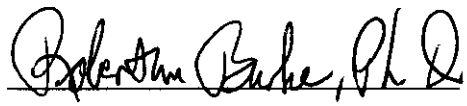
The Performance Battle

An Honors Thesis (HONRS 499)

by

Krista Albertson

Dr. Robert Burke



Ball State University
Muncie, Indiana

March 2004

Graduation Date: May 8, 2004

Abstract

The development of "The Performance Battle" began as an interest in the standardized testing issue within our country. The debate continues over performance-based testing versus multiple-choice tests. In this thesis, the history of standardized testing is discussed as well as the various types of multiple-choice tests that are used to assess our students. Three state assessment programs are examined to survey different assessment strategies. Lastly described is the alternative of performance-based assessment that could be used as another way to answer the public's questions about student performance.

Acknowledgements

- I owe a great deal of thanks to Dr. Burke for all of his help throughout my thesis journey. Thank you for all of your time as well as your wonderful support and guidance.

- My parents also deserve great recognition for their assistance throughout this process as well as the rest of my college career. Thank you so much for all of your loving encouragement.

"Not everything that counts should be counted and not everything that can be counted counts." Our country could take this little bit of advice from Albert Einstein. Our children's education definitely "counts," but should they be "counted" using the country's current system of measurement? Testing has been a major part of our schools for years upon years, but has it gone too far? How are we currently assessing our children? Are there other ways that could be more beneficial? Is someone in the country already taking the lead in new practices? The remainder of this thesis will address these issues and more regarding standardized testing in our country.

The History of Assessment in the United States

Testing our students is by no means a new concept. Assessment has been used in our schools since the beginning of the classroom; however, recent confidence in standardized tests has led to less confidence in the teacher. The first intelligence tests were created in the late 1800's with interest in the genetics of intelligence as the focus. Individuals were tested and then sorted into intelligence categories based on their results. The United States Army used intelligence tests to sort the soldiers during World War I and, after the war, schools began using similar tests to sort the children. During the 1920's, these tests were used to assign students to various programs and "tracks".

In 1957, fear was sparked within the United States when Russia launched Sputnik. The country began to fear that its students were falling behind other countries in math and science. The schools began new programs to emphasize these subjects and used tests to measure the improvement of the students. The United States recaptured the lead in space in an amazingly short time – ironically (points out Larry E. Frase, professor of educational administration at San Diego State University), with scientists and engineers

trained in the supposedly outdated science and math. Testing became almost universal in the 1960's and 70's when the testing technology became cheaper and easier to use (Cookson & Halderstam, 1998).

"A Nation At Risk" was published in 1983, causing new anxieties in the country. The authors pointed out the inadequacies of American education after comparing our students' scores with children in other countries such as Japan, and they called for more testing. The American public was still fearful of falling behind other countries in math and science and test publishers profited from that fear, putting out more and more tests. These tests were external – created and scored by people outside the school system – and therefore seemed objective. They were also easy to produce and score, and the results were quickly known. Confidence was being lost in the public school system and it seemed that no one could be trusted to explain how well the students were performing; so, the nation turned to tests. It is this lack of trust in public education that keeps our country reliant on standardized tests. Around this same time, the states began to adopt standards in order to have written goals for all schools. This then gave testing another purpose – to ensure that the schools were meeting these standards.

The 1980's were full of other discoveries as well. American businesses realized that they needed a different kind of worker than our schools were producing – one that is creative, can solve problems, and work well in groups. This is far from what the students were learning in school at the time and far from what the tests were assessing. At the same time, school districts were continually judged on the results of the now "high-stakes" standardized tests. A test has high stakes when its results are instrumental in a decision such as moving a student to the next grade, increasing a teacher's pay, or

granting funding to a school (Chase & Katz, 2002). Some teachers began to "teach to the test" in order for their students to perform well.

Also, in the next 10 years, professional development was given to teachers to begin learning about "authentic assessment" (Janesick, 2001). Finally, in 2000, teacher education programs began to use portfolios and the future teachers began to learn more about measuring student performance. Wonderfully, this is leading me, along with a new generation of teachers, to a better understanding of assessment.

The History of ISTEP

As mentioned before, many states in the nation became concerned about the performance of their students, and Indiana was no different. The state developed the Indiana Statewide Testing for Educational Progress, or ISTEP, as a way to measure how well the students were performing. The test was first administered in 1988. The original test included a multiple-choice component and a writing component that was later eliminated. New legislation in 1995 required that the test (now called ISTEP+) include a norm-referenced test as well as a criterion-referenced test. These sections included multiple choice, short answer, and essay questions. In 2002, the Grades 3, 6, and 8 ISTEP+ tests were modified to reflect the state academic standards adopted in November 2000 (ISTEP+ program manual 2002-2003). As of Fall 2002, the norm-referenced section is optional, and with the recent No Child Left Behind Act, the tests are administered at every grade level from second to tenth grade, with the tenth grade test a requirement for graduation.

No Child Left Behind

The latest development in our country's testing history is the passing of the No Child Left Behind Act. This act requires that each state measure every public school student's progress in reading and math (and science by the year 2007) in each grade three through eight, and at least once during grades 10 through 12. The state must decide on a measurement of "adequate yearly progress" and display the results of this measurement each year in the form of a public report card. The scores must be divided by ethnicity, gender, student disability, and income status, and compare annual goals to actual performance for each group. Schools in which all groups perform well may give their teachers a monetary bonus (using Title I funds) and schools that do not meet the adequate yearly progress for two or more consecutive years must take "corrective actions" such as start new programs or fire teachers, and allow parents to move their children to schools that are performing better (Testing: Frequently asked questions).

The response from many schools around the country is that they are not receiving sufficient funds to follow-through with these requirements. Recently, in Utah legislators pushed a bill forward that would leave the No Child Left Behind Act behind as well as the $103 million that it brings to the state. Representative Margaret Dayton says that her bill sends a message to the federal government that they are going too far by no longer leaving the testing and accountability responsibility to the states. She believes that the schools in her state should not be held accountable to the federal government. Of course, many in the state do not want to sacrifice over $100 million to the cause, even if they do agree with Dayton. That would be doing the state's children more harm than good. They are hoping, however, that the bill will get their point across to the federal government.

Similarly, the Virginia House has passed a resolution asking the federal government to allow them to use their state's own accountability system. Many other states are following the lead of Utah and Virginia in their fight against this controversial Act (Lynn, 2004).

Multiple-Choice Tests

Norm-Referenced Tests

The tests that states do use to measure student progress often include multiple-choice norm-referenced and criterion-referenced tests. Norm-referenced tests (NRT) are multiple-choice tests that compare the students' scores to a previously tested group – the "norm" group. The purpose is to see how well each child performed on a test compared to other students his or her age across the nation. These tests are based on the Bell Curve, which means that a "good" test is one in which each question has a 50/50 rate – 50% of the students get that question right and 50% get it wrong (50% pass and 50% fail). This will ensure that the scores will be distributed on the normal, bell-shaped curve. This also means that the questions have to be difficult, because half of the students must get them wrong. For example, let's say that the test is to be given to third graders. If you ask the students questions about second grade material that you know they have covered – material that they should have learned – then more than 50% are going to get it right. Therefore, the test-makers have to ask third and forth grade questions or difficult second grade questions that they know fewer students will be able to correctly answer – not necessarily questions that will test what they should have learned (Bracey, 2000).

The politicians of our time would like to see all students score above the national average. However, with an NRT, that is impossible. NRT's are constructed so that half

the population falls below the midpoint, as is the case with the Bell Curve (Janesick, 2001). There is no way for more than half of the students taking the test to score above the average, yet no one wants to see that half of the students are failing. This seems like a lose-lose type of test. As far as Indiana is concerned, the norm-referenced portion of the ISTEP+ has recently become optional, which in my opinion is a step forward for the state. William Graham Sumner said "We throw all our attention on the utterly idle question of whether A has done as well as B when the only question is whether A has done as well as he could" (Cookson & Halberstam, 1998).

Criterion-Referenced Tests

Criterion-referenced tests are more informative than norm-referenced tests because they compare a student's score to a set level, or criterion, rather than an established norm. In this way, the test score can be used to understand what a child knows or does not know. In Indiana, the state standards are the basis for the criterion to which the students' scores are being compared. The problem, however, is that most of these tests are multiple-choice. In Indiana, a multiple-choice section as well as a short answer and essay section are used, which is much more personal and can be even better used to determine in what areas the child excels or needs further instruction. Rubrics are used to score the essay portions of the ISTEP. These rubrics can be very helpful because if the child can see the rubric before or while she writes her essay, she can know exactly what is expected of her. This is not the case with a multiple-choice test.

Downfalls of Multiple-Choice Tests

Multiple-choice tests are quick to administer and score, and can therefore produce quick results; however, many problems also exist with these tests. First, the selection of a

multiple-choice answer is passive. The students do not need to contribute their own thinking to the answer and in that way the tests do not prepare students for the real world. Second, they teach children that there is a right or wrong answer to every question and problem, which we know is very rarely true in real life. Also, by taking this kind of test children are taught to rely on memorization and test-taking rules to pass. The students learn that they do not have to understand the material, but rather just memorize it and know the tricks of taking the test to get a good score. The attitude seems to be, "Never mind thinking – that will only slow down the response time" (Mitchell, 1992, p. 174).

The majority of the ISTEP and similar tests are multiple-choice and include similar items. Two such items from portions of the third grade ISTEP test that addresses "word recognition, fluency, and vocabulary development" are as follows (ISTEP+ grade 3 item sampler, 2002):

1. Read this sentence.

Dr. Smith waved to us when he rode by.

What does Dr. mean?
a. Dear
b. Doctor
c. Dollar
d. Drive

2. Read this sentence.

There were three _____ and a cat in the cartoon.

Which word BEST completes the sentence?
a. Mouse
b. Mouses
c. Mice
d. Mices

Other portions of the test assess "reading comprehension" by asking several multiple-choice questions after the child reads a passage. These items may ask the student to select specific details from the passage or require the student to infer the thoughts of a character based on the information given.

Similar test items are also used to assess "writing processes." One such item follows (ISTEP+ grade 3 item sampler, 2002):

3. Choose the sentence that BEST completes the following paragraph.

Winter is fun because there are so many things to do. When it is very cold, we can go ice-skating on the lake. _____. We can even make snowballs.

a. We like the month of January best.
b. When it snows, we can go sledding.
c. We can go swimming in the ocean.
d. When summer is here, we never have snow.

What Are Test Results Used For?

When thinking about a standardized test on its own, it really does not sound too bad. The students spend a few days taking tests that may not be very fun, easy, or even fair; but, what can it hurt? The question that must be asked next is: What are the test results used for? Do they offer insight to the teachers to let them know where the students stand on their incoming knowledge? Does it give the teachers information on how the students are progressing? If so, are we assuming that the teachers (college educated individuals, interested in the success of every one of their students) are not already doing this in their classrooms? Do they help parents to see what their children have learned since the last time they took the test? Are we then saying that the teachers are not already sending work home to show student progress? Are they used for administrators to see what problems their teachers need to address? If so, are we saying

that the administrators are not already asking their teachers about what subjects the children need most help on? Again, we end up back at the lack of trust that was mentioned earlier and also land on lack of communication.

Public Evidence

Most often, the results are used to give the public evidence of the quality of the public schools that their children attend. Of course this is important, because a family does not want to send its children to a school where the students and teachers are not performing well. With No Child Left Behind, parents that have children in a low performing school now have the option of sending their child to a school that is producing higher test scores. However, can one multiple-choice test really show the public how well the students are performing? And, if a student produces poor results, does that directly reflect on the performance of the teacher? Why has so much importance been placed on these tests? It seems that improved communication between parents, teachers, administrators and politicians could eliminate the need for a "one-shot test" to show what trained professionals could easily explain.

Curriculum Effectiveness

Some school systems also use these tests: to decide on the effectiveness of the curriculum; to decide whether or not to promote students to the next grade; to place students into specialized education programs; and – in high school – to decide if students should receive a diploma (Hamilton, 2002). Again, I go back to the same question: Can this one test really tell us that much about a student? In my opinion, a multiple-choice test does have a function in the classroom. It can be a quick way to assess where the students are at that very moment and where to continue instruction. However, as the final

assessment of students' knowledge, I do not think that multiple-choice tests can be effective, especially when the stakes are so high.

Outside Effects on Test Performance

A multiple-choice test indicates to you what the student knows, or how well the students can take that test, at the very moment in time when the test is given. However, so many variables can affect the performance of the child at that moment. If the child is sick, even with a slight head cold, his or her performance can be thrown off. If the environment in which the test is given is even slightly distracting – too cold, too hot, too bright, too dark, too noisy – the performance is also impeded. Also, everyone that has taken a test for a grade knows the anxiety that can be involved. The children that take these tests know that they are important and most are not comfortable with the format. The tests are different from the format of most classroom instruction and the material is not presented in the way that it was learned. Rarely will a child use a question book and bubble answer sheet at any other time in a classroom unless it is to prepare for one of these standardized tests. This may cause the children much stress, whether or not they are capable of answering the questions correctly. This is true for any one test that is given, multiple-choice or not. These high stakes should not be placed on one single test.

Performance Assessment

An alternative to the standard multiple-choice test is a performance assessment, or authentic assessment, in which the student is required to perform an activity or produce a product. This may include doing an experiment, working on a project, solving a problem or writing a story; often, the students work together for at least a portion of the test. This kind of assessment is connected to the learner's world and includes multiple tasks. It tells

us about how the student connects the knowledge of the material to a given problem. Authentic assessment takes place over time rather than in a "one-shot test", and it therefore gives the teacher, students, and parents continual feedback and more realistically shows the learner's growth. This continual feedback also allows the student to adjust and improve performance, which is not possible with a one-time test (Janesick, 2001).

The essay portion of the ISTEP is one example of performance assessment. The students are asked to apply their knowledge and use their writing skills by responding to a prompt. However, this is still only one day's writing, without any feedback other than a simple score on the student's performance to help them to improve in the future. A rubric or rating scale is often used with performance assessment; it is used with this portion of the ISTEP and is accessible to the student as they work. This tells the students exactly what is expected of them and how their work will be scored. The following ISTEP prompt, and others like it, also include editing checklists and pre-writing questions to help generate ideas for the student (ISTEP+ grade 3 item sampler, 2002):

Read the information. Then do the writing activity.

What do you like to do on a rainy day? Write a real or make-believe story about how you would like to spend a rainy day. Your story will be included in your classroom's book of stories.

Pre-Writing Activity

- *What would you like to do on a rainy day?*
- *Where would you like to go on a rainy day?*
- *Would you like anyone to be with you on a rainy day?*
- *Be sure your story has a beginning, a middle, and an end.*

Editing Checklist

- *Have you started each sentence with a capital letter?*
- *Have you capitalized names of people and places?*
- *Have you ended each sentence with the correct punctuation mark?*

- *Have you spelled all words correctly?*
- *Does the subject of your sentence agree with the action word (verb)?*
- *Examples: Tom plays at the park.*
  They play at the park.
- *Have you written complete sentences?*

In somewhat the opposite fashion of performance assessment, secrecy is oftentimes involved with a multiple-choice standardized test. The questions that will be asked, how the results will be scored, and how the results will be used are often among the topics kept "secret". Unfortunately for the students, this means they do not know how to prepare, how to answer the questions, or what to expect when the results come in (Bracey, 2000).

State Assessment Variety

Every state must assess its students at nearly every grade level. However, the way that the students are tested varies from state to state. What is interesting is to see whether or not the states use norm-referenced tests or other forms of multiple-choice tests, or if they use a performance-based assessment instead.

Arizona

Arizona educators looked at their norm-referenced tests in 1987 to decide which test to use over the next few years. They found that only about 26% of the skills in their state's curriculum framework documents were measured by any of the tests (Kane & Mitchell, 1996). The legislators who used the information from these tests to measure student achievement were appalled. They also found that the tests largely measured lower-level skills, rather than the higher-level skills that they valued. They were worried that teachers were spending so much time working toward the small part of the curriculum framework that was tested. The tests in Arizona were soon changed, and they

required slightly more than just a selection of a correct answer. Currently, based on information obtained from the Arizona Department of Education website, the state tests students partially in multiple-choice format and partially in an extended response piece, or essay question. These tests are referred to as AIMS tests or Arizona's Instrument to Measure Standards. The AIMS test is given to students in grades three through eight. Also given to students in grades two through nine is the Stanford 9 test, which is a multiple-choice, norm-referenced test. One positive feature of the tests in Arizona is that the AIMS tests are not timed. Whereas most standardized tests are given to students within a time limit, the AIMS tests allow students to take as much time as necessary to finish. This should help tremendously in taking some of the pressure off of the students (AIMS update for test coordinators, 2003).

Washington

In Washington, the students take tests under the Washington Assessment of Student Learning (WASL). The standards-based tests include criterion-referenced tests in reading, writing, listening, and mathematics and involve multiple-choice items, short constructed response, and extended constructed response. The students take the WASL in grades four, seven, and ten. In the remaining grades, they must take the Iowa Tests of Basic Skills (ITBS) and the National Assessment of Educational Progress (NAEP), which are both norm-referenced tests. The Office of Superintendent of Public Instruction also provides teachers with assessment tools to be used throughout the year to continually monitor the progress of their students (Assessment: Overview...). This is a state that understands that one assessment tool is not enough to tell you everything about a child's performance; however, the multiple-choice format is still the main factor in their system.

Maryland

In May of 1990, in the state of Maryland, the Maryland School Performance Assessment Program (MSPAP) was first tested statewide (Mitchell, 1992). These tests were designed to simulate classroom instruction and apply the students' knowledge to real life situations through open-ended questions and performance tasks. The students were required to complete writing samples in which they went through the entire writing process and also discussed issues with their classmates before continuing. One portion of the test involved students working together on a hands-on project and writing about the hypotheses and results. These tests were administered over the course of a few days. For the students, the tests were nearly indistinguishable from regular classroom activities. For the teachers, this type of test is worth teaching to. "Teachers improve students' performance on MSPAP by teaching students to analyze what they read, apply skills and knowledge to solve problems, integrate knowledge from different content areas, and work independently and in groups. In this sense, "teaching to" the test is good instruction, the kind of instruction that results in understanding, not in the mere rote recall of isolated facts" (Maryland State Department of Education).

During a previous year's MSPAP math test, third graders were asked to imagine that they were planning a new zoo for their community. They were given the floor plans for three different animal cages and a calculator. Among the various required tasks, the students were asked to find the cost of flooring materials and fencing as well as draw in swimming pools and feed boxes. Lastly, the students were asked to design a repeating pattern for the floor of the information center and explain their drawing. None of the

activities included filling in a bubble and all required the use of problem solving and other skills they had gained from class (Maryland State Department of Education).

Unfortunately, the MSPAP is no longer a part of Maryland's testing program. This state seemed to fully understand that the students should be tested based on actual performances rather than passive answers to questions. The test makers seemed to be at the forefront of the testing movement and leading the rest of the country in the right direction. Why would the state abandon this test? It seems that it did not comply with the No Child Left Behind Act. Maryland was using the results to understand the students as a group and what changes needed to be made in instruction. No Child Left Behind states that each child must be individually tested and compared to other students to ensure that "progress" is being made. Therefore, the emphasis has gone from determining where improvement needs to be made in the schools, to seeing where each students stands in relationship to the rest of the students in the country.

In order to administer a test that does fit the blueprint of No Child Left Behind, Maryland created a new test that seems to take a step backward. The new test is called the Maryland School Assessment (MSA). The MSA tests each student in the form of a criterion-referenced and norm-referenced test. The students are tested using multiple-choice and short answer questions. The students' parents will receive individual results and see how their child compares to students across the nation.

The following is a sample math question from a third grade MSA test that varies greatly from the MSPAP example that was explained previously (Maryland State Department of Education):

4. Some of the letters of the alphabet have lines of symmetry. For which letter of the alphabet can you draw a line of symmetry?

a. K
b. G
c. A
d. S

Teaching to the Test

Needless to say, when a state uses high-stakes, standardized, multiple-choice tests, the teachers want the children to do as well as they can. What often happens is that the teacher spends much of her time teaching the children the format and tricks of these tests – time that would be better spent teaching the curriculum. Often, the administrators are in support of their teachers doing this kind of "teaching to the test". As a matter of fact, in my research I found a book dedicated to the subject. The book actually describes a workshop that teachers can attend that trains them how to best prepare their students for these tests, and it does not include spending more time working on the material that will be covered on the test. Instead, teachers are encouraged to spend eight to ten days, 30-60 minutes each day, teaching the children test-taking skills. The students take practice tests and discuss their feelings, answers, and test-taking strategies. The authors state that they are not "teaching to the test" but rather giving the children the skills to more accurately reveal knowledge they already had. They boast that most students perform better after going through this workshop (Taylor & Walton, 1998). My question, then, is: what does this tell us about the tests that the students are taking? Is the test really measuring the students' knowledge of the content or their ability to take that test? The assessments that we use to gauge students' knowledge should be in an actual situation that the students are familiar with. They should be showing us what they already can do, meaning that they have already done it. They should not have to learn a new procedure to perform well.

The test should come naturally to the students and should ideally not even seem different from regular classroom instruction.

Such is the case with performance-based assessment. In a perfect world, students around the country would be doing projects and hands-on activities, group work and discussions to prove their vast knowledge of the content that they have learned in school. The teachers would share with the children the ways that they would be assessed, so that the students know exactly what to expect. These assessments would come in short sessions over a period of several days or weeks, giving the students feedback and allowing them time to digest the information and improve upon their performance. These assessments would flow with the natural course of the classroom so as not to disrupt the routine, and they would seem to the students like regular classroom activity. Education professionals from across the state would score these tests. Yes, it would take a tremendous effort but it would be worth it to know that the students' progress has been successfully measured. The teachers would be able to see exactly where the students are struggling and know on what areas to work. No longer would they wonder if a child simply answered correctly on a guess or if he missed problems because he mistakenly marked the wrong bubble.

Conclusion

The states in our country have come a long way from the IQ tests used for tracking, and hopefully our system of assessment will continue to progress. Norm-referenced and criterion-referenced tests are not sufficient to measure the ability of our students. State professionals and educators must continue to search, much in the way of

Maryland, for a new kind of assessment that truly assesses a child's performance. Performance-based assessment can be a wonderful addition a state's testing program.

This kind of system is possible even with the passing of No Child Left Behind. Our nation's citizens are worried about their schools and their children and they want to see tangible results; so, we give it to them. Performance assessment seems to be the best way to ensure that we are giving them the desired results, based on what the students are able to do and what they can apply to their real lives. If we ask the students to perform, rather than bubble, we may be able to follow Albert Einstein's advice and count what really counts.

References

AIMS update for test coordinators. (2003, December 12). Retrieved July 15, 2003 from http://www.ade.state.az.us/standards/AIMSUpdates/12-2003.pdf

Assessment: Overview of the Washington state assessment of student learning. (n.d.) Retrieved January 16, 2004 from http://www.k12.wa.us/assessment/WASL/overview.aspx

Bracey, G. W. (2000). Fastback: A short guide to standardized testing. Bloomington, IN: Phi Delta Kappa.

Burns, K. (1999). How to assess authentic learning. Arlington Heights, IL: SkyLight Professional Development.

Chase, R. & Katz, R. (2002). The new public school parent: How to get the best education for your child. New York: Penguin Books.

Cookson, P. W., Jr. & Halberstam, J. (1998). A parent's guide to standardized tests in school. New York: Learning Express.

Frase, L. E. & Streshly, W. (2000). Top 10 myths in education. Lanham, MD: The Scarecrow Press, Inc.

Hamilton, L. S., Stecher, B. M., & Klein, S. P. (Eds.) (2002). Making sense of test-based accountability in education. Pittsburgh: Rand.

ISTEP+ grade 3 item sampler. (2002). Retrieved February 6, 2004 from http://www.doe.state.in.us/istep/pdf/41357-WEB_03_Sampler_02IN.pdf

ISTEP+ program manual 2002-2003. (n.d.) Retrieved July 15, 2003 from http://doe.state.in.us/publications/pdf_istep/progman2002v1.pdf

Janesick, V. J. (2001). The assessment debate: A reference handbook. Santa Barbara: ABC-CLIO, Inc.

Kane, M. B. & Mitchell, R. (Eds.) (1996). Implementing performance assessment: Promises, problems, and challenges. Mahwah, NJ: Lawrence Erlbaum Associates.

Lynn, R. (2004, January 30). Panel votes to leave ed plan behind. Retrieved from The Salt Lake City Tribune Web site: http://www.sltrib.com/2004/Jan/01302004/utah/133940.asp

Maryland State Department of Education. (n.d.) Retrieved February 6, 2004 from http://marylandpublicschools.org/

Mitchell, R. (1992). Testing for learning: How new approaches to evaluation can improve American schools. New York: The Free Press.

Taylor, K. & Walton, S. (1998). Children at the center: A workshop approach to standardized test preparation, K-8. Portsmouth, NH: Heineman.

Testing: Frequently asked questions. (n.d.) Retrieved February 6, 2004 from U.S. Department of Education Web site: http://www.ed.gov/nclb/accountability/ayp/testing-faq.html