WALDEN UNIVERSITY
A higher degree. A higher purpose.

Effect of Vocabulary Test Preparation on Low-Income Black Middle School Students' Reading Scores

Ingrid Mitchell
Walden University

Nicolae Nistor
Walden University

Beate Baltes
Walden University

Michelle Brown
Walden University

Black middle school students in the United States continue to perform poorly on standardized reading achievement tests in comparison to other racial and ethnic groups. The purpose of this research study was to examine the effectiveness of a vocabulary-focused test preparation program for Black middle school students. The theoretical framework consisted of Thorndike's concept of test-wiseness, a test-taking capacity. Teachers at the research site were trained on Larry Bell's 12 Powerful Words strategy that aims to make students test-wise, that is, to familiarize them with key vocabulary terms related to tests. An intact-group comparison was conducted, involving a total of N = 679 Black students in Grades 6, 7, and 8 with 370 girls and 309 boys. An analysis of covariance showed significant effects for Grade 6, marginally significant effects for Grade 7, and nonsignificant effects for Grade 8. These findings suggest that the 12 Powerful Words are effective and that their effect decreases with students' age. As a practical consequence, instructional leaders will be able to make more informed decisions regarding test preparation and potentially reduce the number of underperforming students in classrooms.

Keywords: *test preparation, low-income students, vocabulary, reading scores*

Problem Statement

At both state and national levels, Black students in the United States perform poorly on measures of reading achievement. According to the Tennessee State Department of Education (n.d.), 20% of Grades 3–8 students scored below basic, 48% scored basic, 25% scored proficient, and 5% scored advanced on the Tennessee Comprehensive Achievement Test (TCAT) in the 2011–2012 school year. Of 241 Black males who completed the reading test, 205 (85%) scored below basic or basic, 30 (13%) scored proficient, and six (2%) scored advanced. Black females performed slightly better with 62 (49%) scoring below basic or basic, 35 (27%) scoring proficient, and 30 (24%) scoring advanced.

According to the National Assessment of Educational Progress, Black students in U.S. fourth grade achieved a scale score of 205 out of a possible 500 in their state exams in 2011, whereas White

students averaged a scale score of 231. In 2013, Black fourth-grade students averaged a scale score of 206, whereas White students averaged a scale score of 232 (Nation's Report Card, 2013).

The trend continues into the secondary level with Black eighth-grade students scoring an average of 249 on the National Assessment of Educational Progress in 2011 and White students scoring 274. In 2013, Black eighth-grade students averaged a scale score of 250 out of a possible 500 whereas White students averaged a scale score of 276 (Nation's Report Card, 2013).

To help improve the performance of Black students on standardized reading tests, teachers at one Tennessee middle school have implemented a test preparation instructional program designed to familiarize students with formal tests that they are required to take annually. The test preparation program is a daily integration of 12 Powerful Words into core academic content (Bell, 2005). As there was no evaluation data available to attest the effectiveness of this method, this study investigated whether the instructional program based on 12 Powerful Words impacts the reading achievement scores of Black middle school students on their standardized test in Tennessee.

Theoretical Framework

The influence of test-wiseness has been discussed in the literature for over 50 years. Test-wiseness was described as the "capacity to utilize the characteristics and formats of the tests and/or the test-taking situation to receive a high score" (Millman, Bishop, & Ebel, 1965, p. 87). Originally, Thorndike (1951) identified test-wiseness as one of the grounds for variation observed in test score differences. Millman et al. (1965) suggested that test-wiseness is autonomous to the content area that is being tested. Similarly, Oakland and Weilart (1972) defined test-wiseness as a skill set that promoted test proficiency. Finally, Scruggs and Lifson (1985) proposed to apply the concept to Black students as a way to close the racial gap with reading proficiency.

Literature Review

Researchers have suggested that test preparation may improve students' test performance. Alon (2010) conducted a longitudinal research study tracking and comparing racial minority students to White students who were coached separately on SAT test strategies through private tutors and classes. The test preparation included vocabulary and test formatting practice. Alon stated that test preparation does boost test proficiency on content area formative assessments. These results are supported by the findings of Ebanks, Toldson, Richards, and Lemmons (2011) on the outcomes of a program titled Project 2011, which was a focused learning curriculum tailored for Black and Latino students for earned admittance into New York City's exclusive Specialized High School programs. As participants in this program, students were offered test preparation through vocabulary and test motivation. Results showed that students dramatically improved in their test performance. Because of Project 2011, students were able to enroll and participate in the city's most elite educational programs and early college entry (Ebanks et al., 2011). Winsler, Karkhanis, Kim, and Levitt (2013) added that it is imperative that students are immersed in culturally diverse learning communities and that the goal of teachers providing test preparation should be that proficiency gaps in achievement are closed.

Misco (2010) implemented a test preparation exit exam program, specifically for social studies. His research concentrated on ten ethically mixed students who were high school juniors and seniors. These students had previously failed an exit exam and now teachers remediated the course curriculum by tailoring student learning to reasoning, test-taking skills, and verification of answers with text. All 10 students passed the exit exam after completing the test preparation course and

Misco suggested that test preparation should be part of the general curriculum and could be applied to subgroups such as English language learners.

Overall, research results (Ebanks et al., 2011) have indicated that test preparation can improve performance of traditional, exceptional, and English language learner students when a targeted and explicit approach is followed (Alon, 2010; Misco, 2010). Beyond direct studies of test performance, the literature on test preparation has also expanded to include three emerging lines of research: test-taking strategies, the role of technology, and tests-as-genre studies.

Test-Taking Strategies

Stough (1993) improved test scores by teaching 90 college freshmen to avoid items using the words *always* and *never*, to avoid unrelated alternatives, and to pay attention to grammatical clues. Since then, many different test-taking strategies have been taught in a variety of course, often with positive results. Bicak (2013), for example, developed a test preparation course that helped students from various public high schools who wanted to pass college/university entrance exams. Matsumura, Garnier, and Resnick (2010) examined the effect of a test preparation program named Content-Focused Coaching, in which teachers were trained to coach students for tests. The participants were 96 randomly selected fourth- and fifth-grade teachers and 63 students from 15 elementary schools that were all assessed as proficient from the fall of 2006 to end of study in the spring of 2007 after 1 year. Matsumura et al. described the teachers' perceptions of the test preparation training to benefit their students (Lam, 2013; Matsumura et al., 2010). To determine how test preparation may improve performance by modifying the students' strategic approach to the test, Yang (2000) compared test-taking strategies and exam performance with experienced and inexperienced test-takers. Yang probed students to employ test-wiseness in responding to a Test of English as Foreign Language Practice Test B. Twenty-three out of 390 Chinese Test of English as Foreign Language students were identified as test-wise, and 17 were identified as inexperienced test-takers (Yang, 2000). The 23 students who were considered to be experienced in test-wiseness performed better on test items than the 17 inexperienced students (Yang, 2000). Qualitative interviews showed that experienced test-takers were not only test-wise in making meaning from questions but also were more analytical, logical, and systematic in answering test questions than inexperienced students (Yang, 2000). In a similar study, Zhengdong (2009) conducted a study of 146 students who took an International English Language Testing System test preparation course. Participants received 10–20 hr of test preparation (Zhengdong, 2009), and results showed that all students who took the test preparation course scored higher than the students who did not take any test preparation course.

The Role of Technology

A closely related line of research has investigated the role of technology in test preparation. Hughes, Schumaker, Deshler, and Mercer (1988) and Hughes and Schumaker (1991) developed a test-taking strategy for secondary students with disabilities. Lancaster, Schumaker, Lancaster, and Deshler (2009) then adapted the test-taking strategy to a computerized program to teach test preparation to secondary students with disabilities. Classes were isolated to determine the effects of the program which resulted in significant increases in test performance for the students in the computerized program. The use of test-taking strategies, including familiarization, vocabulary, and ability to think or engage aloud with tests, helped students in being test-wise. Similarly, Fede, Pierce, Matthews, and Wells(2013) investigated the effects of a computer-assisted test preparation instructional program that focused on word problem solving skills for fifth-grade students that perform below proficiency. A total of 32 students participated in a randomized controlled study. Sixteen students were given a 12-week computer-assisted test preparation instructional program. The other 16

students received a 12-week general state test preparation. All students completed the Massachusetts Comprehensive Assessment Test. Results showed that the computer-assisted test preparation instructional program group showed greater gains than the control group; however, students in both groups improved their test scores.

Test-as-Genre Studies

Other researchers have focused on improving test scores by introducing students to the concept of tests as a genre. Assessments could be considered a core class teaching test-wiseness. Hornof (2008) proposed that only studying test preparation materials can help students be better test-takers, and therefore, he developed a 2-week unit for fourth-grade students to prepare for standardized reading tests. He taught, for example, how the elimination of incorrect answers can eventually lead to the correct answer. In addition, Hornof emphasized stamina and, therefore, made practice assessments longer over time. The results of his study showed that test-takers scored better on each progressive test they took.

Welsh et al. (2014) corroborated the results in their investigation of the relationship between student test performance and test preparation activities in 32 third- and fifth-grade classrooms. After using test preparation items similar to the state test, test format practice, and teaching test-taking skills, Welsh et al. found that student achievement benefits from general instruction on the state standards rather than from isolated test preparation.

There are supporters and opponents of test preparation and especially courses whose sole focus is to teach test-taking skills. Nevertheless, these courses do exist. Kontovourki and Campis (2010) studied such a test-taking-skills course for urban students. In another example of explicit test preparation, Doe and Fox (2011) prearranged and implemented a genre test course for nonnative English speakers. Similarly, Gadbury-Amyot, Austin, and Overman (2013) implemented an adult distance-learning program for explicit test preparation, and Lancaster et al. (2009) developed a well-organized computer-based test preparation program for students with disabilities.

Research Question and Hypotheses

Although research is moving beyond direct studies of test performance, conclusions regarding test-taking strategies, the role of technology, and tests-as-genre are still tentative. Available empirical research findings are insufficient to demonstrate the effects of test-wiseness on underachieving Black middle school students. Therefore, this study investigated the following research question:

> *Research Question:* What is the difference in reading test scores of underachieving Black middle school students who participated in a key vocabulary test preparation program and those who did not?

> *Hypothesis 0:* There is no significant difference in students' mean scores on a standardized reading assessment between Black middle school students who participated in a key vocabulary test preparation program and those who did not.

> *Hypothesis 1:* There is a significant difference in students' mean scores on a standardized reading assessment between Black middle school students who participated in a key vocabulary test preparation program and those who did not.

Methodology

Research Design and Approach

An intact-group comparison design was used in which the test scores of students receiving the test preparation treatment in School A were compared to test scores from students in an equivalent School B who did not receive the specific treatment. The majority of the student population in both schools was classified as nonproficient in reading. According to Coalition for Evidence-Based Policy (2014), group comparison design can produce valid results if pre-program review match in demographics, measures to seek to improve an implemented program, groups' geographic location is in the same vicinity, data is collected in the same way, similar motivation in participation from both groups, and statistical methods are used to adjust for minor variances between groups.

Setting and Sample

In order to assure equivalence of intact groups, the two schools (School A and School B) were matched for location, neighborhood, ethnicity, socioeconomic status, student achievement, and teacher experience. The state report card of academic achievement showed that in School A, 35% of students scored below basic in 2010, 33% in 2011, and 32% in 2012 (Tennessee State Department of Education, n.d). In School B, 30% of students scored below basic in 2010, 27% in 2011, and 25% in 2012 (Tennessee State Department of Education, n.d.). In other words, both schools have a trend of poor performance on the reading assessment.

The sample comprised 679 Black students in Grades 6, 7, and 8 (370 girls and 309 boys), all receiving free or reduced price lunch (Table 1 and 2). This sample provides greater power than the minimum sample size of 210. All students in the sample received free or reduced priced lunch. Of the selected study participants, 20% were students with disabilities. Detailed demographic data for research participants showing gender, race and ethnicity, English language learners, low socioeconomic status because all students in the sample received free or reduced priced lunch, and students with disabilities are provided in Table 1 for School A and Table 2 for School B.

Table 1. *School A: Demographic Data for Test Preparation Participants*

Demographic Variable	Grade 6		Grade 7		Grade 8	
	n	%	*n*	%	*n*	%
Gender						
Male	86	57	79	59	71	47
Female	64	42	53	40	77	52
Race/Ethnicity						
Black	142	94	122	.92	141	95
Hispanic	8	05	10	.07	7	04
Other	0	0	0	0	0	0
English language learners						
Yes	0	0	1	.007	2	01
No	150	100	131	.99	146	98
Low socioeconomic status						
Yes	150	100	132	100	148	100
No	0	0	0	0	0	0
Disability						
Yes	23	15	19	14	23	15
No	127	84	113	85	125	84
Total	150	100	132	100	148	100

Table 2. *School B: Demographic Data for No Test Preparation Participants*

Demographic	Grade 6		Grade 7		Grade 8	
Variable	*n*	%	*n*	%	*n*	%
Gender						
Male	39	51	49	56	46	52
Female	37	48	37	43	41	47
Race/Ethnicity						
Black	75	98	85	98	87	100
Hispanic	0	0	0	0	0	0
Other	1	01	1	01	0	0
English language learners						
Yes	0	0	0	0	0	0
No	76	100	86	100	87	100
Low socioeconomic status						
Yes	76	100	86	100	87	100
No	0	0	0	0	0	0
Disability						
Yes	14	18	10	11	15	17
No	62	81	76	88	72	82
Total	76	100	86	100	87	100

Teachers instructed both groups using methodologies already approved by the principals, who had autonomy for curricular decisions, with the exception of mandated Title 1 supplemental academic programs, see page 10 (NCLB, 2002; Tennessee State Department of Education, n.d.). As per Title I requirements, absenteeism cannot fall below 93% of daily instructional hours, so students who did not meet the attendance benchmark were reported as truant students to district personnel (NCLB, 2002). Therefore, absenteeism did not affect the data analysis.

Treatment

Students in Grades 6, 7, and 8 at both School A and School B received the same district-mandated reading curriculum throughout the school year and the same district-mandated supplementary reading remediation as part of the Title I program. In addition, students at School A received an auxiliary instruction for test preparation, called the 12 Powerful Words (see Table 3) test preparation program (Bell, 2005). In addition, School A students were instructed each morning for 15 min via the school's intercom on spelling and reciting particular words. To assure fidelity of implementation of the Powerful Words program, each classroom was observed daily as a part of monitoring for district quality checks for instruction. In addition, the instructional coach and administrative team checked the lesson plans from all 17 core content teachers to ensure the 12 Powerful Words and test preparation strategy were included.

Table 3. *12 Powerful Words and Larry Bell's Strategies (Bell, n.d.)*

1	Explain	Make plain or clear; render understandable or intelligible
2	Summarize	Give the short version
3	Compare	Consider or describe as similar; liken
4	Trace	Follow a course, trail, etc.; list in steps
5	Formulate	Devise or develop, as a method, system, etc.; create
6	Predict	Declare or tell in advance; prophesy; foretell; what will happen next
7	Analyze	Examine carefully and in detail so as to identify causes, key factors, possible results, etc.; break apart
8	Infer	Derive by reasoning; come to a conclusion based on evidence; read between the lines
9	Support	Back up with details
10	Contrast	Show unlikeness or differences; note the opposite natures, purposes, etc.
11	Describe	Tell or depict in written or spoken words all about something; give an account of
12	Evaluate	Judge or determine the significance, worth, or quality of; assess

Instrumentation and Materials

The instrument to measure the dependent variable was the reading subtest of the Discovery Education Assessment (DEA), a timed multiple-choice test that measures skills in reading, mathematics, and science (Discovery Education, n.d.). The DEA is administered to all students in the school district three times during the year as both a tool for instruction and as a predictive test for the Tennessee Comprehensive Achievement Test. All schools must adhere to district's testing time and schedule. The reading subtest is administered in two sections of 75 min each. The two sections contain test items categorized into seven areas: language, vocabulary, writing and research, communication and media, logic, informational text, and literature. The number of reading passages presented may have changed from year to year, but in general, there are four to five long passages per section. Students respond to test items in a multiple-choice format (Discovery Education, n.d.). The test is timed and students are credited for the number of correct answers with no penalty for guessing. Each student receives a raw score for the seven reading categories and a reading composite raw score.

Instrument Validity and Reliability

To ensure content validity, the publisher aligns test content to the state's content standards and the content sampled by the state assessment test (Discovery Education, n.d.). The publisher also uses the Webb Alignment Tool to support alignment with state-specific content standards in both the number of standards objectives sampled; and the cognitive complexity of standards and objectives (Discovery Education, n.d.). Consequently, the DEA in reading used for Tennessee matches the eight reporting categories of the reading test of the Tennessee Comprehensive Assessment Program (TCAP): language, communications, writing, research, logic, information text, media, and literature.

Although the DEA is designed as a formative tool to guide classroom instruction, it has been shown to predict scores on the reading test of the TCAP (Discovery Education, n.d.). For Grade 6, the correlations between three administrations of the DEA and the spring TCAP reading test were .73 (Form A), .72 (Form B), and .69 (Form C). For Grade 7, the correlations between three administrations of the DEA and the spring TCAP reading test were .74 (Form A), .71 (Form B), and .71 (Form C). For Grade 8, the correlations between three administrations of the DEA and the spring TCAP reading test were .71 (Form A), .72 (Form B), and .71 (Form C).

A reliability study of the DEA for Grade 6 Reading was conducted in 2009 with a sample size of 29,253 students. The Cronbach's alpha was .81 (Discovery Education, n.d.). For Grade 7, the sample size was 29,182 students. The Cronbach's alpha reliability coefficient was .76, and for Grade 8, the sample size was 29,043 students with a Cronbach's alpha reliability coefficient of .84. Discovery Education reported no data on test-retest or parallel-forms reliability.

Data Collection and Analysis Procedure

Students at both schools took the district required DEA Reading Test at the beginning and middle of the school year. The test results were available and released as archival data. According to the research design presented above, the dependent variable was the midyear DEA reading test score and the covariate was the test score on the same assessment at the beginning of the school year.

All students at School A and School B sites were administered the DEA by their certificated homeroom teacher, in accordance with procedures directed by the Tennessee DOE. The initial test (pretest) was given in October 2012, and the mid-year assessment (posttest) was given in December 2012. Student answer sheets were forwarded to DEA for scoring and reporting of results. Besides descriptive statistics, the null hypothesis was tested by analysis of covariance (Keselman et al., 1998). Data were analyzed using IBM SPSS Statistics Version 24.

Results

Descriptive Statistics

Grade 6 students at School A (the test preparation school) increased their initial pretest mean score of 26.41 (SD = 22.43) to a posttest mean score of 33.79 (SD = 23.43) which is an increase of 7.38. In School B, however, the mean score decreased by 1.12 points from 23.59 (SD = 19.50) to 22.47 (SD = 18.06).

Grade 7 students at School A showed an increase of 1.20 from a pretest mean score of 31.05 (SD = 25.35) to a posttest mean score of 32.25 (SD = 22.37). In School B, the mean score only varied by 0.16 from a pretest 32.77 (SD = 23.97) to a posttest of 32.61 (SD = 20.77).

Grade 8 students at School A had a decrease in their test scores of 8.04 from a pretest mean score of 38.91 (SD = 25.69) to a posttest mean score of 31.87 (SD = 24.27). The results at the nontreatment School B were more moderate as the means scores decreased only by 1.00 from 34.73 (SD = 22.48) to a posttest mean score of 33.73 (SD = 24.50). All pre- and posttest scores are provided in Table 4.

Table 4. *Pretest and Posttest Mean Scores With Standard Deviations for Treatment and Alternative Treatment Groups Plus Results From Analyses of Covariance of Treatment Groups for Grades 6, 7, and 8*

	N	Pretest (T0) M	Pretest (T0) SD	Posttest (T1) M	Posttest (T1) SD	df	F	p
		Treatment						
Grade 6								
Female	64	27.40	19.91	32.09	22.21			
Male	86	29.08	23.49	35.06	24.35			
Students with disabilities	23	18.96	16.92	27.13	24.17			
Overall pretest	10			26.41	22.43			
Overall posttest	150			33.79	23.43	1, 217	5.687	.018
Grade 7								
Female	53	36.29	26.42	36.81	23.21			
Male	79	27.35	23.93	29.20	21.39			
Students with disabilities	19	15.70	14.32	28.78	15.95			
Overall pretest	132			31.05	25.35			
Overall posttest	132			32.25	22.37	1, 209	3.529	.062
Grade 8								
Female	77	44.16	25.85	34.22	22.76			
Male	71	32.95	24.29	29.32	25.72			
Students with disabilities	23	20.33	13.65	10.69	10.42			
Overall pretest	148			39.91	25.69			
Overall posttest	148			31.87	24.27	1, 226	1.352	.246
		Nontreatment						
Grade 6								
Female	37	24.87	19.54	23.08	18.55			
Male	39	21.85	19.57	21.90	17.80			
Students with disabilities	14	13.18	8.99	15.00	12.50			
Overall pretest	76			23.59	19.50			
Overall posttest	76			22.47	18.06			
Grade 7								
Female	37	35.10	22.40	34.64	19.10			
Male	49	31.77	25.65	31.08	22.02			
Students with disabilities	10	18.18	18.29	10.80	8.67			
Overall pretest	86			32.77	23.97			
Overall posttest	86			32.61	20.77			
Grade 8								
Female	42	41.95	23.31	44.48	26.00			
Male	46	28.28	19.50	24.15	18.63			
Students with disabilities	15	19.37	12.89	15.13	14.53			
Overall pretest	88			34.73	22.48			
Overall posttest	88			33.73	24.50			

Group Comparison

For Grade 6, the difference in adjusted means was significant for the 6th grade students, $F(1, 217) = 5.687$, $p = .018$. For Grade 7, the differences were not found to be significant, $F(1, 209) = 3.529$, $p = .062$. For Grade 8, there was also no difference in adjusted means, $F(1, 226) = 1.352$, $p = .246$. analysis of covariance results are presented in Table 4.

Summary and Discussion of Findings

The pre- and posttest scores had moderate mean values with relatively large standard deviations showing, on the one hand, a wide student achievement spectrum, and very good discriminatory power of the employed instruments, on the other. The differences of mean scores between the test preparation School A and the control School B were significant for Grade 6, marginally significant for Grade 7, and not significant for Grade 8. The findings in this specific study suggest that the 12 Powerful Words (Bell, 2005) appeared more effective with younger students than with older students. According to Discovery Education (n.d.), students attain larger achievement gains in elementary intermediate Grades 3–5 and gradually decline in performance on achievement tests when reaching middle school. Therefore, it might be reasonable to start teaching test-wiseness at the elementary Grades and to restructure test preparation into curriculum throughout student learning over many years of time (Biggs, Musewe, & Harvey, 2014; McCormick, O'Connor, Cappella, & McClowry, 2013; Rowley & Wright, 2011).

Conclusions

In order to be effective, test preparation strategies that focus specifically on reading must be planned, explicit, and consistent (Kontovourki & Campis, 2010) in order to be effective. Test-wiseness can improve performance of underachieving Black students on standardized reading tests; however, further research should determine why there are differences in achievement gain for male and female students as well as for the different grade levels.

In the United States, school test accountability for student achievement under the mandates of NCLB (2002) are attained by a student's proficiency score on standardized tests in core subject areas (DiGaetano, 2015; Fagioli, 2014). If test accountability is not met, according to NCLB (2002) guidelines, there must be a school restructuring or closure of Grades K–12 learning institution/s. For Tennessee school districts, the use of test preparation strategy could result in the reduction or elimination of schools being taken over and closed by Departments of Education (see De Witte & Moccia, 2011; Engberg, Gill, Zamarro & Zimmer, 2012; Irwin & Seasons, 2012).

References

Alon, S. (2010). Racial differences in test preparation strategies: A commentary on shadow education, American style: Test preparation, the SAT and college enrollment. *Social Forces, 89*, 463–474. doi:10.1353/sof.2010.0053

Bell, L. I. (2005). *12 powerful words that increase test scores and help close the achievement gap. A resource for educations and parents*. Retrieved from https://www.larry-bell.com/product/12-powerful-words/

Bicak, B. (2013). Scale for test preparation and test-taking strategies. *Educational Sciences: Theory & Practice, 13*, 279–289. Retrieved from ERIC Database (ED EJ1016653).

Biggs, S. A., Musewe, L. O., & Harvey, J. P. (2014). Mentoring and academic performance of Black and under-resourced urban middle grade students. *Negro Educational Review, 65*, 64–86.

Coalition for Evidence-Based Policy. (2014). *Which comparison-group "quasi-experimental" study designs are most likely to produce valid estimates of a program's impact? A brief overview and sample review form*. Retrieved from http://coalition4evidence.org/wp-content/uploads/2014/01/Validity-of-comparison-group-designs-updated-January-2014.pdf

De Witte, P. M., & Moccia, J. (2011). Surviving a school closing. *Educational Leadership, 68*, 54–57.

DiGaetano, A. (2015). Accountability school reform in comparative perspective. *Urban Affairs Review, 51*, 315–357. doi:10.1177/1078087414537606

Discovery Education. (n.d.). *Assessment*. Retrieved from http://static.discoveryeducation.com/de/docs/assessment/14111_Tennessee%20Research.pdf

Doe, C., & Fox, J. (2011). Exploring the testing process: Three test takers' observed and reported strategy use over time and testing contexts. *Canadian Modern Language Review, 67*, 29–54. doi:10.3138/cmlr.67.1.029

Ebanks, M., Toldson, I., Richards, S., & Lemmons, B. (2011). Project 2011 and the preparation of Black and Latino students for admission to specialized high schools in New York City. *Journal of Negro Education, 81*, 241–251. doi:10.7709/jnegroeducation.81.3.0241

Engberg, J., Gill, B., Zamarro, G., & Zimmer, R. (2012). Closing schools in a shrinking district: Do student outcomes depend on which schools are closed? *Journal of Urban Economics, 71*, 189–203. doi:10.1016/j.jue.2011.10.001

Fagioli, L. P. (2014). A comparison between value-added school estimates and currently used metrics of school accountability in California. *Educational Assessment, Evaluation and Accountability, 26*, 203–222. doi:10.1007/s11092-013-9183-9

Fede, J. L., Pierce, M. E., Matthews, W. J., & Wells, C. S. (2013). The effects of a computer-assisted, schema-based instruction intervention on word problem-solving skills of low-performing fifth grade students. *Journal of Special Education Technology, 28*, 9–21. doi:10.1177/016264341302800102

Gadbury-Amyot, C. C., Austin, K., & Overman, P. R. (2013). Development and implementation of online national board dental examination review courses. *Journal of Dental Education, 77*, 1556–1565.

Hornof, M. (2008). Reading tests as a genre study. *The Reading Teacher, 62*, 69–73. doi:10.1598/RT.62.1.8

Hughes, C. A., & Schumaker, J. B. (1991). Test-taking strategy instruction for adolescents with learning disabilities. *Exceptionality, 2*, 205–221. doi:10.1080/ 09362839109524784

Hughes, C., Schumaker, J. B., Deshler, D. D., & Mercer, C. (1988). *Learning strategies curriculum: The test-taking strategy*. Lawrence, KS: Edge.

Irwin, B., & Seasons, M. (2012). School closure decision-making processes: Problems and prospects. *Canadian Journal of Urban Research, 21*, 45–67.

Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., ... Levin, J. R. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research, 68*, 350–386.

Kontovourki, S., & Campis, C. (2010). Meaningful practice: Test prep in a third-grade public school classroom. *Reading Teacher, 64*, 236–245. doi:10.1598/ RT.64.4.2

Lam, R. (2013). Formative use of summative tests: Using test preparation to promote performance and self-regulation. *Asia-Pacific Education Researcher, 22*, 69–78. doi:10.1007/s40299-012-0026-0

Lancaster, P. E., Schumaker, J. B., Lancaster, S. C., & Deshler, D. D. (2009). Effects of a computerized program on use of the test-taking strategy by secondary students with disabilities. *Learning Disability Quarterly, 32*, 165–179. doi:10.2307/27740366

Matsumura, L. C., Garnier, H. E., & Resnick, L. B. (2010). Implementing literacy coaching: The role of school social resources. *Education Evaluation and Policy Analysis, 32*, 249–272. doi:10.3102/0162373710363743

McCormick, M. P., O'Connor, E. E., Cappella, E., & McClowry, S. G. (2013). Teacher-child relationships and academic achievement: A multilevel propensity score model approach. *Journal of School Psychology, 51*, 611–624. doi:10.1016/j.jsp.2013.05.001

Millman J., Bishop, H., & Ebel, R. (1965). An analysis of test-wiseness. *Educational and Psychological Measurement, 25*, 707–726. doi:10.1177/001316446502500304

Misco, T. (2010). Remediation for another high-stakes test. *Kappa Delta Pi Record, 46*, 121–126.

Nation's Report Card. (2013). *Grade 8 national results*. Retrieved from http://nationsreportcard.ed.gov/reading_math_2013/#/executive-summary

National Assessment of Educational Progress. (n.d.). *NAEP*. Retrieved from http://nces.ed.gov/nationsreportcard/

No Child Left Behind Act (NCLB) of 2001, Pub. L. No. 107–110, § 115, Stat. 1425 (2002).

Oakland, L., & Weilart, E. (1972). The effects of test-wiseness materials on standardized test performance of preschool disadvantaged children. *Journal of School Psychology, 10*, 355–360.

Rowley, R. L., & Wright, D. W. (2011). No "White" child left behind: The academic achievement gap between Black and White students. *The Journal of Negro Education, 80*, 93–107.

Scruggs, T. E., & Lifson, S. A. (1985). Current conceptions of test-wiseness: Myths and realities. *School Psychology Review, 14*, 339–350.

Stough, L. M. (1993). *Research on multiple-choice questions: Implications for strategy instruction*. Paper presented at the Annual Convention of the Council for Exceptional Children. San Antonio, TX. Retrieved from http://files.eric.ed.gov/ fulltext/ED366135.pdf

Tennessee State Department of Education. (n.d.). *State report card data*. Retrieved from http://www.tn.gov/education/topic/report-card

Thorndike. E. L. (1951). Reliability. In E. F. Lindquist (Ed.), *Educational measurement* (pp. 560–620). Washington, DC: American Council on Education.

Welsh, M. E., Eastwood, M., & D'Agostino, J. V. (2014). Conceptualizing teaching to the test under standards-based reform. *Applied Measurement in Education, 27*, 98–114. doi:10.1080/08957347.2014.880439

Winsler, A., Karkhanis, D. G., Kim, Y. K., & Levitt, J. (2013). Being Black, male, and gifted in Miami: Prevalence and predictors of placement in elementary school gifted education programs. *Urban Review, 45*, 416–447. doi:10.1007/s11256-013-0259-0

Yang, P. (2000). *Effects of test-wiseness upon performance on the test of English as a foreign language* (Unpublished doctoral dissertation). University of Alberta, Edmonton, Canada.

Zhengdong, G. (2009). IELTS preparation course and student IELTS: A case study in Hong Kong. *RELC Journal, 40*, 23–40. doi:10.1177/0033688208101449