Language Arts Journal of Michigan

Volume 32
Issue 2 *Race, Language, & Privilege*

Article 10

5-2017

The Consequential Validity of the M-STEP and Third-Grade Retention

Michelle L. Sprouse
University of Michigan - Ann Arbor, sprouse@umich.edu

The Consequential Validity of the M-STEP and Third-Grade Retention

Cover Page Footnote
The author thanks Dr. Anne Gere and Emily Wilson for their feedback on early stages of the manuscript.

Research

The Consequential Validity of the M-STEP and Third-Grade Retention

MICHELLE SPROUSE

In August 2015, Michigan House Representative Amanda Price introduced HB 4822 to the Michigan Legislature calling for "attainment of [a] minimum benchmark on state reading assessment" as a requirement "for promotion of third grade students to [the] next grade level" (Price, 2015). The bill quickly proceeded through the House and Senate after the release of disappointing 2015 NAEP scores. Governor Rick Snyder signed HB 4822 "with immediate effect" on October 19, 2016 with the goal "that more pupils will achieve a score of at least proficient in English language arts on the grade 3 state assessment" (Price, 2015, p. 1).

This legislation demands "evidence-based" instructional programs and strategies, but does not require the same for the ELA assessment or retention policy. As our legislators push for reading reform, I ask to what extent the mandated assessment and consequences are themselves evidence-based.

The Roots of Third-Grade Retention

While test-based retention policies date back as far as 1996, the idea has much older roots (Huddleston, 2014). Dee Norman Lloyd's (1978) report "Prediction of School Failure from Third-Grade Data" argues that dropout predictions could be made from test results as early as the third grade. However, Lloyd (1978) admits "the third was chosen as a point for comparison because standardized tests were given in that grade" (p. 1194). The root of retention laws in Michigan and across the country appear based on the existence of a set of data at that grade, not studies that third grade is in fact the best time to retain students.

Today, the 2001 No Child Left Behind Act mandating standardized testing beginning in third grade "provide[s] a ready-to-use mechanism for including the results of standardized testing in retention decisions" (p. 114). Additional studies that show student outcomes are *independent* of the grade in which they are retained reinforce the arbitrariness of third-grade retention (Martin, 2011, p. 755).

Consequential Validity Framework

In this paper, I evaluate the validity and consequences of the Michigan Student Test of Educational Progress (M-STEP) used to make high-stakes decisions for our third graders. The high stakes of this assessment for Michigan students, families, and schools underscore the importance and urgency of examining the validity of the assessment. I will use a "consequential validity framework" that considers nine stages of assessment development and implementation, including construct (re)definition, assessment design, and analysis of both intended and unintended consequences to holistically evaluate the M-STEP and its intended consequences (Slomp, Corrigan, & Sugimoto, 2014).

Construct Definition

Because HB 4822 is rooted in reading proficiency, I begin by focusing my attention on the construct of reading presented in the Common Core State Standards (CCSS) that describes what it means to read well as it compares to the construct established in the legislation. While I acknowledge the CCSS definition is not without controversy (see for example Bomer et al., 2009; Gilyard, 2012; Williams et al., 2010), I consider the extent to which the knowledge and skills HB 4822 deems essential align with the Michigan Department of Education (MDE) adopted standards.

Wiggins and McTighe (2005) suggest analyzing such standards by unpacking the skills from the key ideas. Unpacking the CCSS, I find frequent repetition of verbs such as "describe" and "explain" in the reading standards which suggest students will necessarily verbalize or write responses to texts (Michigan Department of Education, 2010, p. 10). As I will demonstrate in more detail later, the M-STEP asks students to demonstrate reading skills not in writing or explanations, but through selected-response questions. Comparing the verbs from the unpacked standards with what the assessment asks students to do reveals a limited and misaligned

representation of the reading construct in the M-STEP.

Gaps between the key ideas of the CCSS standards, HB 4822 legislation, and the M-STEP assessment also exist. While the legislation stipulates "intensive development in the 5 major reading components: phonemic awareness, phonics, fluency, vocabulary and comprehension" (Price, 2015, p. 3), the CCSS do not include phonemic awareness and phonics standards in Grade 3. Neither are those major reading components assessed in the M-STEP. Though foundational reading skills support third-grade reading, the misalignment of the standards, bill, and assessment unnecessarily complicates the work for third-grade teachers. For which key concepts will we and our students be held accountable? I argue that this lack of agreement on the key skills and ideas problematizes the M-STEP construct of reading.

Design Process: Aims and Purposes

Narrowing my focus slightly, I also consider how the M-STEP is designed to measure the construct of reading as defined above for the purposes described in HB 4822 (Slomp et al., 2014). I begin with MDE's (2015b) "crosswalk" which summarizes "Claims and Targets [that] can be used to design classroom lessons and district assessments…[and] as a guide in understanding M-STEP reports" (p.), before shifting focus to available practice tests to develop a fuller picture of the M-STEP's validity for assessing the ELA reading construct.

The M-STEP crosswalk outlines four ELA claims: reading, writing, listening (not paired here with speaking as in the CCSS), and research (Michigan Department of Education, 2015b). Each of these correspond to "performance indicators" made in sample M-STEP parent reports. Because HB 4822 is intended to ensure reading proficiency, I concentrate on the reading portions of the M-STEP design. However, it should be noted that the law stipulates students will attain proficiency on the ELA portion of the M-STEP (Price, 2015). The legislation does not specify if the reading subscore or overall proficiency will be used to for retention decisions.

In the crosswalk, MDE (2015b) positions "read[ing] closely and analytically to comprehend a range of increasingly complex literary and informational texts" as the overarching reading claim (p. 2). This crosswalk supports the broad reading claim with specific goals that align to all CCSS reading literature and informational text standards. Notably missing, however, are any claims related to the reading foundational skills explicitly addressed in HB 4822. I argue that if the crosswalk is indeed intended to provide a foundation for classroom instruction as well as assessment as MDE claims, the omission of foundational reading skills is highly problematic in terms of design validity.

Practice Tests and Scoring Guides: What the M-STEP Assesses

Having problematized the validity of the reading claims related to the legislation's purpose, I use the M-STEP practice test to further explore the validity of the assessment design. I consider the design process essential for determining the extent to which the M-STEP assesses the reading construct defined above (Slomp et al., 2014). However, there are limits to this work as the available items may not fully represent the content of the M-STEP.

I must rely on the paper-and-pencil and computer-adaptive (CAT) practice tests published by MDE, though the CAT is now the official version. In a cautionary note introducing the paper-and-pencil practice test, MDE advises:

> The sample items included in this set can be used by students and teachers to become familiar with the kinds of items students will encounter on the paper/pencil summative assessment. The sample items demonstrate the rigor of Michigan's academic content standards. They are not intended to be interpreted as indicative of the focus of the M-STEP assessments; they are simply a collection of item samples. Every standard is not included in this sample set. (Michigan Department of Education, 2015a, p. ii)

The online practice test does not include a similar cautionary note, so it is not clear to what extent that version represents the full M-STEP ELA assessment.

To better evaluate the full content of the assessment, I submitted several FOIA requests for the 2014-2015 and 2015-2016 M-STEP Technical Reports. However, the FOIA requests were denied on the grounds the MDE does not yet have the documents. MDE has repeatedly postponed the expected availability of these reports, from "early 2017" (Head, 2016), "possibly March or April" (Head, 2017a), and most recently to "this summer" (Head, 2017b).

Thus, my investigation of the computer-adapted questions is limited by a small sample of six practice questions to cover the range of ELA domains: reading, writing, speaking and listening, and research. Missing from the test designers are "analyses [that] should make explicit those aspects of the target domain that the test represents, as well as those aspects the test fails to represent" (AERA, APA, & NCME, 2014, p.

196). Instead, students, parents, and schools receive M-STEP reports with poorly validated claims about the ELA domains.

Using the Grade 3 ELA scoring guide to supplement the paper-and-pencil practice test, I find the first four questions link to specific CCSS reading literature standards; of these, the first three show clear alignment. The fourth question, with its focus on dialogue, aligns with the "text features" aspect of the crosswalk (Michigan Department of Education, 2015b, p. 2), but does not relate to the CCSS "successive parts" defined as "chapter, scene, and stanza" (Michigan Department of Education, 2015b, p. 23). As a structural element, dialogue does not appear in the standards until fourth grade (Michigan Department of Education, 2015b, p. 23). Two crosswalk targets are not addressed in any practice test questions. However, I suggest the CAT reading questions show reasonable, though incomplete, alignment with the targets and standards outlined in the crosswalk.

I find the CAT replicates the first few questions included in the paper-and-pencil test. This version omits all informational text questions and it fails to address listening, research, and writing claims. The only visible CAT "improvement" over the pencil-and-paper version is the addition of a single reading question; students are asked to demonstrate "analysis within text" by dragging several events from the passage into chronological order (Data Recognition Corporation, 2016a). Behold the interactive, next generation improvements over the old-fashioned, fill-in-the bubble tests.

MDE (2016a) argues the CAT is "appropriate" because it "allows students to show what they know in an environment that feels comfortable for them, since many of today's students use computers daily in learning and in life" (p. 1). However, a closer look at the practice test reveals an application that is far from what students might typically experience with electronic media or classroom texts. With large icons at the top of the screen and a great deal of horizontal space dedicated to a question window (see figure 1), the reading passage on which a number of questions depends is confined to a narrow column of text through which students must scroll (Data Recognition Corporation, 2016a). Navigating the question itself is also difficult. The drag-and-drop interface in the sample interactive question may rely as much on manual dexterity as reading comprehension. To address these concerns, test designers have students begin practicing these skills as early as kindergarten with practice tests narrated by a robotic voice asking students to "please confirm your profile information is correct" and if it is not "to raise their hand to notify the assessment administrator" (Data Recognition Corporation, 2016b). The assumption that all third-grade students will have the technical skills to navigate the assessment, despite varying home and school access to computers, remains unacknowledged.

Missing Consequential Validity Elements: Stakeholders, Scoring and Sampling Plans

The critical issue here may not be that the practice CAT fails to replicate the full extent of the M-STEP, for "it is usually not possible to comprehensively measure all of the content standards using a single summative test" (AERA et al., 2014, p. 185). Assessment transparency, despite being explicitly required in the state legislation, is a fundamental missing element. The recently revised State School Aid Act of 1979 requires "the pool of questions for the summative assessments shall be subject to a transparent review process for quality, bias, and sensitive issues involving educator review and comment. The department shall post samples from tests or retired tests featuring questions from this pool for review by the public" (Michigan Legislature, 2016). These missing technical documents, contractually required to be produced annually, might provide more transparent insight into the assessment design (State of Michigan Department of Technology, Management and Budget Procurement, 2015). Failing to produce the technical reports that would include samples of additional question types for public review, Data Recognition Corporation (DRC) and MDE obfuscate the full content of the assessment and severely limit possibilities for public test review and critique.

The state's $41 million contract with DRC (State of Michigan Department of Technology, Management and Budget Procurement, 2015) reminds us that this situation "in which legislators and other educational decision makers are lavishly lobbied by testing industries (Simon, 2015) is too dangerous to be educationally sustainable" (qtd. in Broad, 2016, sec. 5.0). To reduce potential bias, all stakeholders impacted by the assessment, including but not limited to students, their families, and teachers, should be considered in the design and revision of the M-STEP (Slomp et al., 2014). The several concerns related to the M-STEP in terms of scoring and sampling plan validity that should have been evaluated before the passage of HB 4822.

Disaggregated Performance Data

Given the absence of this key technical information, I mistrust the validity of the M-STEP as it concerns student diversity and fairness. Statewide, 54% of third-grade students

earned deficient scores on the M-STEP ELA assessment in the 2015-2016 academic year (Michigan Department of Education, 2016b). As this M-STEP data is disaggregated, troubling trends become apparent. Some subgroups are much more likely than the general population to earn a deficient score (see figure 2). Economic status it is clear, significantly correlates with students' proficiency as those who come from economically disadvantaged families are almost twice as likely as their more advantaged peers to score deficient. Race too, is a significant factor. While white and Asian students have a lower risk of scoring deficient than the whole population, the risk climbs to 80% for black students. That is a higher rate than students with disabilities. The disaggregated data for third-grade Michigan students shows markedly different scores that might be interrogated in the absent sampling plan.

Without a sampling plan, it is unclear if the extreme differences in the disaggregated data should be attributed to either actual achievement or other student factors (Slomp et al., 2014). "Think aloud protocols [and] focus groups" are two methods for understanding how students in different subgroups approach the assessment and to what extent the construct measured in the M-STEP aligns with and validates the intended use of the scores (Slomp et al., 2014, p. 284). While there is no evidence this work has been done, this is an opportunity to improve the assessment transparency and validity in future iterations of the assessment.

Without access to the pool of questions, the underlying test blueprint, or the technical reports, no individual or organization outside MDE and DRC can make informed critiques of the scoring and sampling plans. Even the legislature, which as a user of the M-STEP "is ultimately responsible for evaluating the evidence in the particular setting in which the test is to be used" (AERA et al., 2014, p. 13), cannot ensure the validity of the assessment design. Thus, the M-STEP's lack of state-mandated transparency fails to address another three essential aspects of consequentially valid assessments: stakeholder involvement, scoring, and sampling plan (Slomp et al., 2014).

Intended Consequences

In testimony on the proposed bill, the Michigan State Board of Education opposed the retention policy of HB 4822 (Whiston, 2015). Acknowledging "there are situations where retaining students in their current grade is warranted," the State Board of Education argued "the decision needs to be decided on [a] case-by-case basis, between educators and parents" (Whiston, 2015, p. 4). Additionally, House Representative Kristy Pagan, who has a degree in Education and education policy experience in the US Senate, moved to amend HB 4822 to consider stakeholders' opinions in the retention of third-grade students with the addition to the list of exemptions "the pupil's parents, teacher, and school administrator unanimously agree in writing the potential benefits of retention are outweighed by the potential negative consequences of retention" (Randall, 2015, sec. 3). The proposed amendment failed.

There are significant risks to retention that those with close knowledge of students must weigh. While retained students' scores may show short-term increases, the long-term effect on scores "fade[s] over time" (Huddleston, 2014, p. 19). Research suggests "retention has negative implications for academic motivation, academic achievement, academic self-concept, and general self-esteem" (Martin, 2011, p. 753). Retention also correlates with a higher rate of drop out; studies of retained students find that retention in a single grade increases a student's dropout risk by 18 to 50 percent (Holmes, 2006, pp. 57–58).

Proponents of retention argue "that a 'social promotion' may protect a child's self-esteem, but it does so by advancing a lie" (Randall, 2015, sec. 3). For the sake of the "truth" of a 90-minute multiple choice assessment, we risk long-term detrimental consequences for students which persist "even after controlling for ability" (Martin, 2011, p. 754). If the retention policy in HB 4822 is implemented in the 2019-2020 academic year and current M-STEP proficiency rates do not change, more than half of Michigan third-grade students may be retained based on a single assessment. In comparison, only 799 students, .72% of Michigan third graders, were retained in the 2015-2016 academic year (Michigan Department of Education, 2016c). With current rates so low and the risks of retention for students so high, why is the Michigan legislature targeting that grade level with such a high-stakes policy?

For those subpopulations at risk in the disaggregated data, the long-term consequences of retention may be severe (Huddleston, 2014). AERA standards affirm any valid assessment would need to show "that the test is useful in determining which persons are likely to profit differentially from one treatment group or another" (AERA et al., 2014, p. 30). The unstudied causes of disparate Michigan subgroup scores further threaten to make retention unfair for our least advantaged students while parent, teacher, and school administrator input is denied.

School and District Impact

Perhaps even more alarming than the intended consequences are the unintended consequences of the third-grade retention policy for students, schools, and communities. It is true that HB 4822 includes several "good cause exemptions" including IEP or 504 plans, limited English proficiency, and previous retention, and allows for alternative demonstrations of proficiency through another approved standardized test or a portfolio (Price, 2016, p. 6). The House Fiscal Analysis of HB 4822 claims though these "exemptions…would limit the number of actual students repeating the 3rd grade" and mitigate increased costs to the state, it is likely that the law will have significant consequences for some districts and schools (House Fiscal Agency, 2016, p. 12).

Completing portfolios and alternative standardized tests for up to 80% of students in some schools could place a significant additional burden on the teachers (Michigan Department of Education, 2016b). For students who might not perform well on the M-STEP, the portfolio requirement that they "demonstrat[e] competency in all grade 3 English language arts standards through multiple work samples" would be enormous (Price, 2015, sec. 5.a.iii). Teachers would have to document student achievement in 46 standards, many of which are not assessed on the M-STEP. When multiplied by the number of students expected to not earn proficiency (scores are not released until June each year), teachers who serve students in low-performing areas would face significantly increased responsibilities without corresponding increases in compensation or planning time.

In addition to this increased burden on education agencies, schools, and teachers, many unintended negative consequences of high-stakes testing may arise. These consequences include test-preparation focused curricula, cheating, emotional stress, score inflation (where standards are lowered to show artificial growth), and "teachers focused on near-passing students" rather than those with the greatest needs (Huddleston, 2014, p. 7). I worry these changes may lower the quality of reading instruction and result in higher rates of dropout that will have significant and long-lasting effects on our communities.

A Slower Race to the Top

I urge the MDE and DRC to be more prudent in revising the assessment in pursuit of greater validity. In response to stakeholder feedback calling for a shorter assessment after the initial M-STEP (Michigan Department of Education, 2016d, p. 12), designers omitted all third-grade test questions requiring written responses. This change does not improve the consequential validity of the assessment and provides an insufficient response to teachers, schools, and families seeking less disruption to the learning process. Slowing the assessment down may mean looking toward curriculum-based learning portfolios in place of the 90-minute assessment for all students. These, however, should not be required to address all 46 ELA standards, but rather a core group identified as essential by each community. I suggest rooting the assessment in the local context and instruction would also provide more meaningful formative assessment actionable by both teachers and families to improve student learning.

I do not argue against improving the literacy skills of students, but the significant issues of validity and consequential validity at all levels of my analysis call for a more cautious approach. Punishing young students for the larger structures of racial, economic, and educational inequalities based on an invalid assessment is unethical. In our "race to the top" (U.S. Department of Education, 2016), we may need to slow down so that all students can succeed. It may be necessary to extend instruction for students who do not master foundational literacy skills by third grade, but this does not mean that their progress through the grades must be impeded. Rather, we should adjust our instruction as the exemptions allowed in the legislation admit, just as we would for second language students and students with documented disabilities. Students can continue to develop competency in those basic literacy skills while they also receive high-quality, grade-level instruction with embedded supports.

When the State of Michigan argues that literacy is not a right despite grossly unequal opportunities (Chambers, 2016), third-grade "students should not be held accountable for, or face serious permanent negative consequences from their test results when their school experiences have not provided them the opportunity to learn the subject matter covered by the test" (AERA et al., 2014, p. 56). Until the state has developed a valid assessment, with genuine opportunities for critique and revision, it is unfair and unethical to use the M-STEP as the default tool for high-stakes decisions while denying parents and teachers the authority to make the most appropriate decisions for students.

References

AERA, APA, & NCME. (2014). *Standards for educational and psychological testing.* Washington, D.C: American Educational Research Association.

Bomer, R., Bass, B., Cruz, M., Hesse, D., Kiernan, H., Ochoa, J., … Williamson, K. (2009). *A report of*

the NCTE review team on the July 2009 draft of the common core English language arts state standards. Urbana, IL: NCTE. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.569.3135&rep=rep1&type=pdf

Broad, B. (2016). This is not only a test: Exploring structured ethical blindness in the testing industry. *The Journal of Writing Assessment,* 9(1). Retrieved from http://journalofwritingassessment.org/article.php?article=93

Chambers, J. (2016, November 21). State says literacy not a right in Detroit. *The Detroit News.* Retrieved from http://www.detroitnews.com/story/news/education/2016/11/20/state-says-literacy-right-detroit/94193032/

Data Recognition Corporation. (2016a). ELA grade 3 sample items. Retrieved October 30, 2016, from https://wbte.drcedirect.com/MI/#items/2

Data Recognition Corporation. (2016b). ELA kindergarten sample items. Retrieved October 30, 2016, from https://wbte.drcedirect.com/MI/#items/2

Gilyard, K. (2012, February 9). NCTE president Keith Gilyard talks about NCTE and Common Core Standards. Retrieved April 16, 2017, from http://www.ncte.org/standards/common-core/kg_2-9-12

Head, D. (2016, November 17). Michigan Department of Education FOIA Request.

Head, D. (2017a, January 31). Michigan Department of Education FOIA Request.

Head, D. (2017b, March 29). Michigan Department of Education FOIA Request.

Holmes, C. T. (2006). Low test scores + high retention rates = more dropouts. Kappa Delta Pi Record, 42(2), 56–58.

House Fiscal Agency. (2016). *Smart promotion: 3rd grade reading proficiency, early literacy coaches, and reading assistance programs* (Legislative Analysis) (pp. 1–16). Retrieved from http://www.legislature.mi.gov/documents/2015-2016/billanalysis/House/pdf/2015-HLA-4822-22068DE3.pdf

Huddleston, A. P. (2014). Achievement at whose expense? A literature review of test-based grade retention policies in US schools. *Education Policy Analysis Archives*, 22(18), 1–31.

Lloyd, D. N. (1978). Prediction of school failure from third-grade data. *Educational and Psychological Measurement*, 38(4), 1193–1200. https://doi.org/10.1177/001316447803800442

Martin, A. J. (2011). Holding back and holding behind: Grade retention and students' non-academic and academic outcomes. *British Educational Research Journal*, 37(5), 739–763. https://doi.org/10.1080/01411926.2010.490874

Michigan Department of Education. (2010, June 2). Michigan K-12 standards: English language arts. Michigan Department of Education. Retrieved from https://www.michigan.gov/documents/mde/K-12_MI_ELA_StandardsREV_470029_7.pdf

Michigan Department of Education. (2015a). M-Step English language arts grade 3 sample items. Michigan Department of Education.

Michigan Department of Education. (2015b, September 30). English Language Arts grade 3 crosswalk: Claims-targets-standards. Michigan Department of Education. Retrieved from http://www.michigan.gov/documents/mde/G3_crosswalk_claim_target_standard_500252_7.pdf

Michigan Department of Education. (2016a). Computer adaptive testing (CAT). Michigan Department of Education. Retrieved from https://www.michigan.gov/documents/mde/Computer_Adaptive_Testing-fact_Final_520743_7.pdf

Michigan Department of Education. (2016b). Spring 2015 M-STEP district and school aggregate data file. Michigan Department of Education. Retrieved from http://www.michigan.gov/mde/0,4615,7-140-22709_70117_31530---,00.html

Michigan Department of Education. (2016c). Student count: Retained in grade snapshot. Retrieved October 24, 2016, from https://www.mischooldata.org/DistrictSchoolProfiles/StudentInformation/StudentCounts/NewRetention.aspx

Michigan Department of Education. (2016d, March 4). Spring 2016 M-STEP test administration manual. Michigan Department of Education. Retrieved from http://www.michigan.gov/documents/mde/S16_M-STEP_TAM_3-4-16_516206_7.pdf

Michigan Legislature. The State School Aid Act of 1979, Pub. L. No. 94, § 388.1704c, 94–1979–I (2016). Retrieved from http://www.legislature.mi.gov/(S(4leg5jl04e4emloj125x4k11))/mileg.aspx?page=getObject&objectName=mcl-388-1704c

Price, A. Enrolled House bill no. 4822, Pub. L. No. 4822, § 1280f, 1 (2015). Retrieved from http://www.legislature.mi.gov/(S(zgtwb0a3xumvshpfg5maqw0y))/mileg.aspx?page=getobject&objectname=2015-HB-4822

Randall, G. L. (2015). *Journal of the House of Representatives* (No. 84). Lansing, MI: State of

Michigan. Retrieved from http://www.legislature.mi.gov/(S(ucknfdhvpv4dmgzidzj0febf))/documents/2015-2016/Journal/House/htm/2015-HJ-10-15-084.htm

Slomp, D. H., Corrigan, J. A., & Sugimoto, T. (2014). A framework for using consequential validity evidence in evaluating large-scale writing assessments: A Canadian study. *Research in the Teaching of English*, 48(3), 276–302.

State of Michigan Department of Technology, Management and Budget Procurement. (2015, October 8). Notice of contract no. 071B5500120 between the state of Michigan and Data Recognition Corporation. Retrieved from http://www.michigan.gov/documents/micontractconnect/5500120_504305_7.pdf

U.S. Department of Education. (2016, July 19). Race to the top fund. Retrieved December 13, 2016, from http://www2.ed.gov/programs/racetothetop/index.html

Whiston, B. J. (2015, October 13). State board of education statement on house bill 4822. Retrieved from http://www.michigan.gov/documents/mde/SBE_Statement_on_HB_4822_502933_7.pdf??20160524143009

Wiggins, G., & McTighe, J. (2005). *Understanding by design* (Expanded 2nd Edition). Alexandria, VA: ASCD.

Williams, J., McCraw, B., Harris, K., McCullar, D., Marquez, L., Younker, K., … Director, N. E. (2010). *A report of the national council of teachers of English review team on the February 2010 draft of the standards for English language arts: grades k-12.* National Council of Teachers of English. Retrieved from http://www.ncte.org/library/NCTEFiles/Resources/Standards/NCTE_Report_CoreStds_1-10

Michelle Sprouse began her career as a middle and high school reading teacher. She is currently pursuing her PhD in in English and Education at the University of Michigan.

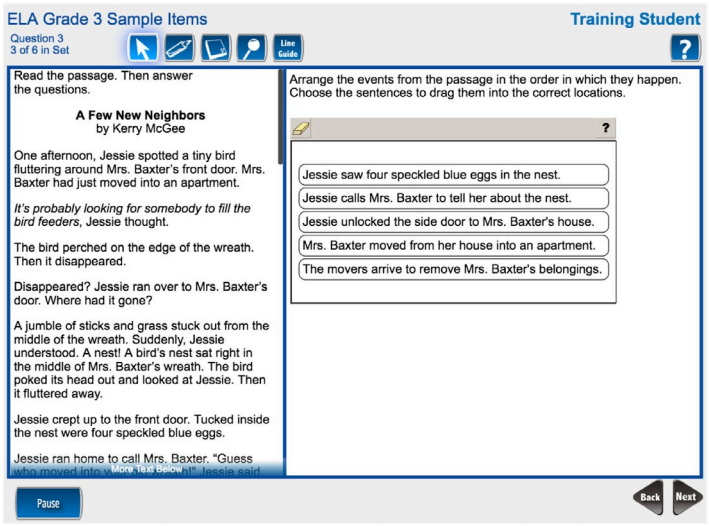

Figure 1. Sample reading question from the third-grade computer-adaptive practice test (Data Recognition Corporation, 2016a).

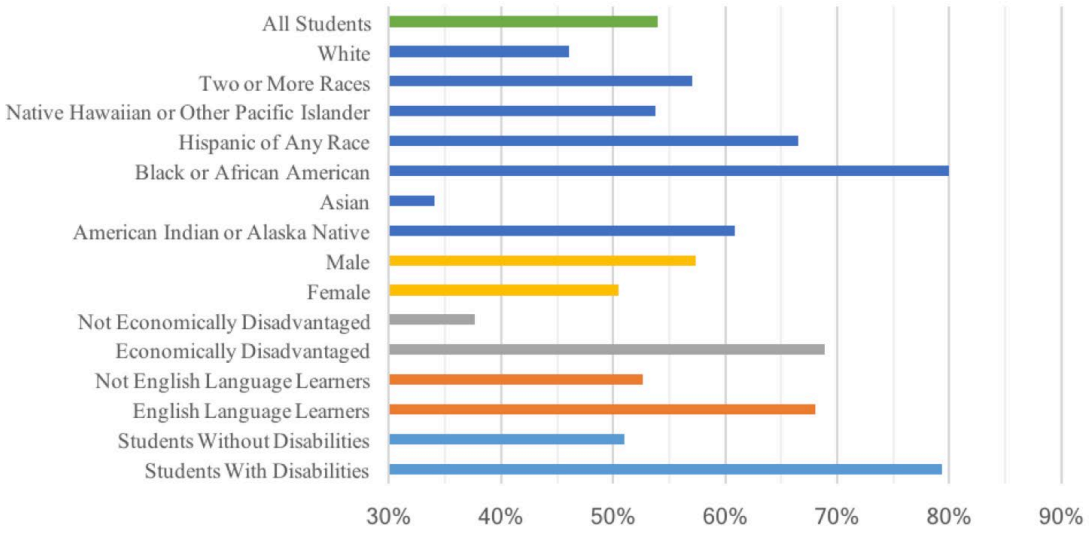

Figure 2. 2015-2016 third-grade M-STEP percent of student scoring not proficient (Michigan Department of Education, 2016b). These students, except for English Language Learners with less than three years of English instruction and students with disabilities would be subject to retention under HB 4822.