International Education and Culture Studies
ISSN 2766-6778
Vol. 6, No. 2, 2026
Published by STSL Press

Original Paper

The Impact of Overtesting on Secondary Students

Erin Campbell

Eastern New Mexico University, Portales, New Mexico

Abstract

Teachers are told by districts and state officials alike that *High-Stakes testing* is the best way to assess a student's present levels of knowledge. This belief has a stronghold in the system that test after test is being added to the mandated yearly or tri-yearly load. There comes a point where we need to step back and look at what is expected of these children. Is the current testing protocol causing more harm than good? This review will examine whether the scores are truly reflective of the student's present levels or are not reliable as a diagnostic tool.

The purpose of this review is to address the possible impact that overtesting is having on the well-being of secondary-level students and their education based on their test scores. If the amount of testing currently required on the state and district levels is putting an unnecessary burden on the students, the data would be unreliable, and therefore, all classroom instruction based on those numbers would not meet the desired educational outcomes.

Statement of the Problem

The statement that overtesting is a problem in today's classroom is no surprise to any current or retired teachers. The amount of assessments mandated at the state and district levels creates an environment where curriculum takes a backseat to assessments for days or, in some cases, even weeks at a time. The data obtained during these assessments is meant to drive instruction and is even considered when assessing the teacher. However, can the data be trusted?

In her article on state testing, Lynn Olson (2020) discusses the impact of overtesting and the lack of consistent data. She also points out that constant shifting between platforms adds to the problem of inconsistent data. This author has seen this both in my previous and my current district; the students are unfamiliar with the platform and therefore do not perform well. This is just one of the many factors that make the assessments unreliable. As the Cambridge Centre for Evaluation and Monitoring (2024) points out, the practice of teaching to the test with unreliable data is a waste of valuable instructional time.

As districts and school administrators rely more and more on the scores obtained during standardized testing as a reflection of the quality of education their students receive, it is becoming increasingly important that the data is as accurate as possible. The fact that instruction is paired so closely with the results is the reason why research on this topic is so important. Instructional time is limited as it is, and as teachers, we must be able to focus on the areas that need the most attention.

A compelling case for formative assessments is given by Ozan and Kincal (2018) in their study of the effects on comprehension of Social Studies skills after an increase in such assessments. This study's significant change between the control group and the formative assessment group shows that there is a correlation that is worth exploring. Amy Grincewicz's thesis (2008) comparing two groups of college students, where one was tested weekly and the other received only three exams, shows similar results.

Definitions

Formative Assessment- Ongoing cycle of assessments and evaluation of learning. Usually, quizzes are used along the way to assess understanding of a subject before moving on.

High-Stakes Testing- Any test where there are significant outcomes based on test performance. These tests include state tests, placement tests, or those required for graduation.

Standardized testing- A test where the questions are uniform across all students at that grade level. This could be at a state or national level.

Summative Assessment- A test given at the end of a large chunk of learning, such as a midterm, final, or end-of-unit assessment.

Several academic journal articles address standardized testing as the best way to assess where students are academically, as well as pointing out where reteaching is needed. These studies fail to take into account the individual student's feelings of anxiety, boredom, or general desire to finish quickly. This lack of looking at the students as individuals has led those in charge to increase the workload on the students in hopes of obtaining better data and therefore a deeper understanding of the child's true needs. This overload is causing the opposite effect when you look closely at the data coming out of these tests.

The Impact of Overtesting on Secondary Students

As teachers, we walk a fine line between pinpointing our students' needs through assessments and testing to the point of skewing the data we rely so heavily on. There will always be the outlier data of those students who didn't really test well the first time, and therefore, the data takes a huge jump, as well as the flipside of that, where the student tanks from one assessment period to another. We must ask ourselves which data point was closest to those students' actual level in order to address their individual needs. This is a tough enough task without adding another layer to this endless spiral, and that layer is more testing. We are told it will help triangulate the data, or that more data helps average it all out, but it looks very different from the perspective of the students. They are the ones who have to sit in session after session, sometimes for weeks at a time. What would you do if you were them? Would you try just as hard on test one as you did on test four?

The literature focuses on addressing different aspects of this question. For the purposes of this review, we will focus on three main categories: the use of classroom assessments to measure success, emotions associated with testing, and the reliability of the tests. We need to look at all of these different moving parts to get a better picture of the minds of the adolescents taking these tests and what is best for them. This author had an old high-school teacher whose name has long been lost to me, who would repeat the phrase *More isn't always better, sometimes it's just more*. That teacher was referring to meeting minimum word counts by writing in circles; however, the same applies here: Is more testing always better, or sometimes is it just more?

The Use of Classroom Assessments to Measure Success

The first study on the category of classroom assessments as a measure of present levels is by Hussain Alkharusi, entitled "Effects of Classroom Assessment Practices on Students' Achievement Goals". This study set out to do exactly what the title suggests: look at the effects of classroom assessments on students' academic achievement goals. The author did this by looking at 1,636 ninth-grade students and 83 science teachers in a public school in Oman. Each class consisted of anywhere from 14 to 21 students, with an average of 20. The gender of the students was 735 males to 901 females. The teachers had an average experience of 5.2 years, with experience ranging from 1 to 13.5 years. Both students and teachers were given a quantitative closed-answer survey. The student survey focused on the assessment environment, personal goals, and self-advocacy, while the teacher survey focused on the frequency of assessments, whether they used any alternative assessments, and their general assessment practices. The study found that the environment plays a large part in the students' perception of the test. In an environment where their test scores and achievements are up for public viewing, students are more motivated to perform for their peers. The most interesting finding is that in classes where the questions are mostly closed-ended, students are more encouraged toward performance avoidance.

This shows that if the assessment is mostly closed-ended, the chances are high that the student will be disengaged. This study supports the hypothesis that the closed-ended, heavy, state testing may be causing a skewed result by disengaging the students, and therefore, they would be more likely to rapidly guess or avoid answering at all. One quote from this study, which was particularly interesting, "From a sociocognitive perspective, students are not social isolates of the influence of those around them…findings showed that the shared perceptions of class members about the assessment environment might influence students' adoption of achievement goals in ways that are consistent with the class shared

perception" (Alkharusi, 2008, pp. 263).

This tells us something that we already know, however, with a unique twist. Students' perception of the task at hand, in this case, testing, is greatly influenced by the feelings of their peers. Therefore, if their peers do not see the value in the testing and choose to take measures to get through it with as little effort as possible, then all or at least most of the students will follow along. This is true in its opposite as well. The teacher's perception of the importance was not mentioned as a factor.

The next study in the category of classroom assessments for success is "The Effects of Formative Assessment on Academic Achievement, Attitudes Toward the Lesson, and Self-regulation Skills" by Ceyhun Ozan and Remzi Kincal. The purpose was to determine the effects of frequent formative assessments on student achievement. It was done on 45 fifth-grade Social Studies students over a 28-week period in 2014-2015. The students were divided into two groups, where one group received significantly more formative assessments than the control group. These assessments were assessed in the normal classwork and were used as an assessment of comprehension of the coursework. The assessments were evaluated, and instruction was altered immediately depending on the results.

The research was a mixed-method study where interviews were combined with the data collected through the assessments. The author looked not only at the data but also at the students' attitudes toward the coursework and self-regulation within the classroom as compared to the control group. The findings of this study were that the students in the assessment group had an overall better experience than those in the control group. They had an increase in their attitude towards the content as well as their academic achievement in the class. The instructor also found that it was easier to address misunderstandings with more frequent assessments rather than waiting till the end of the unit.

This fits by supporting more frequent, smaller assessments of understanding rather than relying so heavily on three times a year larger assessments to measure student growth and comprehension. Ozan and Kincal's conclusion, "As a result of the research, it was determined that the experimental group in which the formative assessment practices were performed had significantly higher academic achievement levels and better attitudes toward the class than the students did in the control group" (Ozan & Kincal, 2018) reveals that test frequency has a positive effect on attitude when it is a classroom assessment.

The final study in this category is Amy Marie Grincewicz's study, "The Impact of Frequent Testing on Student Achievement in an Introductory Level General Education Course," which correlates directly with the difference between classroom assessments vs three times-a-year state assessments. The purpose of which was to assess the effect of weekly exams versus unit exams on student success. This study was done on undergraduate students enrolled in two introductory Biology courses (603 and 654 students), where one section received weekly quizzes, and the other received three assessments. The courses were identical except for the frequency of the testing, including the instructor and course materials. The quizzes covered the same material as the larger tests; however, the questions were not the same to keep the students from sharing answers between the two groups and skewing the results. The results were that the quiz group scored about 3% higher than the assessment group. The researcher compared several factors between the two groups, including GPA and SAT scores, to see if these factors had any impact on the overall testing differences, but found the two groups to be very similar. Thus showing that there is a correlation between more frequent testing and higher scores. This also supports the reliance on classroom assessments and grades as a measure of competency rather than the three major assessments.

Emotions Associated with Testing

The next aspect of testing that needs to be considered is the individual feelings of both the students and the teaching staff who are surrounded by such high-stakes testing. Some tests have more weight than others; for example, the SAT is required for graduation as well as entrance into several colleges. The pressure surrounding taking such a test can and does prove to be one of the factors in the success seen in such testing. Three studies that look at testing from a social/ emotional standpoint look at the anxiety surrounding testing as well as test boredom. Considering these factors is important when considering whether it is worth it to increase the testing and, therefore, increase the prevalence of such occurrences.

Erum Aslam Khan and others discuss anxiety in their study, "Influence of Test Anxiety on Students' Academic Achievement at Secondary Level". Which looked at the effects of test anxiety on academic

achievement. The participants were secondary students at 8 private schools and 4 public schools who were in their 10th year, had completed and passed their previous year's exam in 2028. Both males and females were chosen at random, with 187 students chosen in all. It mentions that the students were a part of the Multan district, but does not mention a specific country; looking up that district resulted in the country most likely being Pakistan. The researchers conducted their research by using a quantitative survey with 34 closed-ended questions related to anxiety felt before, during, and after testing. The study revealed that there was significant anxiety felt by the students regarding the fear of failing and competition with their classmates. This anxiety was the same for males and females.

The study focused on a significant assessment for the children of that area, and the anxiety shown is understandable, given that it is a requirement to move on to the next grade. It is assessments like this that cause stress, which can potentially skew the scores. Anxiety is a universal issue when it comes to high-stakes testing, regardless of location, as this author shows.

Anxiety is also studied by John Jerrim, Rebecca Allen, and Sam Sims in their paper, “High Stakes Assessments in Primary Schools and Teachers’ Anxiety About Work”.

Their study focuses on the impact of such testing on the teachers of 10 and 11-year-old (6th-grade) students as they take the assessments needed to move on to their secondary education. The researchers gathered qualitative data from around 1,000 teachers from across England, looking at those who teach 6th grade, as this is the grade where the pressure is at its peak, because the students must pass the test to move on to secondary school. They looked into whether teachers at this level experience more anxiety than other teachers at other grade levels, and if that anxiety is at its peak around testing time. The data was collected using a survey of 3 questions that needed to be completed daily, with additional questions regarding anxiety at 16 separate points during the school year. The data was analyzed to assess the difference between teachers of different grade levels and the proximity to the testing date.

They used that data to create data values on anxiety levels, and that quantitative data was analyzed. The conclusion was that although the teachers did experience increased stress, it was short-lived. The author argued that the impact on the staff is evident, but a stronger correlation exists between the stress and the students. The stress of the staff is less than that of the students and, therefore, not as statistically significant.

The last study in the social/ emotional category addresses test boredom. The paper by Thomas Goetz and others entitled “Test Boredom: a Neglected Emotion” looks not only at the occurrence of test boredom but also at possible causes. Two distinct groups were used for this study: 208 Eighth graders and 1,612 fifth to tenth graders. A mixed method was used in two stages. Students were given a test with varying difficulty to assess if boredom happens more often when students are under- or over-challenged. Students were then given a survey where they answered questions regarding their boredom with a numerical rating. The answers were assessed for statistical significance. The findings were that students were just as likely to become bored when they were overcharged as they were if they were undercharged; therefore, the causation of the boredom had no correlation with the students’ ability to answer the questions. Test boredom was found to have an impact on the scores of those students with those children who were overcharged. The implications discussed by the authors included increasing the importance of the test in the mind of the student and decreasing the amount of boredom by increasing focus. Also, in the future, a study should look at the differences between boredom in high-stakes testing situations since this study looks at a low-stakes environment. This is precisely what my study looks to do.

The Reliability of the Tests

The final category looks at the reliability of the tests, the frequency of testing, and their benefit to student learning. The first study in this category is “Same tests, same results: Multi-year correlations of ESSA-mandated standardized tests in Texas and Nebraska” by Norman Gibbs, Margarita Pivovarova, and David Berliner. This study analyzed 10 years of state assessment data in Texas and Nebraska to look for correlations in the test results from year to year. The analysis of this quantitative data showed that little changed from year to year, suggesting that frequent testing yields no more benefit than infrequent testing would. This data comes from the analysis of the end-of-the-year data for reading and math on 7,920 elementary and middle school students in Texas and Nebraska from 2010 to 2019.

They found that there was so little change in the data that there was nothing new to be gained from frequent testing, and made the case for testing every second or even third year, and extrapolating the data would yield the same data.

This study supports the point that the amount of testing required of students does not yield enough significant data to make it worth the time it takes to conduct.

By decreasing the testing to every third year, we would be gaining more valuable instructional time. Their conclusion of "reduction in the frequency of testing would simultaneously reduce teacher anxiety and provide additional time for improving teaching and learning—the very process about which accountability measures are most concerned" (Gibbs, N. P., Pivovarova, M., & Berliner, D. C., 2023) is directly in line with my hypothesis.

The next study is "High-Stakes Testing, Uncertainty and Student Learning" by Audrey Amrein and David Berliner, and it analyzes the high-stakes testing data of 18 states to determine if their testing was affecting student performance and learning. The tests used were the ACT, SAT, NAEP, and AP exams. Smaller state-specific tests were not used due to the ability to alter that data through focused test preparation. This quantitative study looked at the data for each test separately across all states and reported the results test-by-test. The researchers wanted to know if the scores they were seeing on the smaller state testing were representative of learning or not. Using the larger standardized testing and analysis of those tests as compared to the state-specific testing, the researchers were able to get a better picture of the present levels of the students as represented in the two categories.

The findings were that, assuming the ACT, SAT, NAEP, and AP exams were representative of the actual levels of the students, then the reported levels seen on state tests were overinflated and did not, in fact, represent actual knowledge growth on the part of the students studied. This study brings the validity of state tests under scrutiny and thus brings into question whether or not such testing is even warranted for purposes other than to overinflate each state's testing scores and therefore their reported student achievement. The authors' conclusion goes even deeper than that and calls the entire system and the policies associated with it into question, the ethics of such testing, "high-stakes testing policy is more than a benign error in political judgment.

It is an error in policy that results in structural and institutional mechanisms that discriminate against all of America's poor and many of America's minority students" (Amrein & Berliner, 2002, p. 58).

The next study in this category is "Opportunities and Obstacles to Making Innovation a Priority in Education" by Robert W. Smith & Kayce Anne Smith. This study recognizes that standardized testing is flawed, and the authors analyzed new ways of teaching and measuring learning to see if there is a better way. These methods were tested by 397 teachers who introduced methods that were considered innovative by the teachers and responded via survey with the results. This study serves several purposes. First, it identified those practices that the teachers considered to be innovative, and then it put these methods to the test to see if they truly were a better way to teach and measure learning. After the surveys were analyzed, four were randomly selected for an interview to gain a deeper understanding. The results indicated a high percentage (43%) of the teachers who completed reported that they included programs and instruction in their classroom that they considered to be innovative. This seemed unusual to the authors, as the state where the study had taken place, North Carolina, was big on testing and state testing in particular.

However, it was hypothesized that teachers who were more likely to be involved in teaching innovative programs in their classrooms were more likely to complete the survey in the first place. Although the respondents reported that their admin was supportive of them developing such programs, they reported not receiving any additional funds or time to develop such programs. The true test of whether something should be implemented or even supported is whether it works or not. According to 91% of respondents, they saw some significant change in their classroom achievement as a result of these programs.

This opens the door to a discussion of a transition to more innovative programs and the results from these programs to represent the true measure of students' abilities over state standardized testing.

The final study in this category is by Sharon Nichols, titled "High-Stakes Testing and Student

Achievement: Does Accountability Pressure Increase Student Learning?" It looks at student achievement and testing in 25 states. An accountability and rating symptom was developed by graduate students using the data reported in all 25 states. This quantitative data was then compared and analyzed for pressure and achievement. They used the NAEP standardized test as the measurement indicator and included data for fourth and eighth-grade math and reading as the data points. They found that high-stakes testing pressure had an impact on the math scores slightly because of the increased accountability; however, there was no difference when it came to the reading scores. This led to the conclusion that there was not enough impact overall to say that there was any correlation; therefore, high-stakes testing and pressure do not increase student achievement in any statistically significant way. Such a study would imply that the reliance on such testing may be misplaced if there is no significant increase in student achievement as a result.

High-stakes testing takes a toll on all those involved. It represents the culmination of many hours and even months of preparation on the part of the admin, teachers, and students, but is it worth it? If these tests cause undue stress because the results do not truly represent the student's abilities, it makes us wonder if that effort can be reallocated to something more representative of true academic achievement. According to these sources, state-regulated testing causes teachers to teach to the test, and even then, the data is not as reliable as classroom assessments done regularly. Even when classroom assessments are done with increasing frequency, the results are more reliable within the classroom than they are when left in the hands of the state.

Conclusion

Teachers must consider not only the test results but also the human component of testing. When those taking the test are suffering from anxiety, lack of perception of value, and/or overall test fatigue, those scores are less likely to be reflective of that student's true levels. When there is a discrepancy between the scores on the first and the last test in a cycle, which scores do we trust? Does averaging those scores do the student a disservice by artificially lowering their score? Those questions and more are constantly on the minds of educators during testing season as scores roll in and we see scores well out of range with what we know our students' true abilities are. That is why this close look not only into the scores themselves is so important, but also to look at the human perceptions behind those scores to get the big picture that we so desperately need in order to help our students thrive.

References

Alkharusi, H. (2008). Effects of classroom assessment practices on students' achievement goals. *Educational Assessment*, *13*(4), 243–266. https://doi.org/10.1080/10627190802602509

Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing & student learning. *Education Policy Analysis Archives*, *10*, 18. https://doi.org/10.14507/epaa.v10n18.2002

Cambridge CEM. (2024, October). *Still too much testing in schools?*. Cambridge CEM - Formative Assessments for Schools. https://www.cem.org/blog/too-much-testing-in-schools

Gibbs, N. P., Pivovarova, M., & Berliner, D. C. (2023). Same tests, same results: Multi-year correlations of ESSA-mandated standardized tests in Texas and Nebraska. *Education Policy Analysis Archives*, *31*(10). https://doi.org/10.14507/epaa.31.7696

Goetz, T., Bieleke, M., Yanagida, T., Krannich, M., Roos, A.-L., Frenzel, A. C., Lipnevich, A. A., & Pekrun, R. (2023). Test boredom: Exploring a neglected emotion. *Journal of Educational Psychology*, *115*(7), 911–931. https://doi.org/10.1037/edu0000807

Grincewicz, A. M. (2008). *The impact of frequent testing on student achievement in an introductory level general education course* [Unpublished thesis]. The Ohio State University.

Jerrim, J., Allen, R., & Sims, S. (2024). High-stakes assessments in primary schools and teachers' anxiety about work. *Educational Assessment*, *29*(2), 59–74. https://doi.org/10.1080/10627197.2024.2350961

Khan, E. A., Munir, Dr. H., Afzam, Dr. A., & Ansari, M. M. (2021). Influence of test anxiety on students' academic achievement at the Secondary Level. *İlköğretim Online*, *20*(2), 303–316. https://doi.org/10.17051/ilkonline.2020.02.31

Mertler, C. A. (2020). *Action Research: Improving Schools and Empowering Educators* (Sixth ed.). SAGE Publications.

Nichols, S. L., Glass, G. V., & Berliner, D. C. (2006). High-stakes testing and student achievement: Does Accountability Pressure Increase Student Learning? *Education Policy Analysis Archives*, *14*. https://doi.org/10.14507/epaa.v14n1.2006

Olson, L. (2020). *A shifting landscape for state testing*. The State Education Standard: National Association of State Boards of Education. Next Generation Assessment, 20(3), 7–42.

Ozan, C., & Kincal, R. (2018). The effects of formative assessment on academic achievement, attitudes toward the lesson, and self-regulation skills. *Educational Sciences: Theory & Practice*, *18*(1), 85–118. https://doi.org/10.12738/estp.2018.1.0216

Smith, R., & Smith, K. A. (2020). Opportunities and obstacles to making innovation a priority in education. *Critical Questions in Education*, *11*(2), 167–178. https://doi.org/10.3102/1569234