Jesse L. M. Wilkins
Brett D. Jones

Planning and Changing
Vol. 40, No. 3&4, 2009, pp. 194–206

AN EXAMINATION OF A STATE-LEVEL MATHEMATICS TEST: WHAT TYPE OF LEARNING DOES THE TEST ACTUALLY MEASURE?

Introduction

Most U.S. states have developed sophisticated assessment programs to evaluate student achievement and hold schools, teachers, and students accountable for learning important content. The No Child Left Behind Act of 2001 (NCLB, 2002) has placed further pressure on educators and administrators to ensure that all students are learning. Yet, some educators and researchers claim that test-based accountability has had detrimental effects on students' learning by causing teachers to focus on low-level knowledge and skills, resulting in less in-depth understanding and less focus on higher-order thinking skills (Jones, Jones, & Hargrove, 2003; Kohn, 2000). The purpose of this article is to examine one high-stakes test in Virginia to provide some evidence as to whether the test focuses primarily on basic skills or includes high-order thinking such as conceptual understanding and problem solving.

The Effects of Testing on Teaching and Learning

Educators are often the most outspoken about how test-based accountability has prevented them from being effective. In fact, over 90% of teachers in one study believed that students would learn the same amount or more without the statewide standardized tests (Jones & Egley, 2007). These teachers' beliefs appear to be consistent with the results of one major study that compared students' scores on high-stakes tests to those of other common assessments (i.e., the ACT, SAT, NAEP and AP tests) and found that student learning had stayed the same or decreased since the implementation of high-stakes testing (Amrein & Berliner, 2002). One reason for the stagnant or decreasing test scores may be that the standardized tests cover a wide range of topics in the curriculum areas tested, thus, giving teachers less time to devote to in-depth exploration of a topic. As one teacher in Texas reported: "We try to do hands-on kinds of things actively involving students, but we realize we have to spend lots of time on drill and practice with paper and pencil because of the way the test is formatted" (Gordon & Reese, 1997, p. 353). A teacher in Florida also explained:

> Before [standardized testing] I was a better teacher. I was exposing my children to a wide range of science and social studies experiences. I taught using themes that really immersed the children into learning about a topic using their reading, writing, math, and technology skills. Now I'm basically afraid to NOT teach to the

المنارة للاستشارات

> test. I know that the way I was teaching was building a better foundation for my kids as well as a love of learning. Now each year I can't wait until [the testing] is over so I can spend the last two and a half months of school teaching the way I want to teach, the way I know students will be excited about. (Jones & Egley, 2004, Themes 2 and 7: Effects on the Curriculum)

The fact that at least some (if not many) teachers are spending more time on "skill and drill" can be problematic because current learning theories emphasize the importance of understanding, as opposed to the rote memorization of facts (National Research Council, 2000). This is not to imply that it is unimportant to learn facts about a subject, only that this should not be the sole focus of learning. Instead, experts in any field of study organize their problem solving around big and important concepts (Voss, Greene, Post, & Penner, 1983). As the National Research Council (2000) has stated:

> The new science of learning does not deny that facts are important for thinking and problem solving…. However, the research also shows clearly that "usable knowledge" is not the same as a mere list of disconnected facts. Experts' knowledge is connected and organized around important concepts (e.g., Newton's second law of motion); it is "conditionalized" to specify the contexts in which it is applicable; it supports understanding and transfer (to other contexts) rather than only the ability to remember. (p. 9)

It is also important to note that it takes major investments of time to develop understanding and expertise in an area (National Research Council, 2000, p. 56). This point is well stated by the National Research Council (2000):

> Learning with understanding is often harder to accomplish than simply memorizing, and it takes more time. Many curricula fail to support learning with understanding because they present too many disconnected facts in too short a time – the "mile wide, inch deep" problem. Tests often reinforce memorizing rather than understanding. (p. 24)

In sum, current learning research emphasizes the importance of teaching for understanding, which occurs as facts are connected and organized around concepts. It takes students time to reach a level of deep understanding in a subject and to have the ability to transfer this understanding to different contexts.

Although some teachers have reported that high-stakes testing has encouraged them to focus on low-level knowledge and skills (Barksdale-Ladd & Thomas, 2000; Hoffman, Assaf, & Paris, 2001), it is unclear whether high-stakes testing has affected teaching practices in any systematic manner that can be generalized to all teachers and states (Cimbricz, 2002; Jones et al., 2003). We believe that research based on educators' beliefs should be taken seriously because they are in the best position to make judgments about students' learning given that they work with students on a daily ba-

المنارة للاستشارات

sis. Yet, to provide a more complete picture of how test-based accountability programs are affecting educators and students, we believe that it is also important to examine test scores and whether they indicate that students are learning higher-order knowledge in addition to basic skills. Consequently, in the present study, we investigated whether the scores from a statewide, high-stakes mathematics test were valid indicators of not only students' basic skills, but also their conceptual understanding and problem solving.

The Use of Test Scores for Test-Based Accountability

The Commonwealth of Virginia has developed its *Standards of Learning* (SOL) (Board of Education, BOE, 1995) and associated tests to evaluate the success of schools and teachers in enabling their students to meet these standards. At the time of this study, students in Virginia were administered mathematics achievement tests in grades 3, 5, and 8 and end-of-course tests for high school courses (e.g., algebra I, algebra II, and geometry).[1] These tests are based on the Virginia SOL (BOE, 1995) and are meant to assess students' proficiency related to the SOL. At that time, for Virginia schools to reach and maintain accreditation they were expected to have a school-wide passing rate of at least 70% on each of the mathematics tests.[2]

The mathematics SOL (BOE, 1995) identifies six content strands of mathematics important for students to learn: Number and Number Sense, Computation and Estimation, Measurement, Geometry, Probability and Statistics, and Patterns. In addition, the SOL point out that, "students must gain an understanding of fundamental ideas in [mathematics]…and develop proficiency in mathematical skills" (p. 1). Further, the SOL declare that: "Problem solving has been integrated throughout the six content strands. The development of problem-solving skills should be a major goal of the mathematics program at every grade level" (p. 4). The assessments were developed to assess what students are learning in relation to the SOL. The validity of these test scores was initially evaluated by a committee of testing and measurement experts brought together by the Virginia Board of Education (Hambleton et al., n.d.).

The validity of test scores is related to the degree to which the scores measure what they are intended to measure (Carmines & Zeller, 1979). Based on the statements from the SOL document (BOE, 1995), the intent of the SOL tests is to measure the degree to which students develop their mathematical skills, conceptual understanding, and problem-solving in the six content strands in mathematics. Content validity is related to the degree to which test scores provide evidence about a particular content domain, such as, for example, the degree to which the items on the Grade 3 SOL mathematics test match the Grade 3 SOL. Content validity also attends to the broadness with which the SOL are being assessed. In other words, at what process levels are the SOL being evaluated? Hambleton et al. (n.d.) have pointed out that the multiple-choice format may not be suitable for

المنارة للاستشارات

assessing all of the intended goals of the SOL. For example, if the SOL called for students to show their conceptual understanding of a mathematical idea or show their ability to solve a nonroutine problem, the design of the SOL tests may not be adequate to assess these process standards. Hambleton et al. (n.d.) recommended that further evidence be provided

> on the extent to which the test specifications match the SOL, and where they do not, [indicate]…the steps that are in place for insuring that the areas of the SOL not covered on the assessments are taught and assessed in other ways. (p. 3–4)

The construct validity of a test is a measure of the degree to which the test scores can be said to measure the underlying "construct" that it was intended to measure, in this case, conceptual understanding and problem solving knowledge in mathematics. One way of assessing the construct validity of a mathematics test is to correlate test scores with a known and accepted standardized mathematics achievement test. For example, using students' Stanford 9 achievement test scores, positive correlations between SOL test scores and the Stanford 9 test scores provide evidence that the tests have some level of construct validity in terms of general mathematical knowledge (Hambleton et al., n.d.). In the present study, we examined whether the scores from a Grade 3 Virginia SOL mathematics test were valid indicators of students' basic skills, conceptual understanding, and problem solving by comparing these SOL scores to scores on another standardized mathematics test.

Method

Sample

This study included four successive cohorts of fourth grade students from one school district in Virginia. Only students with available test scores were used in the study, resulting in a working sample of $N = 3150$. There were 779, 798, 800, and 773 students in Cohorts 1, 2, 3, and 4, respectively.

Measures

Student scaled scores and proficiency ratings from the Grade 3 SOL mathematics test were used as one measure of mathematic achievement. The Grade 3 SOL test is administered in the spring of the school year and the items are in a multiple-choice format. SOL scaled scores range from 0–600, with scores ranging from 0–399 labeled "failure," scores ranging from 400–499 labeled "pass/proficient," and scores ranging from 500–600 labeled "pass/advanced." For the purpose of this study, student scores were collapsed into two categories: "proficient" (i.e., pass/proficient and pass/advanced) and "not proficient" (i.e., failure).

A second measure of student achievement was gathered using the

المنارة للاستشارات

New Standards Reference Examination (NSRE), published by Harcourt, Inc. The NSRE was administered to students in the spring of their fourth grade school year. The NSRE is based on the *New Standards Performance Standards* (National Center on Education and the Economy, NCEE, 1997) and assesses students' mathematical knowledge in the different content strands of mathematics advocated by the National Council of Teachers of Mathematics (NCTM) (2000) and other professional organizations. In addition, the NSRE assesses student performance for three process levels: skills, conceptual understanding, and problem solving. The skills component assesses students' performance on basic mathematical procedures and techniques. The conceptual understanding component assesses students' understanding of mathematical processes and ideas. The problem solving component assesses students' ability to reason mathematically and apply their mathematical knowledge to problem situations. The items on the NSRE use multiple formats including multiple-choice, short answer, and open-ended. Student scores as well as performance standard levels were used in this study. Student scores for each of the components are measured on a continuous metric. Although the range for student scores differed by test form each year, the scores are used to determine performance standard levels for each student that are comparable across time and different test forms. Performance standard levels are based on the *New Standards Performance Standards* and are reported in five levels: "achieved the standard with honors," "achieved the standard," "nearly achieved the standard," "below the standard," and "little evidence of achievement." For the purpose of this study, students' performance levels were collapsed into two categories: "proficient" (i.e., achieved the standard with honors and achieved the standard) and "not proficient" (i.e., nearly achieved the standard, below the standard, and little evidence of achievement).

Results and Discussion

Correlations Between SOL and NRSE Test Scores

Our goal was to assess whether passing a statewide, high-stakes mathematics test was indicative of students' mathematical conceptual understanding and problem solving knowledge. To do so, we compared the scaled scores for a Grade 3 SOL mathematics test to the scores on the three components of the NSRE (skills, conceptual understanding, and problem solving). Using data from all four cohorts of students, we calculated Pearson correlation coefficients for the relationship between the scores. Because the score range for NSRE differed each year for each cohort (i.e., test form), we calculated correlations between the SOL and the three components of the NSRE for each cohort independently. We found all of the correlations to be positive, statistically significant, and within the range of .62 to .81 (see Table 1).

المنارة للاستشارات

Table 1

Correlations Between Scaled Scores on the Grade 3 SOL Mathematics Test and Scores on the Three Components of the NSRE

	N	Skills	Conceptual	Problem solving
Cohort 1 SOL	779	.64	.81	.74
Cohort 2 SOL	798	.68	.77	.68
Cohort 3 SOL	800	.69	.78	.66
Cohort 4 SOL	773	.71	.76	.62

Note. All correlation coefficients were statistically significant at $\alpha = .001$.

Based on the relatively high correlations between the SOL test scores and the three NSRE component test scores, one might conclude that the SOL test scores have construct validity in terms of skills, conceptual understanding, and problem solving such that success on the SOL test indicates high skills, conceptual understanding, and problem solving ability. However, correlations between two scores merely indicate the strength of the relationship between the scores in terms of rank order, that is, how well one can predict one score from another. Correlations do not take into account any criterion or standard; therefore, whereas high scores on one test may be predictive of high scores on a second test, it may be the case that higher scores on the second test are not high enough to meet the criterion set for proficiency. For example, students who pass the SOL mathematics test may or may not score high enough to be deemed proficient on the skills, conceptual understanding, and/or problem solving components of the NSRE test.

Proficiency Ratings for the SOL and NRSE Tests

We compared proficiency ratings (i.e., whether students were proficient or not) for the Grade 3 SOL mathematics test to proficiency ratings for the three components of the NSRE: skills, conceptual, and problem solving. Percentages for those students proficient on the SOL test and each of the components of NSRE are presented in Table 2. In the discussion that follows we focus on the overall sample of students; however, information by cohort is also presented in Table 2. Because we compared proportions across the same set of students, we used a McNemar test for related samples. Overall, 84% of the sample passed the Grade 3 SOL mathematics test, which is greater than the 70% that is needed for the school district to be labeled proficient. Thus, the district was labeled as proficient for Grade 3 mathematics. However, for the skills component of the NSRE, overall, 68% of the students were found to be proficient. This proportion was found to be significantly lower than that for the SOL test, $\chi^2(1, N = 3{,}150) = 373.89$, $p < .0001$. Less than half of the students, 41%, were found to be proficient for conceptual understanding, significantly lower than the pro-

المنارة للاستشارات

portion of students passing the SOL test, $\chi^2(1, N = 3{,}150) = 1{,}337.24$, $p < .0001$. Even fewer students, 26%, were found to be proficient in problem-solving, again, significantly lower than the proportion of students passing the SOL test, $\chi^2(1, N = 3{,}150) = 1{,}831.03$, $p < .0001$.

Table 2

Percentage of Students Rated 'Proficient' for the Grade 3 SOL Mathematics Test and Percentage of Students 'Proficient' for the Three Components of the NSRE

	N	SOL	NSRE		
			Skills	Conceptual	Problem solving
Cohort 1	779	80	68	36	22
Cohort 2	798	82	69	42	25
Cohort 3	800	87	71	45	31
Cohort 4	773	88	66	43	27
Total	3,150	84	68	41	26

Note. Only students with both SOL and NSRE scores are included.

To better understand what a proficient score on the SOL test indicates in terms of students' skills, conceptual understanding, and problem solving, we further examined the level of proficiency on the NSRE components for only those students who passed the SOL test (see Table 3).

Table 3

Percentage of Students Rated 'Proficient' on the Grade 3 SOL Mathematics Test Who Were Also Rated 'Proficient' for the Three Components of the NSRE

	N	NSRE		
		Skills	Conceptual	Problem solving
Cohort 1	621	80	45	27
Cohort 2	656	79	50	29
Cohort 3	695	80	51	35
Cohort 4	684	73	48	30
Total	2,656	78	49	31

Note. Only students with both SOL and NSRE scores are included.

We found that 78% of those students who passed the Grade 3 SOL mathematics test were proficient for the skills component of the NSRE. However, significantly fewer students, only 49%, were proficient on the conceptual component, $\chi^2(1, N = 2{,}656) = 660.28$, $p < .0001$; and even fewer

المنارة للاستشارات

students, 31%, were proficient on the problem-solving component, $\chi^2(1, N = 2{,}656) = 1{,}199.56$, $p < .0001$. The proportion of students who were proficient for problem solving was also significantly lower than that for the conceptual component, $\chi^2(1, N = 2{,}656) = 317.35$, $p < .0001$.

Based on the above findings, although there was not a one-to-one agreement between proficiency on the SOL test and the skills component of the NSRE, a large proportion of students who passed the SOL mathematics test was also proficient on the skills component of the NSRE. Passing the SOL test, however, was not a good indicator of student success for the conceptual and problem solving component of the NSRE. Less than half of those students passing the SOL test were proficient for the conceptual understanding component, and less than a third of those students were proficient for the problem-solving component. At best, the SOL test seems to be a reasonable indicator of basic skills and procedures.

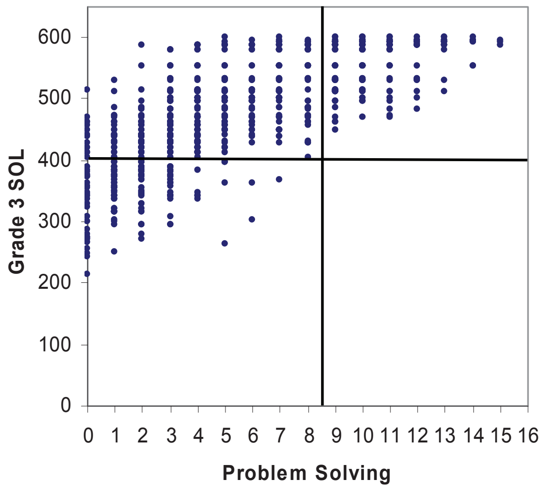
Interpretation of the Correlations and Passing Rates

What do the findings from the correlations in Table 1 and the proficiency ratings in Tables 2 and 3 tell us? As the correlations in Table 1 show, the SOL mathematics test scores are correlated with skills, problem solving, and conceptual understanding as measured by the NSRE. Based on the magnitude and positive direction of these correlations, students who have higher scores on the SOL mathematics test tend to have higher skills, conceptual understanding, and problem solving NRSE scores. When we reviewed the proficiency ratings presented in Table 3, however, we found that a passing score on the SOL test did not necessarily indicate that students had developed the conceptual understanding or problem solving skills necessary to be proficient on these components of the NSRE (i.e., the students did not achieve the standard as defined by the *New Standards Performance Standards*).

To understand this concept, it can be helpful to refer to visuals such as those in Figure 1. Figure 1 presents scatterplots of students' SOL scaled scores versus their scores on the problem solving and conceptual understanding component of the NSRE for students in Cohort 1. Students who have passed the SOL test are represented by dots above the horizontal lines in Figure 1 (scaled score of 400 and above). Students who were not proficient on the NSRE are represented by dots to the left of the vertical lines (representing a score of 9 for problem solving and 37 for conceptual understanding). Thus, students in the upper-left quadrant of the two scatterplots in Figure 1 were proficient on the SOL test but not proficient on the problem solving and conceptual understanding components of the NSRE; students in the upper-right quadrant of the two scatterplots were proficient on the SOL test and the problem solving and conceptual understanding components of the NSRE. These later students are also represented in Table 3 as the proportion of students that was proficient on the SOL

المنارة للاستشارات

test and proficient on the conceptual understanding or problem solving components of the NSRE.



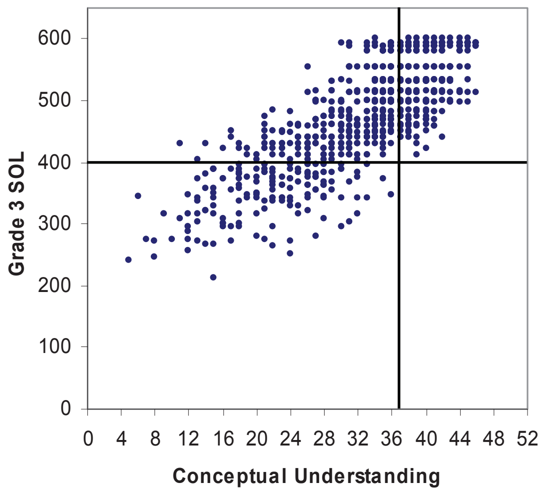


Figure 1. Scatterplots of Grade 3 SOL scaled scores versus Problem Solving (r = .74) and Conceptual Understanding scores (r = .81) for Cohort 1. Inserted lines represent cut-score values for which students were considered “proficient:” SOL (400), Problem-Solving (9), and Conceptual Understanding (37).

المنارة للاستشارات

These findings suggest that the scores from the Grade 3 SOL mathematics test lack construct validity and general content validity because the test scores are not indicative of the degree to which students have developed their conceptual understanding and problem solving as purported by the Virginia mathematics standards (BOE, 1995). The evidence for this is that many more students are proficient on the SOL mathematics test than are proficient on the conceptual understanding and problem solving components of the NSRE. At best, the Grade 3 SOL mathematics test measures students' proficiency at performing basic mathematics skills and procedures. However, the proficiency ratings for the test overestimate the extent of students' conceptual understanding and problem solving abilities.

Implications and Conclusion

It is reasonable for students, parents, teachers, and principals to expect that the Grade 3 SOL mathematics test measures students' mathematical skills, conceptual understanding, and problem solving in the six content strands of mathematics as specified in the Virginia SOL (BOE, 1995). Unfortunately, based on our findings, a passing score on the test does not indicate that students have mastered the conceptual understanding and problem solving specified in the mathematics SOL. As a result, the test scores must be interpreted and used with caution.

At a minimum, the Grade 3 SOL mathematics test is a measure of basic mathematics skills and procedures and should be modified so that the scores received from it are better indicators of students' higher-order thinking (such as conceptual understanding and problem solving). Without being privy to the psychometric properties of the Grade 3 SOL mathematics test, we cannot make specific recommendations for improving this test. However, ideas from a recent position statement on high-stakes testing from the NCTM (2006) might be worth considering. To ensure that students are not missing out on important mathematical competencies, states must balance their assessment programs to include multiple tools to tap the different levels of understanding; and that additional evidence, beyond a single multiple-choice test, can be collected to indicate whether students are learning higher-order thinking skills (NCTM, 2006). Alternatively, the test could be renamed the "Grade 3 SOL mathematics test of basic skills" to reflect the fact that it is an indicator of students' basic skills (at best) and not an indicator of their conceptual understanding or problem solving abilities.

Without changes to the test, schools and school districts that have successfully met the 70% passing rate (and Adequate Yearly Progress [AYP] as presently required by NCLB) would likely be lulled into a false sense of success, even though many of these students would not have achieved proficiency in conceptual understanding and problem solving. Teachers, administrators, parents, politicians, and the general public would sensibly assume

المنارة للاستشارات

that students are obtaining mathematical knowledge in public schools, when in fact, they may be learning lower-level skills but not developing true conceptual understanding and problem solving abilities.

An obvious implication for principals would be not to rely too heavily on the Grade 3 SOL mathematics test scores, as they often do in high-stakes testing environments. Rather, principals should ensure that teachers are using appropriate methods and teaching the appropriate content to help students develop their skills and higher-order thinking abilities. Otherwise, principals might mistakenly believe that teachers whose students score highly on the SOL mathematics test are using the best methods and those whose students are not scoring highly are not. But given that the Grade 3 SOL mathematics test does not measure students' high-order thinking skills, teachers whose students score highly on the SOL mathematics test may or may not be the "best" teachers.

Consider a case where a third grade teacher in Virginia is teaching basic mathematical skills, but rarely teaching conceptual understanding or problem solving. Many of his students might pass the Grade 3 SOL mathematics test if it measures basic skills, but not conceptual understanding or problem solving. Consequently, he will continue teaching in a similar manner in the future because he has received acceptable results and the parents and administrators are pleased with the test scores. In this case, the teacher is limiting students' mathematical development by focusing primarily on skills and procedures. In contrast, students in his courses would score lower if they completed a test that truly represented their mathematical skills, conceptual understanding, and problem solving. Faced with poor test scores and the knowledge that he has not appropriately taught conceptual knowledge and problem solving, the teacher would likely alter his methods and/or content and focus more on these areas.

We agree with teachers that holding students, teachers, and schools accountable for meeting state standards is important (Jones & Egley, 2004); however, the test-based accountability programs used by states should serve as a way of helping reach the intended goals and not ultimately limiting what students learn. Further, tests should help diagnose students' misunderstandings (Popham, 2000) instead of providing a false sense of success by indicating that students are mathematically proficient when they have not developed important conceptual understanding and problem solving knowledge.

The results from the present study do not allow us to generalize our findings to other standardized tests. Nonetheless, we believe that the results provide a cautionary tale that the scores from statewide standardized tests should be interpreted carefully. We would advise against administrators providing rewards or sanctions to teachers based on standardized test scores alone. Other types of evidence (e.g., classroom observations) used in combination with test scores would provide a more complete and accurate assessment of a teacher's abilities.

المنارة للاستشارات

We hope that by bringing attention to one example of a test that does not deliver what it purports to, we will encourage others to be cautious when considering how to use the high-stakes test scores. Only through a better understanding of what the tests are actually measuring can we begin to interpret the test scores that they provide.

End Notes

[1] As a result of NCLB, mathematics tests for grades 4, 6, and 7 have been added since the time of data collection for this study.
[2] Currently schools must also meet Adequate Yearly Progress (AYP) as required by NCLB.

Author Note

The research reported in this study was supported in part by National Science Foundation Grant #9911558. Any conclusions stated here are those of the authors and do not necessarily reflect the position of the National Science Foundation.

References

Amrein, A. L., & Berliner, D. C. (2002, March 28). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives, 10*(18). Retrieved November 3, 2009, from http://epaa.asu.edu/epaa/v10n18/

Barksdale-Ladd, M., & Thomas, K. (2000). What's at stake in high-stakes testing: Teachers and parents speak out. *Journal of Teacher Education, 51*(5), 384–397.

Board of Education (BOE). (1995). *Mathematics Standards of Learning for Virginia public schools.* Richmond, VA: Commonwealth of Virginia.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment.* Newbury Park, CA: Sage.

Cimbricz, S. (2002, January 9). State-mandated testing and teachers' beliefs and practice. *Education Policy Analysis Archives, 10*(2). Retrieved November 3, 2009, from http://epaa.asu.edu/epaa/v10n2.html

Gordon, S. P., & Reese, M. (1997). High-stakes testing: Worth the price? *Journal of School Leadership, 7,* 345–368.

Hambleton, R. K., Crocker, L., Cruse, K., Dodd, B., Plake, B. S., & Poggio, J. (n.d.). *Review of selected technical characteristics of the Virginia Standards of Learning (SOL) assessments.* Retrieved October 28, 2009, from http://www.pen.k12.va.us.VDOE/Assessment/virginiareport.pdf

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

المنارة للاستشارات

Jones, B. D., & Egley, R. J. (2004, August 9). Voices from the frontlines: Teachers' perceptions of high-stakes testing. *Education Policy Analysis Archives, 12*(39). Retrieved November 3, 2009, from http://epaa.asu.edu/epaa/v12n39/

Jones, B. D., & Egley, R. J. (2007). Learning to take tests or learning for understanding? Teachers' beliefs about test-based accountability. *The Educational Forum, 71*(3), 232–248.

Jones, M. G., Jones, B. D., & Hargrove, T. Y. (2003). *The unintended consequences of high-stakes testing.* Lanham, MD: Rowman & Littlefield.

Kohn, A. (2000, September 27). Standardized testing and its victims. *Education Week.* Retrieved November 3, 2009, from http://www.edweek.org/ew/articles/2000/09/27/04kohn.h20.html

National Center on Education and the Economy (NCEE). (1997). *New Standards Performance Standards.* Washington, DC: Author.

National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics (NCTM). (2006). *High-stakes tests: A position of the National Council of Teachers of Mathematics.* Retrieved November 3, 2009, from http://www.nctm.org/about/content.aspx?id=6356

National Research Council (NRC). (2000). *How People Learn.* Washington, DC: National Academy Press.

No Child Left Behind Act of 2001. (2002, January 8). Public Law 107-110, 115 Stat. 1425. Retrieved November 3, 2009, from http://www.ed.gov/policy/elsec/leg/esea02/107-110.pdf

Popham, W. J. (2000). *Modern educational measurement: Practical guidelines for educational leaders.* Boston: Allyn & Bacon.

Voss, J. F., Greene, T. R., Post, T. A., & Penner, B. C. (1983). Problemsolving skill in the social sciences. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation: Vol. 17. Advances in Research and Theory* (pp. 165–213). New York: Academic Press.

Jesse L. M. Wilkins is an Associate Professor in the Department of Teaching and Learning at Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Brett D. Jones is an Associate Professor in the Department of Learning Sciences and Technologies at Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

المنارة للاستشارات

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

www.manaraa.com