Comparing Classroom Assessments to District Assessments to Analyze Assessment Guided Instruction Effectiveness

Author: Vanessa Sampogna Western
Faculty Mentor: Jeanne Tunks, Teacher Education and Administration Department, College of Education
Department and College Affiliation: Teacher Education and Administration Department, College of Education
Bio:

Vanessa Sampogna Western graduated in May of 2010 with a degree in Interdisciplinary Studies from the College of Education at the University of North Texas. She also graduated with a Texas Teaching Certificate as an EC-4 Generalist Educator. She received a Hispanic Scholarship Fund Award in 2009. Vanessa presented this research at the University of North Texas’ Scholars Day. She has also been nominated with a fellow educator for an action research award from the PDS special interest group of the American Education Research Association. Vanessa plans to teach in elementary schools serving students from low-socioeconomic backgrounds.
Abstract:
The purpose of our research was to determine the accuracy of our assessments in guiding our instruction. We pursued this by administering tests to one class of fifteen general education students and one of thirteen Spanish bilingual education students upon completing mathematics instructional units. We hypothesized that their scores on these assessments would significantly correlate with District Benchmark scores at the $p < .05$ level. We also believed that mastery observed on weekly post-assessments would also significantly correlate with district benchmark evaluations of mastery at the $p < .05$ level. We discovered that in-class mastery of specific standards did indeed predict mastery of those standards on District assessments. Also, student in-class averages correlated highly with averages on District assessments.
Introduction

Instruction guided by assessment is a current trend in education. Educators use assessments to create, modify, and guide their instruction. However, in my classroom, we wondered if the assessments we were using to guide instruction were actually useful. My mentor teacher and I realized that we did not know if the assessments we used in our classroom were actually useful in preparing our students for high-accountability tests such as District Benchmarks and the state-wide Texas Assessment of Knowledge and Skills (TAKS) test.

District Benchmarks are tests administered to all students in each grade level. These are multiple choice tests similar in format to the TAKS test. In the fourth grade mathematics class, students take a Benchmark every six weeks testing the material that teachers are required to teach according to the District outlines of curriculum. For example, if mathematics teachers in the fourth grade taught multiplication and division during the first six weeks, it would be tested on the first Benchmark. The last mathematics District Benchmark that fourth grade students took this past year was actually a TAKS test that had been administered in a previous year. These state-wide tests are given near the end of the school-year to test all of the mathematic concepts that students are held responsible for depending upon their grade. Tests such as District Benchmarks and the TAKS test are used to hold teachers, schools, and districts accountable for their students’ performance.

So, we raised the question, “Are our in-class assessments comparable to District and State assessments?” We hypothesized that our assessments would be comparable because we usually compiled them from various books with questions in TAKS test
format. We expected that student mastery of specific Texas Essential Knowledge and Skills (TEKS) in class would predict student mastery of those TEKS on the District Benchmarks. We also hypothesized that students’ classroom post-assessment average would predict their District Benchmark average within five percentage points.

**Procedures**

In my departmentalized fourth grade classroom, we taught science and math to two classes of fourth graders. One class contained fifteen general education students and the other contained thirteen bilingual Spanish students. We taught mathematics to both classes using a workshop method similar to guided reading in language arts. At the beginning of each unit (typically one week long), we administered a pre-assessment to both classes. My mentor teacher and I then planned our instruction for the week based on the results from the pre-assessments. Students were grouped by level based on the pre-assessment to complete small group work with the teacher or individual work using the group as a resource. The work was usually based on the weaknesses of the whole class or of the given group. At the conclusion of the unit, we administered a post-assessment to determine if the students had improved in their mastery of the unit concept. The students’ pre- and post-assessment scores were kept on student tracking sheets with teacher notes and copies of student work. We only used the post-assessments to compare student success.

In order to determine if our unit post-assessments were comparable to the District Benchmarks, we specifically compared student success on both. We compared student achievement on both tests in two ways. First, we selected the fifteen units for which we had sufficient data to compare to District Benchmarks. Next, we chose students for whom
we had sufficient data to compare both in-class assessments and District Benchmarks. After narrowing down our testing base to fifteen general education students and eleven bilingual Spanish students, we compared student mastery of individual TEKS on in-class assessments and on District Benchmarks. For example, if a student mastered the TEKS of multiplication of two digits times one digit in class, we looked to see if he or she also mastered it on the District Benchmark. If the student mastered it, we took this as a valid predictor. We defined Mastery in class and on the Benchmark as 70% or more of questions covering that TEKS correct. We also compared students’ averages on the fifteen post-assessments to their averages on the three District Benchmarks administered to date. One of the three District Benchmarks was a released TAKS test. We calculated each student’s average on the fifteen post-assessments and then compared that score to their average of the three District Benchmarks. Lastly, we checked to see if the scores were within 5 percentage points of each other.

Findings

Comparison by Individual TEKS Mastery

Our data showed that, overall, students who mastered a TEKS in class, also mastered the TEKS on a District Benchmark 73% of the time. Students in the general education class mastered TEKS on a District Benchmark 74% of the time, and students in the bilingual class mastered the TEKS 69% of the time. Student TEKS mastery in-class served as a valid predictor for mastery on the District Benchmark.

Comparison by Overall Average

The data showed that overall student averages in class only predicted District Benchmark averages in 31% of students. Thirty-three percent of students’ Benchmark
averages from the general education class were comparable to their in-class averages within five percentage points. Only 27% of students’ classroom assessments average from the bilingual Spanish class predicted District Benchmark average within five percentage points. However, when we compared the overall class averages of in-class assessments to District Benchmarks, there was a very high correlation. In fact, when we compared both classes’ average scores on the Benchmark and in class using the Pearson correlation coefficient, it showed a correlation of 0.811. The class of fifteen general education students (or class G) had a correlation of 0.853 as shown in Figure 1, and the class of eleven bilingual Spanish students (or class B) had a correlation of 0.893 as shown in Figure 2.

Results

From our data, we can conclude that our hypothesis concerning student mastery of individual TEKS was correct. If a student masters a concept in class using our current assessments, then 73% of the time, that student will master it on a high-stakes test such as a District Benchmark or State TAKS test.

We can also conclude from our data that our hypothesis concerning student grade averages was incorrect. It would be fruitless to use a student’s in-class average using the current assessments to predict their average score on a District or State test. However, there are many factors that differentiate weekly in-class assessments from District Benchmarks administered every six-weeks. These contributing factors may have lead to the difference between student post-assessment scores and District Benchmark scores. For example, in-class post-assessments are about ten questions long as opposed to District Benchmarks which are at least 40 questions long. Students that tire of reading or
test-taking quickly may perform well on short tests, but not on long tests such as
Benchmarks. Also, class assessments focus only on one or two TEKS while Benchmarks
cover at least fifteen on each Benchmark.

Another factor to consider is that the accountability level for classroom post-
assessments is not very high compared to the accountability level for which students,
teachers, and the district are held to for District Benchmarks. This pressure sometimes
can affect student view of and performance on the District Benchmarks. Post-assessments
are also taken in the usual, comfortable, classroom environment. Sometimes struggling
students are pulled out of their classroom into a small group or for individual
administration. This change of routine and setting may upset struggling students. Finally,
some students suffer from test anxiety that may appear on large high-stakes tests rather
than on weekly in-class assessments.

The many individual cases within the data bring up several issues. For example,
Student 12G averaged a 68 on the weekly classroom assessments and averaged a 38 on
the District Benchmarks. Based on Student 12G Benchmark scores, the student is
struggling to get to the mastery expectations for the Benchmark, but is very close to
mastery expectations in class. We believe the reason for cases such as this one is
retention. Students like Student 12G grasp concepts in class, but may not remember them
well enough to apply them on the Benchmarks. One solution to this retention problem
may be to spiral curriculum more effectively. Daily practice of skills taught previously
has been shown to help students retain topics more successfully. Perhaps it would be
helpful in our classroom to increase the achievement of those students who struggle to
retain mathematics concepts.
Another concern brought from this data is the general trend of the bilingual Spanish students’ in-class mastery of TEKS that involve word problems not predicting their mastery on the District Benchmarks. We determined that this is most likely because the bilingual Spanish students may struggle more with vocabulary than the general education students. This data will lead us to use more vocabulary intensive instruction in our bilingual Spanish class as well as more practice in word problems.

When analyzing specific units for prediction on the Benchmarks, we discovered that the in-class assessments of the fractions and the measurement units had the lowest percentage of valid predictors. This means that we are not assessing (and therefore not teaching) these topics in the same way that the District is implementing the topics on the Benchmark. Because of this research, we will return to our plans and assessments from these units to find ways to improve our assessments and instructional methods on these topics.

Conclusion

In lieu of this data, we are also planning on pursuing this process with other content areas. For example, in Language Arts, it would also be beneficial to begin a differentiated method based on assessments of the skills required for the District Benchmarks. While we currently differentiate lessons by reading level, we do not always do so on skill level. Developing weekly formative assessments comparable to District Benchmarks for Language Arts may help us create instruction that individual students will benefit from.
Figure 1. The Assessment Comparison of Class G

Class G Comparisons

Scores in Percent

G Class Students

In-class Average
Benchmark Average
Figure 2. Comparisons of Classroom B Averages to Benchmark Averages

Comparisons of Class B Classroom Averages to Benchmark Averages
Suggested Reading

William, D. (2006). Formative assessment: Getting the focus right. *Educational Assessment, 11*(3/4), 283-289. doi:10.1207/s15326977ea1103_7.

Roberts, A. (1984). Group methods? Primary teachers' differentiation policies in mathematics. *Educational Review, 36*(3), 239-48. Retrieved from ERIC database.

Ellis, D., Ellis, K., Huemann, L., & Stolarik, E. (2007). Improving mathematics skills using differentiated instruction with primary and high school students. (Action research paper.) Saint Xavier University, Chicago, IL. Retrieved from ERIC database.

Ferreri, A. (2009). Including Matthew: Assessment-guided differentiated literacy instruction. *Teaching Exceptional Children Plus, 5*(3), 1-11. Retrieved from Education Research Complete database.

Taylor-Cox, J. (2009). *Math Intervention: Building Number Power with Formative Assessments, Differentiation, and Games, Grades 3-5.* Larchmont, NY: Eye on Education.

Heuser, D. (2000). Reworking the workshop for math and science. *Educational Leadership, 58*(1), 34-37. Retrieved from ERIC database.

Schmidt, R. (2009). Assessing our way into instruction: What teachers know and how they know it. In R. Schmidt & P.L. Thomas (Eds.), *Explorations of educational purpose, (Vol. 5), 21st century literacy: If we are scripted, are we literate.* Retrieved from Springerlink database.

Thousand, J., A. Liston, M. McNeil, & a. Nevin. (2006, November). Differentiating instruction: collaborative planning and teaching for universally designed learning. Paper presented at the annual conference of the Teacher Education Division of the Council for Exceptional Children. Retrieved from Google.docs.

Celeste Cusumano; Jonel, M. (n.d). How differentiated instruction helps struggling students. *Leadership, 36*(4), 8. Retrieved from ProQuest Research Library database.

Editor, Carol Anne, Dwyer. (2007). *The Future of Assessment: Shaping, Teaching and Learning.* New York: Taylor and Francis.

Rebecca L Pierce; Cheryll M, A. (n.d). Tiered lessons: One way to differentiate mathematics instruction. *Gifted Child Today, 27*(2), 58. Retrieved from ProQuest Research Library database.