MATHEMATICS TEACHING RESEARCH JOURNAL
VOL 10, N 2
SUMMER 2018



# The Relationship Between Assessment Practices and Hispanic and White Eighth Grade Students' Mathematics Achievement: An Analysis of 2013 NAEP Data

James A. Telese

University of Texas, Rio Grande Valley

Correspondence concerning this manuscript should be addressed to James A. Telese, Department of Teaching and Learning, The University of Texas, Rio Grande Valley, 1 West University Blvd., Brownsville, TX 78520.

E-mail: james.telese@utrgv.edu

***Abstract:*** *This study examined data from the 2013 NAEP eighth grade mathematics assessment. This correlational analysis related to classroom assessment practices, from the teacher survey, with students' mathematics achievement. Multiple regression analyses were also conducted to examine the relationship between the frequency of use various assessment strategies and student mathematics achievement. The results showed that Hispanic students' mathematics performance was at a lower level than non-Hispanic (White) students. Items related to formative assessment strategies were correlated with achievement. There was a negative relationship found when discussing current performance level of students. Both groups of students had lower mean composite scores as the reported frequency of use increased. In addition, a negative relationship was found when teachers reported assessing students by adjusting teaching strategies to meet the needs of students and achievement. The findings suggest that formative assessment strategies have a negative relationship with the Nation's eighth grade mathematics students.*

**Keywords:** NAEP, mathematics, eighth grade, student achievement

There are many views that suggest that the use of assessment strategies impact student achievement. The National Council of Supervisors of Mathematics (2014) and the National Council of Teachers of Mathematics (2014) have issued position papers affirming the use of assessment including formative assessment. Yet, there have been few studies that examine the impact of assessment strategies on middle school students' mathematics achievement. Erasmus (2014) developed an overall assessment system for second language learners, which involved cognitive facets, affective facets and screening tests, but no examination of a particular strategy on mathematics achievement. William, Lee, Harrison, & and Black (2004) studied the impact of teachers developing assessment strategies on student achievement on accountability tests.

The purpose of this study was to examine the relationship between various assessment strategies and Hispanic and White eighth grade mathematics achievement on the 2013 National Assessment of Education Progress (NAEP) test. The research questions were: What is the difference between Hispanic and White students' achievement on the eighth grade NAEP mathematics exam? What is the relationship between assessment practices and performance of Hispanic and White students on the eighth grade NAEP mathematics exam? What is the relationship between formative assessment strategies and performance of Hispanic students and White students on the eighth grade NAEP mathematics exam?

## REVIEW OF RELATED LITERATURE

### Hispanic Education

The educational attainment of United States Hispanics has been changing over the last several years (Krogstad, 2016). This is reflected in a growing number of Hispanics' students in the nations' K-12 Schools. In 2014, it was projected that the public schools in the United States would be majority-minority, with a large percentage being Hispanic (Krogstad & Fry, 2014). Since 1997, the number of Hispanic students has almost doubled to 12.9 million in 2014 According to Krogstad & Fry, most of the growth is due to U.S. born Hispanics.

As noted, the portion of the U.S. Hispanic population is increasing over time. A substantial proportion of this growing Hispanic student population in grades 4 and 8 are English language learners with 37 percent and 21 percent respectively (Hemphill & Vanneman, 2011). Both the growing population of Hispanic students and the larger portion that are English language learners contribute to the achievement gap between Hispanic and White fourth and eighth graders (Hemphill & Vanneman, 2011). The National Center for Education Statistics (NCES) [2015] defines an achievement gap as "when one group of students such as, students grouped by race/ethnicity, or gender outperforms another group, and the difference in average scores for the two groups is statistically significant." Closing an achievement gap can be a challenge especially the one between Hispanic and White students, even though Hispanic students' average scores have

increased across the assessment years, White students had higher scores across all assessments (Hemphill & Vanneman, 2011).

The National Assessment of Educational Progress (NAEP) mathematics exam was first administered to both Hispanic and White public school students in 1990. Hemphill and Fry's (2011) report indicated that on the 2009 NAEP mathematics exam, the scores for both Hispanic and White students in grades 4 and 8 were higher than in 1990. However, from 1990 to 2009, the achievement gap between Hispanic and White students did not change significantly at either grade four or eight. Scores for Hispanic and White fourth-graders remained unchanged from 2007 to 2009, and the gap persisted at 21 points. For eighth-graders, scores increased for both Hispanic and White students from 2007 to 2009, but the gap remained at 26 points, and it was not significantly different from the gap in 1990 or 2007 (Hemphill & Vanneman).

The achievement gap between Hispanic and White students seems to be persistent with little change. Moreover, it is essential to be concerned with Hispanic student achievement in relation to future high school mathematics course taking and in science, technology, engineering, and mathematics (STEM). Hispanic students are underrepresented among students who complete a four-year degree in science, technology, engineering and mathematics (Borman,, et al., 2017; National Center for Education Statistics, 2016). The more advanced high school courses that are taken the greater the likelihood that students will enroll in college and complete a degree (Hinojosa, Rapaport, Jaciw, LiCalsi, & Zacamy, 2016).

Without an adequate foundation in mathematics and by not having access to advanced courses, Hispanic students may be at a disadvantage when seeking STEM related career opportunities (Garland & Rapport, 2018). In 2009 Hispanic employees accounted for 14 percent of the U.S. workforce but held only 6 percent of STEM jobs (Beede et al., 2011). Garland & Rapport (2018) conducted a study in Texas, a state where 51 percent of its students are Hispanic, found that White students completed a slightly greater number of advanced STEM courses. In addition, a greater difference by race/ethnicity was found for the percentage of students who completed three or more advanced STEM courses. They found that among high ability grade 8 students, approximately 52 percent of White students completed three or more advanced mathematics courses during high school compared with 41 percent of Hispanic students. This observation suggests that there is a need to examine mathematical instructional practices that offer the best chance for promoting student achievement and thus advanced course taking in STEM fields (Garland & Rapport).

**Feedback as Formative Assessment**

It is widely believed that formative assessment enlightens students on their strengths and weaknesses (Phelan, Choi, Vendlinski, Baker, & Herman, 2011). Also, it is suggested that

formative assessment fosters large learning gains, especially for low-achieving students (William, Lee, Harrison, & Black, 2004). Controversy exists regarding this issue. In recent meta-analyses that analyzed the impact of formative assessment on student achievement Kingston & Nash (2011) concluded that the use of formative assessment produces a 0.2 median effect size. In response, Briggs, Ruiz-Primo, Furtak, Shepard and Yin (2012) questioned the methodology of the Kingston and Nash study in four areas: (a) its approach, (b) inclusion criteria, (c) biased effect sizes, and (d) the relationship between effect sizes and the magnitude of the outcomes. They concluded that the effect size reported by Kingston and Nash is not accurate. Kingston and Nash's finding suggested that the use of formative assessment strategies may not have as great an impact as indicated by others reporting an effect size ranging from 0.4 to 0.7 (McMillan, Venable & Varier, 2013). Apparently, the promise of formative assessment indicated by some studies may be difficult to fulfill (Phelan et al., 2013).

Formative assessment researchers have widened their scope by examining teacher professional development (Wylie & Lyon, 2015) and special populations. For example, a study conducted by Shayyan, Thurlow, and Liu (2008) on English Language Learners with disabilities attempted to determine effective instructional strategies. They identified 10 strategies upon examination of the literature. Two strategies in particular relate to formative assessment of students: (1) provide feedback that is adapted to the learner's level of language proficiency, and (2) conduct on-going assessment of the effectiveness of instruction. Similarly, students in their study weighted as a highly important instructional strategy the use of random, recurrent assessment. In contrast, educators in the same survey rated highly important the assessment strategy of having students think aloud. These strategies are appropriate for all students. Hudson (2015) suggested that assessment should be on going, dynamic and provide feedback to students.

Furthermore, feedback has also become an area for formative assessment research, which is regarded as an essential component of formative assessment (Hattie & Timperley, 2007; Hattie, 2009). Havnes, Smith, Dysthe, and Ludvigsen (2012) noted that positive effects from the use of feedback do not always occur. They cited a 1996 study by Kluger and DeNisi who found that more than one third of the effects indicated negative impact of feedback on learning. However, Havnes et al contended that in order for feedback to be formative, there is an assumption that the feedback should be provided in a manner that actively engages the learner with the feedback. This suggests that the learner plays a role in how feedback is used and interpreted (Sadler, 2010). Havnes et al examined perceptions of feedback practices of teachers, and how the feedback is used by teachers and students using surveys and focus groups. Their findings suggested that feedback practice is subject related, different content areas have different methods and purposes for feedback, and that weak students are not able to use feedback provided by the teachers. Their other findings suggested that practices related to assessment were found to focus on corrections and grading and that communication about learning processes between the teachers and the students

was viewed as formative assessment. They identified four classroom situations that have potential to be rich in feedback opportunities: working through a test or assignment when corrected and returned to students, student presentations of group projects, group-work, and discussions between the teacher and the student. Feedback is an integral aspect of the classroom, but the era of accountability has affected assessment practices.

**Toward Formative Assessment**

Accountability has emphasized the use of standardized tests often in the form of multiple-choice items for use in high-stakes testing (Kantrov, 2000). Teachers use multiple-choice and similar tests in the classroom (Kantrov, 2000). Assessment is considered any systematic procedure for collecting information that can be used to make inferences about students (American Education Research Association, et al, 1999). High-stakes test results are often used to make decisions such as whether or not a student graduates. Important decisions should not be based on a single test or other assessment procedures (Reynolds, Livingston, & Willson, 2009). Assessment is a tool that provides evidence and feedback concerning what students know, a way to communicate the value of knowing important concepts, and determining program effectiveness. Problems persist with the current accountability system (Koretz, 2008). Recognizing these limitations and current assumptions about learning and knowing has led to tensions with newer learning and assessment paradigms (e.g., Hickey & Anderson, 2007; Shepard, 2008; Stiggins, 2002) which have motivated educational researchers to rethink the role of assessment to one where it supports and documents classroom learning.

Assessment of students' capabilities in mathematics begins in the classroom with the teacher. Moreover, assessment as a broad, general term evolves into formative assessment when teachers and students conduct activities that provide information to be used as feedback to modify teaching and learning activities in order to best meet their needs (Black & William, 1998; Havnes, Smith, Dysthe, & Ludvigsen, 2012). The importance of assessment within the classroom is being recognized, and there is a shift from the 'testing culture' to an 'assessment culture' (Filsecker & Kerres, 2012). Shepard (2000) saw the shift emerging as part of the learning process rather than external tests administered at the end of curricular units. Consequently, the assessment paradigm has moved toward classroom assessment, which is characterized as constructivist and learner centered (Filsecker & Kerres, 2012).

A working group called Demos (2004) examined the issue of learning. The group contended that those teachers who constantly monitor their students, as they set about learning and reflect on actions that do not go as expected, were the better teachers. This is otherwise known as assessment for learning that results in an assessment-centered design for a learning environment (Demos,

2004). This points to the notion that formative assessment is a process that occurs in the moment-to-moment interactions between the teacher and students (Filsecker & Kerres, 2012).

**Assessment in the Mathematics Classroom**

The National Council of Teachers of Mathematics [NCTM] (2014) listed assessment as one of six principles to ensure success for all students, and it was suggested that assessment should support the learning of important mathematics and furnish useful information to both teachers and students. Assessment within an excellent mathematics program ensures that assessment is an integral aspect of instruction providing evidence of what students know about important mathematics content, and includes a variety of data sources from which important decisions can be made concerning students, instruction, and the program (NCTM, 2014). Likewise, Hudson (2015) and Fennel, Swartz, McCord, Kobett, and Wray (2015) contended that formative assessment should be on going, dynamic, and provide detailed and useful feedback. Formative assessment has taken on a more important role in providing this information, and it is viewed as a way to bring about changes in the classroom (Black & William, 2009). It is defined as the extent that actions are taken in the classroom to elicit evidence of student achievement, interpretation of the evidence used by teachers, learners, or their peers to make decisions about the next steps in instruction that would be better founded if the evidence had not been used to make such decisions (Black & William). More succinctly, formative assessment is to inform instruction and provide feedback to students on their learning (Filsecker & Kerres, 2012; & Keeley & Tobey, 2011).

Wylie & Lyon (2015) disaggregated Leahy, Lyon, Thompson, and William's (2005) definition of formative assessment and identified strategies that support the use of assessment by teachers that include clarifying and sharing learning intentions and criteria for success, orchestrating effective discussions, questions, and learning tasks, and providing feedback. Wylie and Lyon (2015) conducted a professional development program, for 200 mathematics and science teachers, in an attempt to evaluate how they implemented the strategies in their classrooms. The program had an introductory two-day workshop for teachers, a follow-up two-day workshop for teacher leaders, and monthly learning community meetings for two years. Participants were asked to complete on-line surveys, supply a daily log indicating which of the 74 assessment strategies where used, and a reflection of implementing one technique with a detailed description of its implementation. Wylie and Lyon (2015) found that the teachers reported a similar frequency of communicating expectations, the use of classroom questions, and tasks prior to the training and two years after the training occurred. The study revealed a change in their teachers' questioning strategies from calling on students with hands raised to a more random approach. There were positive changes in how teachers issued feedback, from commenting only without a grade to mastery grading, the expectation that work will be revised based upon feedback until it reaches an acceptable standard. However, on the reflection piece, they noted that the teachers reported communicating learning

expectations more often than criteria for success and an increased use of collecting student evidence of learning. This study examined the frequency of use of formative assessment strategies following a professional development program. A weakness in the study is that student achievement was not examined relative to the frequency of use of the five strategies. Mathematics teachers often use performance-based, embedded assessments such as problem sets, writing in journals, providing immediate or delayed feedback from the use of summative tests, and questioning (Keely & Tobey, 2011). Consequently, the teacher in an assessment-centered classroom must continuously note how to meet the learning needs of students and build a bridge between their initial ideas and mathematics instructional goals (Keely & Tobey; William, 2011).

The above discussion centers on describing assessment and formative assessment. Good (2011) contended that formative assessment has long been described as a type of assessment, and believes that it is a process; at issue is the timing regardless of the quality of item or its connection to instruction. An example was provided to assess student understanding of the order of operations: $3^2 + 2 \times 4$ which can be used as either a summative, if administered at the end of a unit or formative item if used to identify misconceptions during instruction (Good, 2011). So, assessment items, tasks, or activities like mathematics exercises may be considered formative or summative depending on how they are used.

In the current study, the term assessment is being used to encompass formative assessment strategies. In summary, Dunn & Mulvenon (2009) concluded that there have been few studies that have examined the impact of assessment strategies on student achievement. There have been some studies such as Black and William's (1998) study concluding that formative assessment does improve learning. Others such as Bennett (2011), Filsecker & Kerres, (2012), and Kingston and Nash (2011) have addressed generally, the degree to which formative assessment affects achievement (McMillan, Venable & Varier, 2013). The meta-analysis conducted by Kingston and Nash investigated the relationship and found a median effect size of 0.20, which is lower than the range cited often of 0.4 to 0.7 (McMillan et al., 2013). McMillan et al critiqued Kingston and Nash's study on three grounds: (1) the quality of the studies used, (2) selection criteria for the studies used in the meta-analysis, and (3) and the nature of the formative assessment strategies used. Others have examined the impact of professional development such as, Wylie & Lyon (2015) who examined the fidelity of the implementation of a professional development program for mathematics and science teachers in formative assessment and found that the use of feedback, and its quality improved as a result training in formative assessment strategies. Phelan et al. (2011) conducted a randomized study that examined the implementation of sixth grade formative assessment tasks in various mathematics domains. They found that, overall, the treatment students did not outperform control students, and the students who scored higher on the pretest benefited more from the intervention compared to students who scored lower on the pretest. This result means that higher performing students benefited the most from implementing formative

assessments. The intervention did not help lower performing students develop greater understanding of the mathematics concepts. Given that the Phelan et al. study is an example of determining the impact of formative assessment on student mathematics understanding; there are relatively few studies that have examined the relationship between formative assessment strategies and middle school mathematics students' achievement. The current study aimed to determine the relationship between formative assessment strategies and student achievement.

## METHODS AND PROCEDURES

### Instrument

The 2013 NAEP National Public School database was the data source for the analysis. The sample consisted of eighth grade students. Teachers of students who were administered the assessment were asked to complete a questionnaire. The survey contains items related to classroom instruction and practices. The independent variables were selected from the teacher reported data regarding the modes of instruction/classroom activities, in particular those that were considered assessment strategies and/or formative assessment techniques. These included: Approximately how much mathematics homework do you assign to students in your mathematics class each day? The responses for this item were None, Less Than an Hour, About 1 hour, About 2 – 3 hours, and More Than 3 hours. Teachers were to respond to items by describing the frequency of use for particular strategies, such as:

- Assess math students by discussing current performance level,
- Assess math students by adjusting teaching strategies to meet needs,
- Assess math students by discussing progress toward goals,
- Assess math students by setting goals for specific progress,
- Assess math with individual or group projects,
- Assess math with multiple-choice tests,
- Assess math with problem sets,
- Assess math with short or long written responses.

For these items the response choices are (1) Never or Hardly Ever, (2) A Few Times a Year, (3) Once or Twice a Month, (4) Once or Twice a Week, (5) Every Day or Almost Every Day. The response options were different for the item, "Amount of math homework assigned per day." They were (1) None, (2) Less than One Hour, (3) About One Hour, (4) About two to three Hours, (4) More than Three Hours. Significance tests and Multiple regression analyses were conducted using

the questionnaire items as independent variables and the Composite Mathematics Plausible Value Score as the dependent variable. The analysis focused on Hispanic and White student ethnic groups.

**NAEP Sampling Procedures**

NAEP (2011) uses complex sampling procedures to limit the error and bias in sampling. Students were randomly selected from schools. Sampling strata for schools are developed from groups like school, Region, Urbanicity, Minority composition and Enrollment. The schools are sampled with probability proportional to size of enrollment. It is noted that each stratum can have a different sampling rate. Sampling weights are then computed to counter-act the unequal probabilities of selection and ensure unbiased estimates for the Nation. In order to account for unequal probabilities of selection, sampling weights were used in the analysis. Variance estimates related to the Clusters are combined to account for cluster dependencies from separate estimates from within clusters. Jacknife procedures are used to handle clustering to create replicate weights. Taylor series is used to obtain an approximation to some nonlinear estimating function that is applied to the Primary Sampling Unit with the stratum; thus creating a final weight, Strata, and cluster identifiers to combine estimates of means and variances within each cluster. Plausible values are created and used in the analysis of scores.

**Results**

The overall performance of the nation's middle school Hispanic students was lower than White students and statistically significant by 23.5 points. Table 1 below presents the mean Composite Scores for items identified as formative feedback techniques. Hispanic students scored lower than White students in each category. It appears that for Hispanic students, when discussing progress toward goals every day or almost every day, the scores were on average 34.6 points lower than not at all in comparison to White students. White students had nearly 9 points higher when their progress was discussed nearly each day compared to not at all. These were statistically significant results.

When setting goals, those teachers who reported Never or Hardly Ever, Hispanic students scored statistically significantly lower than White students 275 and 298 respectively. There were decreases in the means for each category and for each ethnic group. For Hispanic students, their mean was 263 when teachers reported discussing progress nearly each day, a 12-point difference from the category Never to the category Hardly Ever. For White students, the result is similar but with a drop of 8 points.

The pattern continues for assessing math students by discussing their current performance level. Both groups had higher means for the Never or Hardly Ever category and the mean performance tended to decrease within each of the subsequent categories. For Hispanic students

the means fell from 274 in Never to Hardly Ever to 269 in Everyday or Almost Everyday, and for White students the means ranged from 301 to 290 in the same category. The differences in performance when teachers reported assessing math students by adjusting teaching strategies to meet needs were less pronounced. In other words, the drop in scores from Never or Hardly Ever to Everyday or Almost Everyday was smaller. However, Hispanic students still performed at a lower level than White Students. The difference was 6 points for Hispanic students and 5 points for White students.

Table 1

*Mean Mathematics Composite Scores and Standard Errors for Items Related to Feedback*

| Item | Never or Hardly Ever | Few Times a Year | Once or Twice a month | Once or Twice a Week | Everyday or almost Everyday |
|---|---|---|---|---|---|
| Assess math students by discussing progress toward goals. | Hispanic 274 (1.6)<br>White 299 (0.6) | Hispanic 273 (1.6)<br>White 296 (0.6) | Hispanic 274 (0.9)<br>White 294 (0.4) | Hispanic 271 (0.8)<br>White 290 (0.5) | Hispanic 264 (3.1)<br>White 290 (1.7) |
| Assess math students by setting goals for specific progress. | Hispanic 275 (1.9)<br>White 298 (0.6) | Hispanic 273 (0.7)<br>White 296 (0.64 | Hispanic 274 (0.7)<br>White 294 (0.5) | Hispanic 270 (1.3)<br>White 291 (0.8) | Hispanic 263 (3.1)<br>White 290 (1.4) |
| Assess math students by discussing current performance level. | Hispanic 274 (4.6)<br>White 301 (1.8) | Hispanic 273 (1.0)<br>White 297 (0.4) | Hispanic 272 (0.7)<br>White 295 (0.4) | Hispanic 270 (0.9)<br>White 293 (0.5) | Hispanic 269 (1.8)<br>White 290 (1.0) |
| Assess math students by adjusting teaching strategies to meet needs. | Hispanic 276 (2.9)<br>White 299 (1.1) | Hispanic 273 (1.2)<br>White 296 (0.7) | Hispanic 272 (0.8)<br>White 296 (0.6) | Hispanic 272 (0.7)<br>White 292 (0.5) | Hispanic 270 (1.0)<br>White 294 (0.5) |

Table 2 presents Mean Composite Scores for items that may or may not be formative assessment depending upon the timing. However, these are considered by the researcher to be assessment strategies that include performance and objective tasks. The achievement gap persists for these items. For Hispanic students, the gap between when teachers reported assessing math with individual or group projects was smaller than for White students, with 4 points and 9 points respectively. Hispanic students had a higher mean of 274 than any other category when teachers reported doing this One to Two times per year. The result was similar for White students. Students

whose teachers reported rarely using multiple-choice tests one to two times per year to assess them had higher achievement, 299 for White students and 276 for Hispanic students. Teachers who reported using multiple choice tests one to 2 times per week had students with lower scores compared to Never or Hardly Ever using multiple-choice tests. The use of problem sets was found not to show significant differences between the frequencies of use.

Hispanic students whose teachers who reported using short or long written responses for assessing math had similar scores in each of the frequency categories ranging from 270

for Never or Hardly Ever to 272 for 1 to 2 times a Month and 1 to 2 times a week. White students had a similar pattern going from 292 to 295 at 1 to 2 times a year.

Table 2

*Mean Mathematics Composite Scores and Standard Errors for Items Related to Assessment Activities*

| Item | Never or hardly ever | 1 to 2 times a Year | 1 to 2 times a Month | 1 to 2 times a Week |
|---|---|---|---|---|
| Assess math with individual or group projects. | Hispanic 270 (1.0)<br>White 295 (0.6) | Hispanic 274 (0.7)<br>White 296 (0.4) | Hispanic 270 (1.0)<br>White 294 (0.6) | Hispanic 266 (2.6)<br>White 286 (1.8) |
| Assess math with multiple-choice tests. | Hispanic 272 (1.4)<br>White 298 (0.6) | Hispanic 276 (1.4)<br>White 299 (0.5) | Hispanic 271 (0.6)<br>White 293 (0.4) | Hispanic 269 (1.3)<br>White 289(1.0) |
| Assess math with problem sets. | Hispanic 272 (2.3)<br>White 295 (1.4) | Hispanic 270 (2.5)<br>White 295 (1.5) | Hispanic 270 (0.8)<br>White 295 (0.5) | Hispanic 273 (0.6)<br>White 295 (0.3) |
| Assess math with short or long written responses. | Hispanic 270 (1.9)<br>White 292 (1.1) | Hispanic 271 (1.1)<br>White 295 (0.8) | Hispanic 272 (0.7)<br>White 296 (0.4) | Hispanic 272 (0.7)<br>White 295 (0.5) |

Table 3 presents the results for the use of homework. This was selected because traditionally homework can be used as a form of assessment, whether summative or formative. The results are similar in relation to the achievement gap. However, in the category About One Hour of Homework, the mean scores were greater than the other categories regardless of ethnicity.

For Hispanic students the mean was 275 and for White students the mean was 301. Compare these scores to the scores whose teachers reported 'None,' 260 and 273 for Hispanic and White students respectively.

Table 3

*Mean Mathematics Composite Scores and Standard Errors for Frequency of Homework*

| Item | None | Less than One Hour | About One Hour | About 2 to 3 Hours | More than 3 Hours |
|---|---|---|---|---|---|
| Amount of math homework assigned per day. | Hispanic 260 (3.1)<br>White 273 (1.6) | Hispanic 271 (0.4)<br>White 295 (0.3) | Hispanic 275 (1.2)<br>White 301 (0.7) | Hispanic 271 (6.7)<br>White 292 (2.8) | Hispanic 270 (9.2)<br>White missing |

Table 4 presents the multiple regression results for each of the items from the teacher survey related to assessment as selected by the researcher. There was a negative association between adjusting instruction to meet the needs of students and student achievement; the greater the reported frequency, the lower the achievement of students whose teachers were administered the survey. A negative relationship was found for assessing math students by setting goals, discussing current performance level, and discussing progress toward learning goals with student achievement. It appears, taken as a whole, the strategies that may be considered formative assessment strategies such as those that provide feedback to students have a negative relationship with achievement in mathematics.

Assessing students with individual or group projects was found to have a positive relationship with achievement when the use is 1 to 2 times a year. There is a negative relationship for the use of projects beginning at one to two times a month with a regression coefficient of -1.01 and -7.10 for 1 to two times a week. A similar result occurred for the use of multiple-choice tests. When teachers reported using them once or twice a year there was a positive relationship with achievement, but negatively associated with achievement at the frequency of one to two times a month and one to two times a week. Assessing math with short or long written responses at least one or two times a month was positively related to achievement, where there was a regression coefficient of 2.76, and suggested that students whose teachers reported providing this opportunity scored nearly three points higher when compared to students whose teachers reported Never or Hardly Ever requiring long or short written responses.

Table 4

*Multiple Regression Analyses for Teacher Survey Formative Assessment Items Related to Student Achievement*

| Comparison code: Never or Hardly Ever | Intercept | Regression Coefficients | S. E. of Regression Coefficients | P-Value | R-Squared |
|---|---|---|---|---|---|
| Assess math students by adjusting teaching strategies to meet needs. | 299.84 | | | | 0.141 |
| A few times a year | | -3.15 | 1.35 | 0.023 | |
| Once or twice a month | | -4.32 | 1.29 | 0.001 | |
| Once or twice a week | | -6.90 | 1.28 | < 0.001 | |
| Every day or almost every day | | -6.62 | 1.36 | < 0.001 | |
| Assess math students by setting goals for specific progress. | 298.33 | | | | 0.412 |
| A few times a year | | -2.25 | 0.83 | 0.008 | |
| Once or twice a month | | -4.93 | 0.97 | < 0.001 | |
| Once or twice a week | | -6.61 | 1.14 | < 0.001 | |
| Every day or almost every day | | -9.89 | 1.82 | < 0.001 | |
| Assess math students by discussing current performance level. | 299.85 | | | | 0.140 |
| A few times a year | | -3.15 | 1.97 | 0.11 | |
| Once or twice a month | | -4.88 | 1.97 | 0.02 | |
| Once or twice a week | | -7.17 | 2.06 | 0.001 | |
| Every day or almost every day | | -9.12 | 2.07 | <0.001 | |
| Assess math students by discussing progress toward goals. | 298.07 | | | | 0.412 |
| A few times a year | | -1.96 | 0.83 | 0.021 | |
| Once or twice a month | | -4.57 | 0.88 | <0.001 | |
| Once or twice a week | | -7.03 | 1.05 | <0.001 | |
| Every day or almost every day | | -9.74 | 1.78 | <0.001 | |

| | | | | | |
|---|---|---|---|---|---|
| Assess math with individual or group projects. | 294.41 | | | | 0.139 |
| 1-2 times a year | | 1.79 | 0.63 | 0.006 | |
| 1-2 times a month | | -1.01 | 0.76 | 0.192 | |
| 1-2 times a week | | -7.10 | 1.51 | <0.001 | |
| Assess math with multiple-choice tests. | 296.38 | | | | 0.143 |
| 1-2 times a year | | 2.21 | 0.80 | 0.007 | |
| 1-2 times a month | | -3.00 | 0.72 | <0.001 | |
| 1-2 times a week | | -6.51 | 0.95 | <0.001 | |
| Assess math with problem sets. | 294.55 | | | | 0.138 |
| 1-2 times a year | | 1.11 | 2.15 | 0.52 | |
| 1-2 times a month | | -0.30 | 1.39 | -0.21 | |
| 1-2 times a week | | 0.62 | 1.27 | 0.49 | |
| Assess math with short or long written responses. | 292.82 | | | | 0.138 |
| 1-2 times a year | | 1.40 | 1.32 | 0.297 | |
| 1-2 times a month | | 2.76 | 1.10 | 0.014 | |
| 1-2 times a week | | 1.93 | 1.20 | 0.113 | |
| Amount of math homework assigned per day. (contrast: None) | 276.95 | | | | 0.416 |
| Less than 1 hour | | 18.03 | 1.62 | <0.001 | |
| About an hour | | 21.70 | 1.65 | <0.001 | |
| About 2 – 3 hours | | 18.33 | 4.35 | <0.001 | |
| More than 3 hours | | 16.21 | 5.53 | <0.001 | |

**Summary of Results**

The mathematics achievement of Hispanic students was statistically significantly lower than the performance of White students for each of the selected formative assessment items, and in overall performance. Recall that NAEP collects teacher survey data from students who were selected to take the assessment. Multiple regression analyses were conducted, using the frequency, 'Never or Hardly Ever,' as a comparison for each of the above items.

It appears that feedback related activities were negatively associated with mathematics achievement. This parallels Phelan et al.'s (2011) study where formative assessments had little impact on low performing students' achievement in mathematics. When teachers reported a greater frequency of adjusting instruction to meet the needs of students, the scores were lower for Hispanic students. When compared to Never or Hardly Ever, adjusting instruction to meet the needs of students was statistically negatively associated with achievement. Similarly, a negative relationship was found for setting goals for specific progress and achievement. The regression analysis indicated that scores were nearly 10 points lower for doing this nearly every day. Also, discussing current student performance level, and discussing progress toward goals were found to be negatively associated with achievement, with a gap of nearly nine points for each activity. Mathematics achievement was positively related for students whose teachers reported using individual or group projects one to two times a year; and negatively related with mathematics achievement when teachers reported using them more often with students. An infrequent use of multiple-choice tests for assessment was positively associated with mathematics achievement; more frequent use of multiple-choice tests for assessment was negatively related to achievement. The use of problem sets is not related to achievement. This seems to contradict Shirvani's (2009) findings, if quizzes are considered problem sets. The use of written responses once or twice a month was positively related to achievement; however, more infrequent use or a much more greater use of written responses for assessment were negatively related to achievement. Finally, the frequency of Homework was positively related to achievement. Students of teachers who reported assigning about one hour of homework scored nearly 21 points higher than those students whose teachers Never or Hardly Ever assigned homework.

**Implications**

The findings of the current study showed that both Hispanic and White students had similar patterns of achievement in relation to the selected items, although Hispanic students had lower achievement than White students. The nation's Hispanic students scored lower than White students in almost every category. The current study revealed that assessment practices such as discussing with mathematics students their current performance levels, setting learning goals and discussing their progress toward those learning goals are not generally effective for assisting Hispanic students.

These findings, on the surface, appear contrary to current thought about assessment in mathematics, in particular, formative assessment related to feedback. The current study's findings suggest that there is a negative relationship between assessment and mathematics achievement of the nation's eighth graders. Prior studies, including Kingston and Nash's (2011) meta-analysis, although controversial suggested a 0.20 effect size, which indicated higher achievement when formative assessment is used. Ruiz-Primo & Furtak (2007) observed teachers' use of formative

assessment strategies in science and found a 0.92 effect size, which also suggested higher achievement when feedback is used.

The current study's results were determined through the use of multiple regression analysis on the NAEP teacher survey items and their students' scores comparing the frequencies to not doing any of the actions. The findings may be explained by noting that perhaps the teachers who self-reported doing various techniques more often had lower performing students in general. This brings into question what is the quality of the discussions and feedback shared with students in a way that they contributed to what they should do to improve their scores measured by the survey. Hence, it is the quality not the quantity of feedback and discussions that can assist all students to improve their achievement.

Phelan et al. (2011) indicated challenges when trying to fulfill the promise of formative assessment including teachers who often have limited background and capacity to develop or engage in high quality assessment practices. The criteria for formative assessments must be made clear and quality feedback must be provided to students in a timely manner so that they can understand what they need to do to improve at the same time as being engaged in the feedback process.

Other findings suggested that assigning moderate amount homework, about one hour each day, is positively related to achievement. A frequent use, at least once or twice a month, of short or long writing appeared to be positively related to achievement. This demonstrates the efficacy of such a strategy as the best way to provide feedback through moderate amounts of hands-on mathematical tasks of all the listed strategies. It is not known how much effort does the teacher put forth to have students go to the board to provide answers during class and the degree to which students are engaged in rigorous mathematics tasks (NCTM, 2014), which ties into the homework process.

The results related to providing feedback, as noted above was perplexing. A possible explanation may be due to the teachers who reported using formative assessment more often might be teaching lower achieving students. It is contrary to what other studies have found regarding the potential benefit of offering feedback. Perhaps this is due to the way the NAEP items are written. It is unclear whether the teachers conduct those actions with students or other teachers. The students of teachers who reported greater frequency of adjusting instruction to meet student needs had lower achievement. This is a surprising finding. Also, discussing the performance levels and using discussions both were negatively related to achievement, the greater the frequency of occurrence was related to sharply lower achievement scores, although the weakest students can have instruction adjusted and still not be engaged. Teachers, who reported a moderate level of assigning projects to assess mathematics, Once or Twice a Month, had students with higher achievement scores compared to Not Hardly Ever assigning projects. This finding suggests that projects have

the potential to improve student achievement and should be regularly incorporated into instruction. Perhaps there should be greater use of more student-centered assessment strategies.

The current study relied on self-reporting of the frequency of doing particular formative assessment related activities by teachers of students who were selected to take the 2015 NAEP exam. Further probing, similar to the Phelan et al. (2011) study, is necessary to gain insight into the impact of formative assessment on middle school students' mathematics achievement. A finer grained analysis could include the examination of teacher implementing formative assessment techniques, teacher knowledge related to assessment and observations of teachers, and student growth while implanting formative assessments (Phelan et al.). With quality feedback and appropriate adjustments to instruction, student achievement in mathematics can be further enhanced through the use of student-centered, hands-on, rigorous mathematics tasks or projects that have the potential to maximize student engagement. These activities are beneficial for both the general population and the Hispanic student population.

## REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Black, P., & William, D. (1998). *Inside the black box: Raising standards through classroom*

Beede, D., Julian, T., Khan, B., Lehrman, R., McKittrick, G., Langdon, D., & Doms, M. (2011). *Education supports racial and ethnic equality in STEM* (ESA Issue Brief No. 05–11). Washington, DC: U.S. Department of Commerce. http://eric.ed.gov/?id=ED523768 assessment. King's College: London.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy and Practice*, 18(1), 5-25.

Borman, T., Margolin, J., Garland, M., Rapaport, A., Park, S. J., & LiCalsi, C. (2017). *Associations between predictive indicators and postsecondary science, technology, engineering, and math success among Hispanic students in Texas* (REL 2018–279). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from http://ies.ed.gov/ncee/edlabs.

Briggs, D. C., Ruiz-Primo, M. A., Furtak, E., Shepard, L. & Yin, Y. (2012). Meta-analytic methodology and inferences about the efficacy of formative assessment. *Educational Measurement: Issues and Practice*, 31(4), 13-17.

Demos. (2004). *About learning: A report of the learning working group.* London: United Kingdom. Retrieved from Demos website: http://www.demos.co.uk/?s=about+learning

Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessments: The limited scientific evidence of the impact of formative assessments in education. *Practical Assessment &Research and Evaluation*, 14(7), 1-11.

Erasmus, P. (2014). Assessment strategy in mathematics for second language learners. *Literacy Information and Computer Education Journal*, 5(4), 1164-1169.

Fennel, S., Swartz, B. A., McCord Kobett, B., & Wray, J. A. (2015). Classroom-based formative assessments: Guiding teaching and learning. *Teaching Children Mathematics*, 21(6), 325-327.

Filsecker, M., & Kerres, M. (2012). Repositioning formative assessment from an educational assessement perspective: A response to Dunn & Mulvenon (2009). *Practical Assessment, Research & Evaluation,* 17(16). Retrieved from http://pareonline.net/getvn.asp?v-17&n=16.

Garland, M., & Rapaport, A. (2018). *Advanced course offerings and completion in science, technology, engineering, and math in Texas public high schools* (REL 2018–276). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from http://ies.ed.gov/ncee/edlabs.

Good, R. (2011). Formative *use* of assessment information: It's a process, so let's say what we mean. *Practical Assessment, Research, and Evaluation*, 16(3). Retrieved from http://pareonline.net/getvn.asp?v-16&n=3.

Hattie, J. A., (2009). *Visible learning. A synthesis of over 800 meta-analyses related to achievement*. New York: Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112.

Havnes, K., Smith, K., Dysthe, O., & Ludvigsen, K. (2012). Formative assessment and feedback: Making learning visible. *Studies in Educational Evaluation*, 38, 21-27.

Hemphill, F.C., and Vanneman, A. (2011). *Achievement Gaps: How Hispanic and White Students in Public Schools Perform in Mathematics and Reading on the National Assessment of Educational Progress* (NCES 2011-459). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Hickey, & Aderson, (2007). Situative approaches to assessment for resolving problems in educational testing and transforming communitites of educational practice. In P. Moss (Ed.),

*Evidence and descision making. The 103rd NSSE Yearbook* (pp. 269 – 293). National Society for the Study of Education: University of Chicago Press.

Hinojosa, T., Rapaport, A., Jaciw, A., LiCalsi, C., & Zacamy, J. (2016*). Exploring the foundations of the future STEM workforce: K-12 indicators postsecondary STEM success*. Washington, DC: Regional Education Laboratory Southwest, Institute of Education Sciences, U.S. Department of Education. http://eric.ed.gov/?id=ED565641

Hudson, T. (2015). *Six strategies to reach, teach, and close math gaps for Latino English language learners in elementary and middle school*. Bellvue, WA: Dreambox Learning, Inc.

Kantrov, I. (2000). *Assessing students' mathematics learning*. Education Development Center. Retrieved from http://mcc.edc.org/pubs.php

Keeley, P., & Tobey, C. R. (2011). *Mathematics formative assessment: 75 practical strategies for linking assessment, teaching, and learning.* Thousand Oaks, CA: Corwin.

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, 30(4), 28-37.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254–284.

Koretz, D. (2008). *Measuring up: What educational testing really tells us*. Cambridge, MA: Harvard University Press.

Krogstad, J. M. (2016*). Five facts about Latino education*. Pew Research Center. Retrieved from http://www.pewresearch.org/fact-tank/2016/07/28/5-facts-about-latinos-and-education/

Krogstad, J. M., & Fry, R. (2014). *Department of Education projects public schools will be 'majority-minority this fall.* Retrieved from http://www.pewresearch.org/fact-tank/2014/08/18/u-s-public-schools-expected-to-be-majority-minority-starting-this-fall/

McMillan, J. H., Venable, J. C., & Varier, D. (2013). Studies of the effect of formative assessment on student achievement: So much more is needed. *Practical Assessment Research & Evaluation*, 18(2). Retrieved from http://pareonline.net/getvn.asp?v=18&n=2

Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute by minute, day by day. *Educational Leadership,* 63(3), 19–24.

National Assessment of Educational Progress. (2011). *NAEP assessment weighting procedures.* Retrieved from https://nces.ed.gov/nationsreportcard/tdw/weighting/

National Center for Education Statistics. (2016). Digest of education statistics: 2015 (NCES No. 2016–014, Tables 318.45 and 322.20). Washington, DC: U.S. Department of Education. Retrieved from https://nces.ed.gov/programs/digest/d15/

National Center for Education Statistics. (2015, September 22). *Achievement gaps*. Retrieved from https://nces.ed.gov/nationsreportcard/studies/gaps/

National Council of Supervisors of Mathematics. (2014*). Improving student achievement through formative assessment in instruction*. Retrieved from http://www.mathedleadership.org/resources/position.html

National Council of Teachers of Mathematics. (2014). *Principles to actions: Ensuring mathematical success for all.* Author: Reston, VA.

National Council of Teachers of Mathematics. (2013). *Formative assessment: A position of the National Council of Teachers of Mathematics*. Retrieved from http://www.nctm.org/Standards-and-Positions/Position-Statements/Formative-Assessment/

Phelan, J., Choi, K., Vendlinski, T., Baker, E., & Herman, J. (2011). Understanding principles following formative assessment intervention. *Journal of Educational Research*, 104, 330-339. doi: 10.1080/00330671.2010.484030

Reynolds, C., Livingston, R. B., & Wilson, V. (2009). *Measurement and assessment in education* (2nd ed.). Boston, MA: Pearson.

Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching*, 44, 57-84.

Sadler, R., (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment and Evaluation in Higher Education,* 35, 535–550.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Education Researcher*, 29(7), 4-14.

Shepard, L. A. (2008). Formative assessment: Caveat emptor. In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning*. New York: Lawrence Erlbaum Associates.

Shayyan, V., Thurlow, M. L. & Liu, K. K. (2008). Instructional strategies for improving achievement in reading, mathematics, and science for English language learners with disabilities. *Assessment for Effective Intervention*, 33(3), 145-155. doi: 10.1177/1534508407313239

Shirvani, H. (2009). Examining an assessment strategy on high school mathematics achievement: Daily quizzes vs. weekly tests. *American Secondary Education*, 28(1), 34-45.

Stiggins, R. (2002). Assessment crisis: The absence of assessment FOR learning. *Phi Delta Kappan*, 83, 758-765.

William, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.

William, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education*, 11(1), 49-65.

Wylie, E. C., & Lyon, C. J. (2015). The fidelity of formative assessment implementation: Issues of breadth and quality. *Assessment in Education: Principles, Policy, and Practice*, 22(1), 140-160. doi: 10.1080/0969594X.2014.990416