in this case, literacy practices that do not conform to traditional expectations of Whiteness and monolingualism.

Yet unlike Foucault, who considers the surveillance of the panopticon to operate "abstracted from any obstacle, resistance, or friction" (p. 205), I, like Greene (2010), "object to

closed systems, to fixities, to finalities" (n.p.), acknowledging the agency of individuals and groups to resist the control of the state. In addition to naming and recognizing structures that control and punish, actors within the system "strain toward the normative, toward what might be, what ought to be…a social vision of a more humane, more pluralist, more just, and more joyful community" (p. 61). Critiquing post-structural theorists like Foucault, Greene's call for social vision implores educators to "bring renewal to that conversation [about power] to do what we can to include within it the voices of the long silent or unheard" (p. 56). In the case of education policy, the unheard are often classroom teachers; it is in those voices that I find resistance to and friction against institutional control. It is critical to document how teachers work against the harmful effects of standardized testing described above in the review of research literature, providing spaces in their classrooms free from the oppressive gaze of the test. Documenting instances of curricular openness and literacy practices where a plurality of students can find their voices, rather than conform to the demands of testing, provides a vision of hope in politically contentious times when public education is under attack.

Methodology

This study is part of a larger year-long single site embedded case study (Yin, 2014) documenting how one school's English department conceptualized and enacted their vision of English Language Arts, as well as how students understood the work they did in English class. Here the analysis focuses on teachers, using the initial school-level case, the staff of the English department (n = 26). Further embedded cases are the English I (n = 12) and II (n = 8) teams of teachers, as well as one focal class of pre-AP English II taught by Mr. Roman. Mr. Roman was the 10th grade professional learning community (PLC) lead and taught sections of pre-AP English II as well as co-taught sections of special education inclusion English II with Mr. Gomez, an inclusion teacher.

Data Collection

I engaged in participant observation during the 2016-17 school year, first observing teachers in their professional and planning contexts, administering a survey to the entire department in November, and writing ethnographic fieldnotes (Emerson, Fretz, & Shaw, 1995). I selected focal classrooms for observations mid-year, continuing participant observation in professional spaces and rotating between focal classrooms as well as conducting two semi-structured interviews each with the department head and the focal teacher. While on site, I also collected a variety of types of data that represented the ideological environment of the school (Rogers & Mosley, 2006), documenting the texts used in planning and instruction, as well as resources brought to teachers in professional development (PD) contexts.

Data Analysis

Throughout the data collection process I wrote periodic analytic memos to explore emergent themes in the data such as the role of multilingualism, patterns in staffing, and the existence of tracking within the student population. As I engaged in member-checking around tracking policies it became apparent that the hierarchical leveling of students into advanced, on-level, and remedial courses was not an official school policy, rather lingering effects of the high-

stakes literacy assessment. This, combined with an accumulating number of artifacts used in teaching and planning that referenced STAAR, indicated the influence of high-stakes testing within the teaching and learning environment of Midgard High. Having identified an analytic focus, I engaged in data condensation (Miles, Huberman, & Saldaña, 2014) of the full data corpus to identify relevant items that would focus attention on this important factor and allow to me to explore the discourses circulating within the department around testing. I then conducted a two-phase critical discourse analysis of fieldnotes, interview data, and artifacts to account first for the influence of both external structures shaping curriculum and secondarily for the agency teachers exhibited in shaping curriculum to their own ends. Upon entering the first focal classroom mid-year, I noted an absence of test-focused activities. This contrasted with professional spaces such as PD sessions and professional learning community (PLC) meetings, where talk and texts referencing STAAR continued. This two-phased analysis represents the complexity of instructional priorities at the local school level.

Panoptic effects of standardized testing. I coded the data corpus for references to testing, starting with fieldnotes, as they represented the daily social context of the school, as well as interviews, which captured teachers' discourse around testing and their interpretations of policy. I then included any artifacts referenced in interviews or field observations, such as final exams, PD material, and district benchmarks. To verify which assessments were mandated, I engaged in member-checking with the department head and focal teachers, also collecting publicly available samples of official state and district accountability policy.

Finally, I returned to Foucault's conceptualization of the panopticon to link the data to the theory, noticing data that both confirmed and contradicted the theory of the standardized test acting as a panoptic device, structuring the analysis across levels of administrative authority, looking first at the ideological source of the policy: the state, and then, following McLaughlin's recommendations to "link macro and micro levels of policy, analysis and action" (p. 73) throughout an administrative system, grouped the data into district, school administrative, and classroom contexts to highlight the diffuse nature the surveillance structure.

Transforming policy at the local level: Imagining things otherwise. The STAAR and its reverberating power was strong, but neither omnipotent nor omnipresent at Midgard High. To account for the lack of influence in certain spaces, I returned to the data to identify absences of test-centered curriculum, where students had opportunities "to tell their own stories, to pose their own questions, to be present—from their own perspectives" (Greene, 2000, p. 34). The moments where this happened most frequently were during writing instruction, in both English I and II classroom contexts. I coded classroom observations for procedures in writing instruction across multiple classrooms to highlight the "intentionality and concreteness of everyday life" (Greene, 2000, p.10) in these spaces, including genres and topics in students' writing. This second phase of analysis emerged as an important complement to the top-down control of the state test, critical for understanding the localized power teachers had over aspects of their work.

Finally, to examine the intentionality of teachers' decision-making around alternatives to test-focused instruction, I selected the case of Mr. Roman as the next unit of analysis. As the English II PLC he represents the "competing story" of student-focused writing instruction, that, when put into "dynamic relation with [other stories]" (Dyson & Genishi, 2005, p. 111) yields insights about the possibilities for other types of instruction in an urban context: "alternative social relationships and possibilities of things being otherwise" (Greene, 2000, p. 24).

Findings

Concurrent with past research in elementary contexts (Davis & Willson, 2015; Hoffman et al., 2001), findings suggest that the mandated high-stakes ELA exit exams had a strong shaping influence on curriculum and instruction at Midgard High. In the initial findings section I will examine how these two exams served to regulate curriculum decisions, with effects magnifying at each administrative level, from district, school administration, and finally individual classroom contexts. The second findings section addresses how, within this highly restricted context, teachers shielded their writing curriculum from the harsh gaze of the test. As noted previously, it is important in this age of standardization to highlight where teachers are able to create spaces for students to develop their voices and make choices about their literacy learning.

Part I: Panoptic Effects of the State's High-Stakes Standardized Test

State level disciplinary mechanisms. In its official publicity material, the state frames the function of the STAAR as a tool for measuring students' skills in various core content areas, with its primary purpose being "first and foremost…to improve student performance" (Texas Education Agency, 2016, p. 4). Secondarily, the test is officially marketed as a step towards achieving equity in education, framed in vague language regarding disparities in achievement: "closing advanced academic performance level gaps among student groups" and "address(ing) the diversity of student populations" (Texas Education Agency, 2016, p. 4). This acknowledgement of differential achievement as a "gap" is far different than Ladson Billings' call to treat current conditions in education as the results of an "education debt" (Ladson-Billings, 2006, p. 3) owed to students of color. Instead, it privileges hegemonic definitions of literacy achievement centered on White monolingual expectations of performance, pathologizing other groups and mandating documentation of these groups to the centralized state data center. In this way, the structure of surveillance meets the requirements for a panopticon: "registration of the pathological must be constantly centralized" (Foucault, 1977/1995, p. 195) through the yearly (and for re-testers, quarterly) reports to the state's education agency.

From a legislative perspective, the high-stakes nature of the end-of-course (EOC) exam occupies a small portion of the massive Texas accountability apparatus. The Texas Education Code has a mere sentence in the policy invoking the consequences for not meeting the changing scale score requirement, belying the severity of real-life consequences for high school students: "A student may not receive a high school diploma until the student has performed satisfactorily on end-of-course instruments in the manner provided under this subsection" (2H Texas Education Code §39.025). In this way, the assessment reflects how official power can be exercised through relatively small demands. Foucault (1977/1995) notes that in a panopticon, participants might be "surprised that panoptic institutions could be so light" (p. 202), and indeed the secondary literacy accountability structure at the state level is relatively light: two EOC exams, one taken for the first time after English I, traditionally in the 9th grade, and the second taken after English II, the 10th grade.

At the level of the state, mandating two tests over the course of four years is a reduction of the number of tests required in previous assessment structures, and the 2013 revision to the code removed mandated assessments in other content areas. It is not until this policy cascades through localized levels of administration that it acquires its panoptic effects, as I will discuss in

the following sections. Without observing the reality of policy implementation in context, a legislator might be inclined to believe the state's claim of "minim[al] unintended consequences."

District level disciplinary mechanisms. It is important to note that the state delegates surveillance and reporting to individual school districts in its accountability implementation manual: "the system relies on local districts to develop and implement local accountability systems that complement the system" (Texas Education Agency, 2016). In the case of Midgard's local district policy, this translated to PD based on the STAAR, as well as a series of district-wide benchmark exams created in-house by the district's assessment team.

The department head characterized Midgard's supervision from the district as relatively minimal, in comparison to other demographically similar schools at the district: "We're in an interesting place, because most of the district Title 1 schools have a lot more control than we do, because we've managed to keep our scores at a high enough level" (Interview, 2/9/17). In fact, during the 2016-17 school year I only observed the district ELA representative deliver district-sponsored PD three times, all of which focused on STAAR. The resources the district representative presented circulated throughout the department, particularly an item analysis of the past three years of released STAAR tests, a Powerpoint titled "Planning with Purpose" (see Appendix A for sample slide), shown to the department head, then subsequently presented at an English II PD session a week later. Somewhat ironically, "Planning with Purpose" represents the kind of test-centered planning and unfounded inferences about what instruction students might need, documented in an elementary context by Davis and Willson (2015). While there were occasional references to one-on-one reading conferences as an assessment and teaching tool (Fieldnotes 10/12/16), the primary message communicated by the district was that of standardized assessment practices resembling STAAR: multiple-choice questions and practice of the 26-line persuasive and expository essays. With limited physical presence the district had minimal direct supervision of the department, but planning documents circulating throughout the discursive sub-communities of English I and II shows how district surveillance continued through the distribution of resources, not the literal presence of an observer.

Perhaps most influential was the district's complementary local assessment system, a collection of benchmarks related to STAAR but not comprised of released questions, intended to be administered in all district schools after the 11th and 21st week of class. As I will discuss in detail in the next findings sections, the English II team chose not to administer the first middle of the year (MOY) benchmark, but did participate in the second district MOY in January 2017. The benchmark system is an example of how the district acts as a secondary enforcement mechanism, magnifying the surveillance effect of the high-stakes test and ensuring that "each individual is constantly located, examined, and distributed" (Foucault, 1977/1995) within the system. This district-level surveillance through multiple-choice, standard-based tests is something teachers in Texas also noted under the TAKS test: "that's what the district cares more about: are they doing well on the test" (Palmer & Rangel, 2011) and also in Dutro and Valencia's (2004) cross-case policy study: "local views of, and approaches to, the alignment between state and local standards were heavily shaped by local test scores" (p. 33).

While the district claims that the benchmarks were "high quality" (District website, 2017) and focused on students' mastery of skills, teachers saw these assessments as conveying at best mixed messages. The English I PLC met to discuss the results of the first MOY benchmark, having done frequency analysis of correct and incorrect responses on their own planning time. The PLC leaders asked the team to collectively analyze the questions students struggled with most:

Ms. Castle very astutely notes that a huge reason why students may have gotten the incorrect answer to question #8 on the benchmark was that there was a specific vocabulary word embedded in the question: juxtaposition. If you didn't know that word (along with the other words for the answer choices), you could not answer the question. It is therefore equally as likely that the "skill" tested is a particular vocabulary set, not the skill of literary analysis and identifying the literary element (Fieldnotes 10/20/16).

Here a teacher in her analysis observes that the standard the district has assigned to the question, a literary analysis task, is potentially not the skill students needed to have in order to answer the question. The item acts simultaneously as a vocabulary and skill assessment. Without further investigation by the teachers, it would be impossible to tell what students actually struggled with: the cognitively complex skill of literary analysis or the relatively simple recall task of knowing the definition of juxtaposition. This is just one example of many where the skill or correct answer to the question was in doubt, a phenomenon that had become normalized into the department's professional life.

The disciplinary power of the STAAR exam was amplified by its ambiguity, both in the content and skills assessed, as well as in the incredibly influential cutoff scale score for passing, determined by a closed committee after each year's testing. Laden with mystery, the STAAR fulfilled Foucault's description of the panoptic device exercising a "power [that] should be visible and unverifiable" (p. 201). The benchmarks were visible reminders of the upcoming high-stakes assessment, but the individuals responsible for it and its construction (as well as its significance) are hidden throughout an opaque bureaucratic system. Thus, teachers and students in the peripheral edges of the system are "seen [by the test], without ever seeing" the clear meaning of the assessment, whereas state and district officials "in the central tower [see] everything without being seen" (Foucault 1977/1995, p. 202).

School level disciplinary mechanisms. Additionally, the magnified effects of the standardized test appeared in the local school administration's focus on test-preparation, specifically prioritizing standardized test data as a metric for school and student success, linking English course performance to STAAR performance, as well as relying on STAAR data for student scheduling purposes and course offerings.

While state policy claims the accountability system has minimal effects, all of the students enrolled in English I and II classes saw their academic grade in the course fluctuate based on their ability to perform on STAAR exams. School policy was to have each final exam at the end of the fall and spring semesters be a prior year's released STAAR test. This was the case for all sections of 9th and 10th grade English. Despite teachers' enacted curriculum including choice-based independent reading and various writing units from poetry to memoir and short fiction, the final exam for the course was STAAR based, a collection of reading comprehension passages and writing convention questions with multiple-choice answers. While there were logistical constraints in place limiting the possibilities for assessment (final grades were due less than 24 hours after the last final exam, prohibiting more qualitative forms of assessment), the fact remains that students' overall course grades became linked to the supposedly separate state accountability system, evidence the "disciplinary mechanism" (Foucault 1977/1995, p. 197) of STAAR effectively evaluated the students from multiple vantage points, not just the official end of year test.

Teachers curved final exam grades in an attempt to mitigate effects of poor performance, also incorporating corrections or other process-based grades such as identifying keywords in questions and analyzing answer selections, when STAAR release tests or STAAR-like materials

were used in daily instruction, effectively bumping up students' overall grades. Yet regardless of adjustments, students at multiple points of the year found their academic grades influenced by their STAAR performance. This double surveillance of students, through official yearly scores and additionally through course grades, is evidence "the panopticon is a marvellous [sic] machine which…produces homogenous effects of power" (Foucault 1977/1995, p. 202).

An even more consequential effect of the high-stakes standardized test was the effect that it had on staffing and course offerings. Here official STAAR scores served as a tool for "disciplinary partitioning" (Foucault 1977/1995, p. 199) of the student body. All 9th grade students, regardless of performance, received a "double block" of ELA instruction amounting to 7.5 hours per week, with low-performing students enrolled in an additional course. 10th grade students enrolled in pre-AP sections of English II enrolled in one section of ELA for a maximum of 4.5 hours of ELA instruction per week, whereas all regular and special education inclusion classes were double blocked. 10th grade also had additional remedial sections for students who had not passed their English I EOC exam. The consequences of this scheduling is that students who received low STAAR scores had up to 12 hours per week, or 40% of their entire schedule, devoted to ELA instruction. Scheduling created a "constant division between the normal" students who had passed their EOC exams "and the abnormal" (Foucault 1977/1995, p. 199) who had not. This tracking was not official school policy, but as noted earlier, indirect regulatory effects of the test.

The effects of scheduling extended beyond students' schedules (keeping low-performing students from participating in more varied courses and electives) but increased teacher workload elsewhere. Class sizes in non-tested ELA courses (English III and IV) were significantly higher than tested grades; there were up to 36 students enrolled in each English IV class, and the one English IV teacher was responsible for teaching over 200 students. In contrast, the English I PLC consisted of 12 staff members, each of whom, because of double-blocking, taught no more than 75 individual students. The increasing depersonalization of ELA instruction beyond tested classes was evidence of the value the administration placed on test results, as the material resources of teachers' salaries were overwhelmingly directed towards students in English I and English II (see Appendix B for a full staffing chart), limiting the possibility of offering electives and upper-level English courses.

Classroom level disciplinary mechanisms. Because the focus of the larger study was not high-stakes testing, my research design intentionally avoided observing in classrooms with frequent test preparation, specifically English I and II "End of Course" (EOC) classes required of students who had failed one English STAAR: "All 3 EOC teachers share that they do open ended response practice every day in their EOC courses" (Fieldnotes 10/12/16).

However, STAAR-like instructional tools were used throughout the department, not just in official test-preparation classes. In teachers' response to 3-point Likert-scale questions on the fall survey (n = 22 responses), 16 rated multiple-choice questions somewhat important in their instruction, and 3 very important; 14 rated short-answer questions (SAQ's) very important. From these responses, it is clear that many teachers valued the forms of assessment seen in the high-stakes test and saw using these assessments in their instruction as important for their jobs. I frequently observed teachers in the English I PLC reference multiple-choice questions as shared assessments they used in the teaching of literature.

In addition to some teachers using the forms of assessment privileged on the test, some teachers also used the test to define their jobs and the content they taught. An English III and EOC prep teacher declared in his survey responses "they have to pass a series of tests to

graduate…they MUST learn to write to the formula that is required for the cadre of tests and they need to learn to express and communicate." When asked why students should have to take an English class, a 9th grade teacher wrote simply "test preparation." An ESL teacher defined her influences on teaching as "my past teachers and what is expected of my students on STAAR." To contextualize these responses within the department's general understanding of their content and professional goals, the teacher survey did not reference the STAAR, and only six teachers explicitly referenced standardized testing in their free responses. So while some teachers' understandings of their work as English teachers were strongly influenced by the exam, this was only the case for a minority of teachers.

The test was also present in professional spaces from the beginning of the year. I observed an early meeting of the 9th grade PLC:

> Discussion shifts to the next diagnostic assessment, which will take place on 10/11 or 10/12, and appears to be a voluntary benchmark…There is talk about doing a STAAR blitz prior to the benchmark, and the necessity to "see where they are" with reference to the end of year assessment. Teachers work to determine which sections of pre-prepared materials should form this initial benchmark, drawing upon a STAAR prep booklet. (Fieldnotes, 9/26/16)

In this planning session it is possible to observe how the surveillance effect of the test radiates out from the state's centralized mandates, showing how the institutional surveillance mechanism of the test can be in practice "democratically controlled" (Foucault, 1977/1995, p. 207) by individuals within the system without much institutional power. These teachers were thinking long-term about the course of the year, using the state exam as the central focus of their planning and mimicking the district's behavior of administering benchmark assessments.

By complying with all district-mandated assessments and additionally adding in another diagnosis and testing cycle, the English I team's decisions show that "anyone may come and exercise in the central tower the functions of surveillance" (Foucault, 1977/1995, p. 207). To maintain a state of surveillance it is not necessary to have a central oppressive figure exercising direct control over regulatory systems; all that is necessary is for ordinary actors in the system to maintain the "disciplinary programme" with interim testing, voluntarily engaging in surveillance themselves.

Conclusion: panoptic effects across the system. The STAAR's influence emerged across and within each of these different levels of power, solidifying the panoptic effect: "so to arrange things that the surveillance is permanent even if it is discontinuous in its action" (Foucault, 1977/1995). The state need not mandate any more assessments when the system is so solidly in place to structure teaching and assessment around an understanding of reading and writing as constructed by the state's standards and in the form of the exam itself, regardless of the fact that the test is administered only as an annual exam.

The reality of the test acting as a disciplinary mechanism is that even an advanced 9th grade student in a pre-AP English class was assessed by STAAR-like tests in a formal capacity five times: two district MOY benchmarks, two final exams, and the test itself, representing at least 15 instructional hours, the equivalent of 1½ weeks of school. As 9th graders, these students are under "the constant pressure [of the test] even before the offences, mistakes, or crimes have been committed" (Foucault, 1977/1995, p. 206), marked as "individuals on whom a task or a particular form of behavior must be imposed [in the] panoptic schema" (p. 205).

A student in 10th grade who did not pass the English I STAAR would have taken one district MOY, the English I December STAAR re-test, two final exams, and contingent upon the

December re-test, both the English I and II end of year STAAR. This is up to six test administrations, or two full weeks of school spent on assessment, not teaching. Considering the case of the 10th grade team, the least test-focused in its planning and instructional choices, students had minimally 6 weeks (four weeks in the spring and two weeks around the December re-test) of STAAR instruction plus one week of test-taking. This means that at least one sixth of the school year was devoted to test-centered instruction. The frequency of testing served to restrict students' curriculum within a school year and the effects of low performance restricted students' access to rigorous curriculum and elective opportunities across their high school experience.

Teachers demonstrated an awareness of surveillance to different degrees and incorporated talk about the STAAR into their regular professional discourse within PLC meetings. This finding contradicts some historical research on policy implementation, particularly McLaughlin's work with federal reform programs, which found that "teachers rarely saw policy or organizational boundaries as critical influences on their work" (McLaughlin, 1998, p. 74). In the local context of Midgard, state accountability policy was very much on educators' minds and influenced their work.

Part II: Imagining Something Different for English

As Greene (2000) notes, schools are "largely hierarchical, bureaucratic institutions [which]…by their very nature make it difficult for openings to be explored and critical thinking to take place" (p. 56), as demonstrated in the first findings section. While STAAR shaped much of the curriculum and instruction at Midgard High, there were also significant ways in which teachers thought both beyond and critically about the test. In this way, the "implementation problem" (McLaughlin, 1998, p. 70) became a success in policymaking. Teachers redefined their instruction and vision for teaching according to their own professional standards and knowledge, not according to the state's decontextualized and fragmented vision for what ELA teaching should be.

Writing workshop: amplifying students' voices and minimizing STAAR. The context where the ELA curriculum was the most open and responsive to students was in English I and II writing instruction. While there were moments in other contexts that appeared free from the test, it was when teachers planned and enacted their writing curriculum that showed the most contrast to STAAR, because it was a space where students could "tell their own stories…pose their own questions...[and] be present from their own perspectives" (Greene, 2000, p. 24). Within writing instruction, many teachers practiced a workshop approach to teaching, incorporating student choice in writing topics and independent work time in the classroom. They also made clear distinctions between test-preparation and real-world writing tasks, and limited students' exposure to the STAAR writing genre: a 26-line timed essay written in response to a prompt.

While teachers enacted a workshop approach to teaching writing (Bomer, 2011) to varying degrees, many of combined English I and II faculty (n = 20) chose to teach writing in ways that supported student independence and highly individualized teaching, the antithesis of the pre-determined and standardized curriculum related to the high-stakes test. I observed this writing instruction on a regular basis in Mr. Roman's 10th grade pre-AP class, occasionally in Ms. Abramson's 9th grade special education inclusion class, and in five other classrooms on pop-in visits with a PD consultant.

The most striking feature of these writing classes was that teachers framed their writing units as genre study (Ray, 2006) of texts that exist in the world. Often school writing products are texts intended for the audience of the teacher, never leaving the world of the classroom. As Whitney (2011) notes "what are the cousins of the 'research' paper that live and breathe outside of school?" (p. 60). Students at Midgard wrote in living and breathing genres of memoir, op-eds, horror stories, short fiction, poetry, and children's books. Four teachers taught a unit at the end of the year where students created texts in a genre of their own choice, focusing on how writers make texts that work for social change. There were no arcane rubrics or writing formulas attached to these real-world genres: teachers and students studied examples of "mentor texts" (Bomer, 2011, p. 236), establishing patterns they saw as readers rather than giving that authority to test-makers. Here, teachers treated students as "potential active learners who can best learn if they are faced with real tasks and if they discover models of craftsmanship and honest work" (Greene, 2000, p. 13), free from the factory-style monotony of A, B, C, or D answer choices. They accepted "the challenge…to refuse artificial separations of the school from the surrounding environment, to refuse the decontextualizations that falsify so much" (Greene, 2000, p. 11) in education.

Another way teachers added context and humanity into the writing curriculum was by encouraging students to choose topics of their own interest. Across the aforementioned genres of text, students wrote about the rising costs of living in their gentrifying city, visiting Mexico for the Feria de San José, soccer cleats, the births of siblings, rusty trucks, problems in the foster care system, a house fire, Jim Crow segregation, sports cars, immigration journeys, basketball, the newly-elected president, and many more ideas that they decided were worthy of their time. During independent work time, I observed:

> Mr. Gomez talks to a student about choosing a topic, and he asks that student, "What do you know? You know manga, cheerleading, baseball, you know your phone"…I hear Ms. Novak conference with a student, explaining, like Mr. Gomez has just done, how to draw on topics that are personally meaningful and close to the student's experience. She says "If it was my life, [I'd write about] my daughter's sippy cup—the spills and stains. But that's my life. What's your life?" (Fieldnotes, 1/20/17)

Conversations like these allowed students to practice choosing topics of importance, a high-value skill valued by college-level instructors that is not often taught in secondary contexts (ACT, Inc., 2016).

Test preparation as genre study: limiting exposure to STAAR. Test preparation in English II and in Mr. Roman's classroom in particular was another genre study (Hornof, 2008): as modeled by a writing PD consultant, the team designed a unit in which students did item analysis, discovering patterns within the way questions were asked and the kinds of answers the test privileged. Additionally, his English II classes immersed themselves in STAAR persuasive essays, developing patterns and trends in this writing genre as they practiced writing responses themselves. The test-preparation unit, 4 weeks out of the yearly curriculum, was structurally and procedurally similar to the overall workshop approach he used during the rest of the school year. Rather than mechanically completing worksheets and relying on commercial STAAR preparation materials, students deconstructed the actual released tests, with guidance from their teachers.

The way Mr. Roman reflected on the purpose of English classes was:

> If I'm thinking like grand scheme of life, I'm trying to make them lifelong readers and writers, and increase communication skills, get them ready for the real world…And short term, and depending on the time of year, pass the STAAR test. (Interview 12/16/16)

Here he positions the test as bounded by time, with its concomitant effects also representing "short term" influences on students' reading and writing lives. His perception of the test and its role in his classroom was not as an omnipresent guiding force, but a reality that could be addressed within a specific time window with as little fear and intimidation as possible.

Teacher decision-making around the test and its "fake world." Mr. Roman, the English II PLC lead, was highly aware of the ideological contrast between the writing curriculum he and his team enacted, and the artificiality of the high-stakes test. He openly acknowledged this tension, also examining how the test operated in the sociopolitical context of working at a Title 1 school with few material resources:

> As a Title 1 school, at the end of the day, those scores do matter to some degree, both in the public eye and what policies are gonna come through here...And so there's like a practical aspect of, we can't always just prepare for the real world because there's also this fake world of testing that we have to get ready for. So that's kind of, a weird push and pull. (Interview 12/10/16)

Teaching for the "real world" of reading and writing meant using models of texts that circulate in specific social contexts discussed in the prior section. Attending to the "fake world" of testing meant doing what he could to prepare students to be successful in composing a persuasive text for an anonymous audience of test readers, timed, in a silent testing room.

Additionally, Mr. Roman exposed another underlying tension in test-preparation curricula: If students do not perform well after formulaic practice, it is possible to blame students for their low achievement, rather than placing blame on a system that privileges particular types of knowledge over others and has historically underserved certain populations of students, particularly with respect to writing assessment (Behizadeh, 2014; Ladson-Billings, 2006). With a standardized curriculum, he explains a possible teacher rationale: "we can write [low scores] off as well, 'nothing really works, it's these damn kids.' It's just you know 'we're giving them the info, it's just not sinking in, what do you want us to do?' Okay move on" (Interview 12/10/16). This statement contrasts with his own thinking and practice about teaching: that he's moved away from a banking-style (Freire, 1970) orientation towards learning where students were expected to copy notes about literary terms from district-made Powerpoint slide, and instead, has taken up a more individualized way of teaching. In this sarcastic aside, Mr. Roman also articulates an awareness of deficit perspectives directed at his students as a result of their institutional labeling by the STAAR, showing his "wide awakeness, awareness of what it means to be in the world" Greene, 2000, 25) charged with negative assumptions about the cognitive ability of "these damn kids." An educational system that blames the students for failing to measure up to normative expectations of schooling is not what Mr. Roman believes to be right and just.

Finally, Mr. Roman explained that he felt ethically and professionally validated by this decision to abandon rigid standardized curriculum:

> When a kid asks why are we doing this, [and I say] it's for the test, I don't have to say that as much [this year], cause sometimes that kills a little part of me inside…Are we just finding filler things all the time that even though we know on a deep level don't really work? (Interview 12/10/16)

In transitioning to genre study units and workshop instruction, he found better answers for students about why they're doing the work of their English class. Biding time and doing "filler" assignments were not fulfilling to him as a teacher, and he did not see these tasks as beneficial to student learning. In contrast to teaching with worksheets and repetitive practice as he had done before, he saw this new writing workshop instruction as positive for students' development and learning, more intellectually and personally meaningful.

Conclusion: imagining things otherwise. Like Palmer and Rangel (2011) found in elementary bilingual contexts, this study contradicts the notion that teachers are automatons dominated by state power structures. Greene (2000) theorizes that teachers can "refuse artificial separations of the school from the surrounding environment, to refuse the decontextualizations that falsify so much" (p. 11) in education. Teachers at Midgard were able to advocate for and enact a vision of writing instruction that was individualized, flexible, grounded in real-world text genres, and centered on students' lived experience, rather than standardized and imposed by the state. Through writing, students were invited to bring their lives and realities into the classroom, to be acknowledged as having voices and meaningful things to say to the world.

In the bustling world of secondary school, education operates at an industrial scale. Education is too often dehumanizing in this context, particularly when the institution privileges normative ways of being, erasing the experiences of historically marginalized students. As Greene notes, "the invisibility of too many students has somehow to be broken through" (p. 15). Writing made students visible to their teachers as unique individuals with histories, complexity, and literate abilities, and teachers like Mr. Roman felt validated by this process-based approach to teaching writing, connected to text genres that existed in social worlds beyond school.

It is important to remember that not all of the teachers at Midgard structured their writing curriculum with genre study, and some classes finished fewer writing cycles than others. A key factor that seemed to facilitate Mr. Roman's transition to workshop and genre study was his flexible orientation towards curriculum and planning, as well as his willingness to take risks in allowing for curricular openness. For example, he explains that as students collectively create rules or "noticings" about a genre, the teacher can't entirely pre-plan instruction:

> To put it on the kids to notice things, that's a big risk, it's really scary to me every time still…but I think it works out a surprisingly high percentage of the time. I shouldn't have worried, it came out fine. (Interview 3/10/17)

The shift in teaching to "provoke and release [ideas] rather than to impose and control" (Greene, 2000, p. 57) means that teachers must be able to sustain uncertainty and respond to their students in the moment of teaching, rather than come into teaching contexts with answers and products predetermined. As he notes: teachers can "rehearse it, but once it actually starts, like the kid brains get to work and they start noticing what they notice, it's different every time" (Interview 3/10/17). If teachers work with diverse set of human beings, it would make sense that this would be complicated. Further, "It is difficult to affirm the values of plurality and difference while working to build a community of persons who have a feeling of agency, who are ready to speak for themselves" (Greene, 2000, p. 42). It is difficult work, but it is possible—it's work the teachers at Midgard did all year long.

Implications

If we are committed to equity in schools and loosening the testing regime's stranglehold on curriculum and instruction, it is imperative to work not only at the state legislative level, but

also within school districts and individual campus administrations to advocate for more rigorous and responsive ways of assessing what students can do with language, reading, and writing. As the case of Midgard High school has shown, state policy is but a small part of the system of surveillance and discipline that shapes schooling. Districts, serving as the state's proxy enforcers of discipline, played a much larger role in intimidating and pressuring teachers to use standardized measures of literacy to structure their teaching, yet rarely followed up to provide teachers with meaningful information about students, confirming Davis and Willson's (2015) disturbing findings about the ambiguity embedded in these types of assessments and their dubious usefulness as meaningful tools to inform instruction.

If educators truly seek to cultivate skills that relate to students' success in higher education and develop students' ability to see writing as a valuable tool for communication and expression, writing instruction must relate to real-world text genres and take place over time, embedded in social contexts, rather than be isolated to the 6 hours and 26 lines of the STAAR. The teachers at Midgard taught in this way often; more schools should follow their lead. Additionally, moving to portfolio-based writing assessments would allow educators to account for a multiplicity of voices and linguistic registers (Behizadeh, 2014), rather than a vaguely defined notion of standard 'correctness' as enforced by the anonymous STAAR readers who have no personal knowledge of the writers whose work they assess. This kind of anonymous writing assessment is loaded with discriminatory ideologies that penalize non-dominant language practices (Davila, 2012).

Reading instruction remained traditional for most of the English I team, frequently assessed with multiple-choice questions, and while there was material support for reading from the school district in the form of high-interest classroom libraries of young adult literature, there was no accompanying professional development to support teachers using choice reading in their daily instruction. Literacy professional development provided for Midgard's English department through a two-year grant from the state was intended to focus on writing instruction, however teachers also saw the potential in moving towards a workshop approach in reading after seeing increases in students' writing motivation and fluency. It is odd that a state and district bureaucracy obsessed with the STAAR, which demands students to complete more reading tasks than writing tasks, would provide almost zero professional support for the teaching of reading. Teachers deserve high quality professional development sustained over time (Skerrett, Warrington, & Williamson, in press), addressing the challenges of teaching reading to diverse learners, a difficult task in schools with scarce resources (Williamson, 2017).

There are currently few incentives to recognize good teaching (Cochran-Smith, 2001), as most accountability systems are structured around punitive measures. Furthermore, many urban schools do not see a commitment to fostering student voice and autonomy in literacy as connected to educational equity and rigorous academic standards (Quartz, 2003). Teachers like Mr. Roman and others in the English department should be recognized for their work in providing high-quality, student centered writing instruction. This is critical, particularly in the context of urban schools, which are often represented as spaces of pedagogical dysfunction and low-quality teaching. More specifically, there should be institutional mechanisms to support this kind of professional commitment to high-quality literacy instruction that allows for student autonomy and decision-making in writing, skills that college instructors across the country indicate are lacking in current incoming undergraduates (ACT, Inc., 2016).

Change in professional practice is slow, and oftentimes not immediately captured by standardize measures of literacy. However, research indicates that commitments to choice based

literacy instruction can have durable effects on students' achievement outcomes (Francois, 2012, 2013). The 2016-17 school year was the second in a two-year professional development program for writing and was Mr. Roman's first full school year implementing this kind of instruction. With administrative support and continued professional development, Midgard might be able to sustain the change that was under way. Without explicit support, it is likely that instruction will return to the status quo, this year's changes engulfed by the powerful gaze of the test.

––––––––––

Thea Williamson is a doctoral candidate in the Language and Literacy Studies program at the University of Texas at Austin. Her research relates to the teaching and learning of reading and writing in culturally and linguistically diverse contexts, literacy policy, and teacher education.

References

Achinstein, B., & Ogawa, R. (2006). (In)Fidelity: What the resistance of new teachers reveals about professional principles and prescriptive educational policies. *Harvard Educational Review*, *76*(1), 30–63.

ACT, Inc. (2016). *ACT national curriculum survey 2016*. Iowa City, IA: ACT Inc.

Agee, J. (2004). Negotiating a teaching identity: An African American teacher's struggle to teach in test-driven contexts. *Teachers College Record*, *106*(4), 747.

Anyon, J. (1981). Social class and school knowledge. *Curriculum Inquiry*, *11*(1), 3–42.

Applebee, A. N., & Langer, J. A. (2011). "EJ" extra: A snapshot of writing instruction in middle schools and high schools. *English Journal*, *100*(6), 14–27.

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, *36*(5), 258–267.

Au, W. (2011). Teaching under the new Taylorism: High-stakes testing and the standardization of the 21st century curriculum. *Journal of Curriculum Studies*, *43*(1), 25–45.

Behizadeh, N. (2014). Mitigating the dangers of a single story: Creating large-scale writing assessments aligned with sociocultural theory. *Educational Researcher*, *43*(3), 125–136.

Bomer, R. (2011). *Building adolescent literacy in today's English classrooms*. Portsmouth, NH: Heinemann.

Cochran-Smith, M. (1991). Learning to teach against the grain. *Harvard Educational Review*, *61*(3), 279–310.

Cochran-Smith, M. (2001). Learning to teach against the (new) grain. *Journal of Teacher Education*, *52*(1), 3.

Collins, J., & Blot, R. (2003). *Literacy and literacies: Texts, power, and identity*. New York: Cambridge University Press.

Cruse, K. L., & Twing, J. S. (2000). The history of statewide achievement testing in Texas. *Applied Measurement in Education*, *13*(4), 327–331.

Cuban, L. (2013). Why so many structural changes in schools and so little reform in teaching practice? *Journal of Educational Administration*, *51*(2), 109–125.
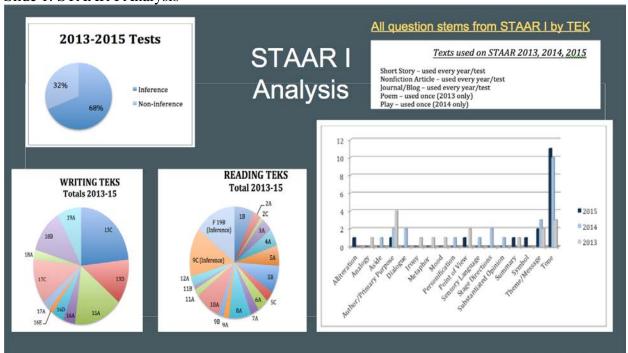
Davila, B. (2012). Indexicality and "standard" edited American English: Examining the link between conceptions of standardness and perceived authorial identity. *Written Communication*, *29*(2), 180–207.

Davis, D. S., & Willson, A. (2015). Practices and commitments of test-centric literacy instruction: Lessons from a testing transition. *Reading Research Quarterly*, *50*(3), 357–359.

Dutro, E., & Valencia, S. (2004). The relationship between state and district content standards: Issues of alignment, influence and utility. *Education Policy Analysis Archives*, *12*(45), 1-49.

Dyson, A. H., & Genishi, C. (2005). *On the case: approaches to language and literacy research*. New York: Teachers College Press/NCRLL.

Emerson, R. M., Fretz, R. I., & Shaw, L. L. (1995). *Writing ethnographic fieldnotes* (1st ed.). Chicago, IL: University of Chicago Press.

Foucault, M. (1995). Panopticism. In A. Lane (Trans.), *Discipline and punish: the birth of the prison* (2nd Vintage Books ed, pp. 198–228). New York: Vintage Books (Original work published 1977).

Francois, C. (2012). Getting at the core of literacy improvement: A case study of an urban secondary school. *Education and Urban Society,* *46*(5), 580-605.

Francois, C. (2013). Reading in the crawl space: A study of an urban school's literacy-focused community of practice. *Teachers College Record*, *115*(5), 1-35.

Freire, P. (1970). *Pedagogy of the oppressed*. (M. B. Ramos, Trans.) (30th Anniversary Ed.). New York, NY: Bloomsbury Academic.

Gillespie, A., Graham, S., Kiuhara, S., & Hebert, M. (2014). High school teachers' use of writing to support students' learning: A national survey. *Reading and Writing*, *27*(6), 1043–1072.

Greene, M. (2000). *Releasing the imagination: Essays on education, the arts, and social change*. San Francisco, CA: Jossey-Bass.

Greene, M. (2010). Prologue to art, social imagination and action. *Journal of Educational Controversy*, *5*(1), n.p.

Hoffman, J. V., Assaf, L. C., & Paris, S. G. (2001). High-stakes testing in reading: Today in Texas, tomorrow? *The Reading Teacher*, *54*(5), 482–492.

Hornof, M. (2008). Reading tests as a genre study. *The Reading Teacher*, *62*(1), 69–73.

House Bill 5. §74.1021 Curriculum requirements; Subchapter BB. Commissioner's rules concerning high school graduation, §28.025(h) and (h-2) Texas Education Code § 83rd Texas Legislature (2013). Retrieved from http://ritter.tea.state.tx.us/rules/tac/chapter074/ch074bb.html

Kena, G., Hussar, W., McFarland, J., de Brey, C., Musu-Gillette, L., Wang, X.,…Dunlop Velez, E. (2016). *The condition of education 2016* (No. NCES 2016144). Washington, DC: National Center for Education Statistics.

Kinloch, V. (2010). "To not be a traitor of Black English" : Youth perceptions of language rights in an urban context. *Teachers College Record*, *112*(1), 103–141.

Kiuhara, S. A., Graham, S., & Hawken, L. S. (2009). Teaching writing to high school students: A national survey. *Journal of Educational Psychology*, *101*(1), 136–160.

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

Maniates, H., & Mahiri, J. (2011). Post-scripts: Teaching reading in the aftermath of prescriptive curriculum policies. *Language Arts*, *89*(1), 10–21.

McLaughlin, M. W. (1998). Listening and learning from the field: Tales of policy implementation and situated practice. In A. Hargreaves, A. Lieberman, M. Fullan, & D. Hopkins (Eds.), *International Handbook of Educational Change* (pp. 70–84). Boston, MA: Kluwer Academic Publishers.

McNeil, L. (2005). Faking equity: High-stakes testing and the education of Latino youth. In A. Valenzuela (Ed.), *Leaving children behind: How "Texas style" accountability fails Latino youth* (pp. 57–111). Albany, NY: State University of New York Press.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Thousand Oaks, CA: SAGE Publications, Inc.

Milner, H. R. (2012). But what is urban education? *Urban Education*, *47*(3), 556–561.

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.

Paris, D., & Alim, H. S. (2014). What are we seeking to sustain through culturally sustaining pedagogy? A loving critique forward. *Harvard Educational Review*, *84*(1), 85–100.

Quartz, K. H. (2003). "Too angry to leave:" Supporting new teachers' commitment to transform urban schools. *Journal of Teacher Education*, *54*(2), 99–111.

Ray, K. W. (2006). *Study driven: A framework for planning units of study in the writing workshop*. Portsmouth, NH: Heinemann.

Rogers, R., & Mosley, M. (2006). Racial literacy in a second-grade classroom: Critical race theory, Whiteness studies, and literacy research. *Reading Research Quarterly*, *41*(4), 462–495.

Skerrett, A., Warrington, A., & Williamson, T. (in press). Generative principles for professional learning for equity-oriented urban English teachers. *English Education*.

Smagorinsky, P., Lakly, A., & Johnson, T. S. (2002). Acquiescence, accommodation, and resistance in learning to teach within a prescribed curriculum. *English Education*, *34*(3), 187–213.

Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research*, *72*(3), 387–431.

Street, B. (1985). *Literacy in theory and practice*. New York, NY: Cambridge University Press.

Street, B. (2013). Literacy in theory and practice: Challenges and debates over 50 years. *Theory into Practice*, *52*(1), 52-62.

Texas Education Agency. (2016). *2016 Accountability manual for Texas public school districts and campuses*. Austin, TX: Texas Education Agency.

Tyack, D. (1974). *The one best system: A history of American urban education*. Cambridge, MA: Harvard University Press.

Tyack, D., & Tobin, W. (1994). The "grammar" of schooling: Why has it been so hard to change? *American Educational Research Journal*, *31*(3), 453–479.

Whitney, A. E. (2011). In Search of the authentic English classroom: Facing the schoolishness of school. *English Education*, *44*(1), 51–62.

Williamson, T. (2017). Listening to many voices: Enacting social justice literacy curriculum. *Teaching and Teacher Education*, *61*, 104-114.
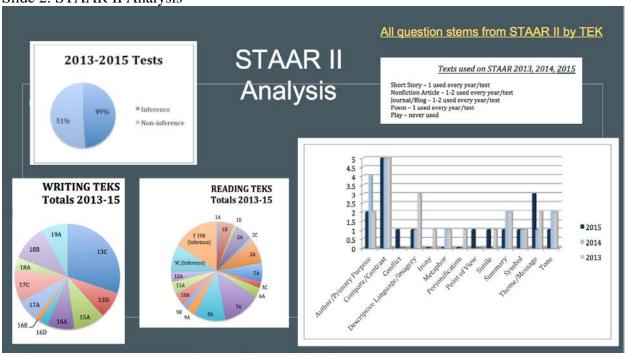
Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). Los Angeles, CA: SAGE
　　　Publications, Inc.

Yoon, H. (2013). Rewriting the curricular script: Teachers and children translating writing
　　　practices in a kindergarten classroom. *Research in the Teaching of English*, *48*(2), 148–
　　　174.

Zoch, M. (2015). "It's important for them to know who they are": Teachers' efforts to sustain
　　　students' cultural competence in an age of high-stakes testing. *Urban Education*, *52*(5),
　　　610-636.

Appendix A: Excerpts from "Planning with Purpose" Powerpoint

Slide 1: STAAR I Analysis

Slide 2: STAAR II Analysis

Appendix B: Staffing Chart

	Course Title	**Track Description**	**Teachers**
9th grade	Pre-AP English I (all students scheduled in double-blocked sections, with Literary Genres)	Advanced	Ms. Elliot Mrs. Abramson Mrs. Henderson
	English I (all students scheduled in double-blocked sections, with Literary Genres)	On-level	Mrs. Abramson Mr. Black Ms. Castle Ms. Greenwood Mrs. Henderson
	SPED Inclusion English I (all students scheduled in double-blocked sections, with Literary Genres)	Low-level	Mrs. Abramson Mrs. Henderson *Mr. Rossi (Inclusion)* *Mr. Harris (Inclusion)*
	Literary Genres (the double block for English I)	Both Low, Advanced, and On-Level	Mrs. Abramson Mr. Black Ms. Castle Ms. Elliot Ms. Greenwood Mrs. Henderson Mrs. Henderson *Mr. Rossi (Inclusion)* *Mr. Harris (Inclusion)*
	Creative Writing (Additional remedial class)	Low-level (struggling)	Ms. Elliot Ms. Greenwood Ms. Bianchi Ms. Prairie
	ESOL Reading	Low-level	Ms. Willis Ms. DiFara
	ESOL Sheltered English I	Low-level	Ms. Washington Mr. Uhr Mrs. DiFara
10th grade	Pre-AP English II	Advanced	Mr. Roman Ms. Roy
	English II (all students scheduled in double-blocked sections, with Practical Writing)	On-level	Ms. Bianchi Ms. Rin Mrs. MacDuff\Ms. Roy* Ms. Smart
	SPED Inclusion English II (all students scheduled in double-blocked sections, with Practical Writing)	Low-level	Ms. Smart Mrs. Novak Mr. Roman *Mr. Gomez (Inclusion)*

	Technical Writing (additional remedial class)	Low-level (struggling)	Ms. Fox Ms. Prairie
	Practical Writing (the double block for ELA II)	Both On-Level and Low	Ms. Bianchi Mrs. MacDuff\Ms. Roy* Mrs. Novak Ms. Rin Mr. Roman Ms. Smart *Mr. Gomez (Inclusion)*
	SPED Sheltered English II	Low-level	Ms. Christie
	ESOL Sheltered English II	Low-level	Mr. Uhr
11th grade	Regular Track English III	On-level	Mr. Finch Mr. Hirschorn
	EOC Prep (for students who still need to pass state test)	Low-level (struggling)	Mr. Finch Mr. Hirschorn
	(see ESOL II)	Low-level	Same as above for 10th grade
	SPED Sheltered Combined 11th and 12th	Low-level	Ms. Christie
12th grade	Regular Track English IV	On-level	Mr. O'Connor
	EOC Prep (for students who still need to pass state test)	Low-level (struggling)	Ms. Fox
	SPED Sheltered Combined 11th and 12th		Ms. Christie
ELA Courses offered to students at Midgard High, not by staff in the English Department			
	Credit Recovery 1. DELTA, 2. Twilight: evening program for online credit options	Remedial	Online content delivery (no teachers, just proctors)
	Credit Recovery: Jumpstart	Remedial	Certified teacher, any content area, i.e. Ms. Castle (1st year) or art teacher
	Dual Credit (ACC Instructor)	Advanced	ACC instructors

Mrs. MacDuff passed away in January 2017; Ms. Roy, hired as a supplemental instructor in January teaching reduced-sized classes, took over Mrs. MacDuff's regular teaching schedule.

---

[i] I taught English Language Arts, reading, and a remedial test-preparation course at an urban intensive (Milner, 2012) school in Miami, Florida in the mid 2000s. Florida, like Texas, has a long history of standardized literacy

assessment, and the Florida Comprehensive Assessment Test (FCAT) affected the school, students, and curriculum much like the STAAR did in Texas. The test-preparation curriculum at the high school where I taught consisted of students completing a corporate-branded FCAT practice booklet.