STEM EXPEDITIONS® PILOT STUDY REPORT EXCERPT

PARTNERS

The mission of the Friday Institute is to advance education through innovation in teaching, learning, and leadership. Bringing together educational professionals, researchers, policy-makers, and other community members, the Friday Institute is a center for fostering collaborations to improve education. We conduct research, develop educational resources, provide professional development programs for educators, advocate to improve teaching and learning, and help inform policy-making.

THE WILLIAM & IDA FRIDAY INSTITUTE FOR EDUCATIONAL INNOVATION
NC STATE UNIVERSITY | COLLEGE of EDUCATION

NWEA® is a research-based, not-for-profit organization that supports students and educators worldwide by creating assessment solutions that precisely measure growth and proficiency—and provide insights to help tailor instruction. For 40 years, NWEA has developed Pre-K–12 assessments and professional learning offerings to help advance all students along their optimal learning paths. Our tools are trusted by educators in 145 countries and more than half the schools in the US.

nwea™
Measuring What Matters

STEM EXPEDITIONS BOOST SCIENCE SCORES

According to a December 2017 report from the Friday Institute for Educational Innovation at North Carolina State University's College of Education, STEM Expeditions® helped students achieve statistically significant results on national tests.

Sixth- and seventh-grade STEM Expeditions students at Tucker Creek Middle School in Havelock, NC, scored 20.1 percent higher on the 2016-17 Measures of Academic Progress (MAP) science test than did members of a control group at a nearby middle school.

Eighth graders at Pittsburg (KS) Community Middle School (PCMS) who completed the Expeditions as part of a blended science program showed a statistically significant difference in growth in MAP science and reading than did the virtual comparison group.

According to the Friday Institute's executive summary in the report, "Students in these analysis clusters tended to show more academic growth in science when compared to their peers in a comparison group or when compared to the national median."

Tucker Creek STEM Expeditions teacher Michelle Smith was not surprised at her students' stellar performance. "**I think the results are directly related to the amount of student engagement with the content. Students know from the beginning of the Expeditions that they will be using the scientific principles they are taught to solve the lesson Essential Question, which is an engineering design challenge. Because they know that the information will be immediately valuable to them in designing their solution, they pay more attention and get more involved from the very start."**

An excerpt of the report is contained in the following pages.

Pitsco Education's
2016-17 STEM Expeditions Pilot
Study Report

Andrew Weedfall, Malinda Faber, Jeni Corn
December 2017

THE WILLIAM & IDA FRIDAY INSTITUTE FOR EDUCATIONAL INNOVATION
NC STATE UNIVERSITY COLLEGE of EDUCATION

Results

Student Academic Progress

Student academic progress was measured using the Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP) assessments for Science, Mathematics, and Reading.[1] Specifically, these assessments use each student's assessment scores at multiple time points to yield a single Conditional Growth Percentile (CGP) metric for that student. This metric indicates how much the student's score improved over time compared to their peers of similar grade and age across the country. In other words, it provides a percentile of their learning progress compared to the national median (or 50th percentile). Table 1 summarizes the MAP assessment data analyses and comparisons completed in this study. This section of the report will use the labels in that table when referring to each analysis. Within each of these analyses, data were also analyzed by gender (male/female) in order to detect any differences in academic progress between male and female students.

Table 1: *Descriptions of each analysis cluster used to examine student academic progress*

Analysis Name	Pitsco Education Curriculum Piloted	Analysis Description
Craven County MAP	STEM Expeditions	Student MAP assessment data from a STEM elective class at Tucker Creek Middle School (Pilot Group) were analyzed and compared to student data from a STEM class at Havelock Middle School (Comparison Group).
PCMS MAP	STEM Expeditions	MAP assessment results from a STEM class at Pittsburg Community Middle School (PCMS) were analyzed and compared to a virtual comparison group (VCG) of students across the country who shared similar characteristics with the PCMS sample. These analyses were broken down into three different subject areas of academic progress (mathematics, science, and reading).

[1] www.nwea.org

Craven County MAP – STEM Expeditions

Results from the Craven County MAP Comparison analysis suggest that the STEM Expeditions materials may have had a positive effect on students' science learning at Tucker Creek Middle School. Furthermore, results show that female students who experienced the STEM Expeditions curriculum might have experienced a larger increase in their science learning compared to male students. Table 2 displays results of the comparison between the Pilot participant group (Tucker Creek Middle School) and the Comparison participant group (Havelock Middle School).

Table 2: *Craven County MAP pilot and comparison – science academic progress*

Participant Group	**N**	**MAP Science CGP Mean**	**Standard Deviation**	**Difference in Means**
Pilot (Tucker Creek Middle School)	51	46.5	31.9	20.1 ($p<.001$)
Comparison (Havelock Middle School)	61	26.4	29.1	

As seen in Table 2, students in the pilot group made far more academic progress than did students in the comparison group. Specifically, students at Tucker Creek Middle School had a mean CGP that was 20.1 percentile points higher than students at Havelock Middle School had. Independent t-test results show that this was a statistically significant difference.

Tables 3 and 4 display these data broken down further by gender, separated by pilot and comparison groups.

Table 3: *Craven County MAP pilot – science academic progress by gender*

Students in the Pilot Group (Tucker Creek Middle School)	**N**	**MAP Science CGP Mean**	**Standard Deviation**	**Difference in Means**
Male	32	39.0	30.7	-18.0 (p=.06)
Female	18	57.0	30.9	

Table 4: *Craven County MAP comparison – science academic progress by gender*

Students in the Comparison Group (Havelock Middle School)	N	MAP Science CGP Mean	Standard Deviation	Difference in Means
Male	24	28.2	30.2	.58 (*p*=.95)
Female	24	27.6	29.4	

Results in Table 3 show that although not statistically significant, females in the pilot participant group tended to make more academic progress than males in the pilot participant group (by 18 percentile points). Readers should use caution when making inferences from these results, as the sample sizes of these participant groups, especially the female participant group, are very small. When this happens, results may be idiosyncratic, therefore not representative of broader trends.

In contrast, results from Table 4 show that there were almost no differences between male and female participants in the comparison group. Because of such differences in the pilot and comparison groups, these results suggest that the Pitsco Education Pilot may be affecting males and females differently than other curricula – resulting in more even levels of performance from male and female students. However, as mentioned, larger sample sizes for each participant group would be beneficial for producing more reliable results.

PCMS MAP – STEM Expeditions

The second analysis cluster investigated the MAP Science, Mathematics, and Reading assessment data from a STEM course at Pittsburg Community Middle School (PCMS). The students in this course experienced the Pitsco Education STEM Expeditions curriculum, and the results indicate that they tended to experience more learning growth in science and reading, but less in mathematics, when compared to the virtual comparison group (VCG) provided to The FIRE Team by NWEA. This VCG is made up of students throughout the nation who have taken the MAP assessments and who have demographic and socioeconomic characteristics that match those of the students in the PCMS sample. In addition to demographic and socioeconomic status variables, this VCG contains students who matched the PCMS students based on initial MAP scores. The VCG is a reasonably strong comparison group, because it accounts for many extraneous variables that may affect student performance. However, the VCG does not account for local differences in educational programming. An ideal comparison group would contain all the demographic, socioeconomic, and academic performance characteristics of the VCG and the students in this group would be from the same school, district, or region as the PCMS students. When comparing genders across groups, findings indicate that larger group differences tended to exist among males in science and mathematics, whereas there were no significant differences among the females. In contrast, both males and females in the PCMS group tended to show significantly more growth in reading than the VCG.

Table 5: *PCMS MAP pilot - mathematics, science, and reading academic progress*

MAP Assessment	N	PCMS Average Raw Growth (RIT Score)	VCG Average Growth (RIT Score)	Growth Difference
Mathematics	200	4.4	5.9	-1.5*
Science	181	5.2	3.9	1.3*
Reading	193	5.9	3.4	2.5*

Note: The RIT Score is a measure of student performance on the MAP tests created by the NWEA. For an explanation of the RIT Score, see Appendix A.
* $p < .05$

Table 5 displays the average growth in mathematics, science, and reading of students who were exposed to the STEM Expeditions curriculum, and the average growth of students in a virtual comparison group (VCG). The growth scores listed in Table 5 are on the NWEA Rasch Unit (RIT) scale. For more information about the RIT scale and its development see Appendix A.

The final column of Table 5 shows the difference in growth between students at PCMS who experienced the STEM Expeditions curriculum, and students who were in the VCG. For mathematics, the students at PCMS demonstrated less growth than the students in the VCG did. Although this difference was statistically significant, the effect size was small ($d = -.20$). In contrast, students who experienced the STEM Expeditions curriculum showed larger growth in science and reading than the virtual comparison group. However, although these differences were also both statistically significant, the effect sizes for science ($d = .23$), and reading ($d = .33$) were also small.

These results indicate that for science and reading, the STEM Expeditions curriculum may have had a significantly greater positive effect on student achievement than that of a typical curriculum that students in the VCG may have experienced. In contrast, these results also suggest that the STEM Expeditions curriculum had less of a positive effect on growth in mathematics compared to students in the VCG. Although it is possible that the effect of the STEM Expeditions curriculum is stronger for science and reading than it is for mathematics, there is also the possibility that the nature of the subject content has a longer implementation dip. In other words, perhaps both students and teachers required more time with the new implementation strategies of the STEM Expeditions curriculum in order for them to reach their full potential, and in adjusting to this new form of instruction, students growth was slightly hindered. Longer duration of study could help to show this implementation dip. Additionally, variation in the classroom, school, or district settings and/or educational programming between PCMS and the environments of the VCG students could be driving some of the differences in student performance. A comparison group with the matching demographic, socioeconomic, and academic performance characteristics who are also from a similar location would be an even better test.

Table 6: *PCMS MAP pilot – mathematics academic progress by gender*

Students in pilot group (PCMS)	N	PCMS Average Raw Growth	Average VCG Growth	Growth Difference
Male	109	3.8	6.0	-2.2*
Female	91	5.0	5.7	-.7

Note: * *p* <.05

Table 7: *PCMS MAP pilot – science academic progress by gender*

Students in pilot group (PCMS)	N	PCMS Average Raw Growth	Average VCG Growth	Growth Difference
Male	94	5.7	4.2	1.6*
Female	87	4.7	3.7	1.0

Note: * *p* <.05

Table 8: *PCMS MAP pilot – reading academic progress by gender*

Students in pilot group (PCMS)	N	PCMS Average Raw Growth	Average VCG Growth	Growth Difference
Male	103	5.8	3.4	2.4*
Female	90	5.9	3.4	2.5*

Note: * *p* <.05

As Tables 6 shows, although both male and female students at PCMS tended to show less growth in mathematics than their VCG counterparts, only males showed a statistically significant difference. In other words, it appears that although the STEM Expeditions curriculum was less effective than the curricula that VCG students experienced, this was more so the case for males than it was for females. Table 7 shows a similar, but opposite pattern for science progress. Specifically, although male and female students at PCMS both demonstrated more science growth than students in the VCG did, this difference was only statistically significant for males. Finally, table 8 shows that both male and females students at PCMS demonstrated a statistically significant difference in reading growth compared to students in the VCG. It should be noted that like the aggregated findings above, all of these statistically significant findings had small effect sizes.

Appendix A

The following sections come from pages 23 -24 of Northwest Evaluation Association's 2011 *Technical Manual for Measures of Academic Progress® (MAP®) and Measures of Academic Progress for Primary Grades (MPG).*

The Measurement Model

Item Response Theory (IRT) (Lord & Novick, 1968; Lord, 1980; Rasch, 1980) defines models that guide both the theoretical and practical aspects of NWEA scale development. IRT provides a basis for measurement scale development that rests on the relationship between a student's achievement level and item characteristics. MAP and MAP for Primary Grades use the one-parameter logistic IRT model (1PL), also known as the Rasch model. This model estimates the probability, P_{ij}, that a student, j, with an achievement level of θ_j, will correctly answer a test question, i, of difficulty δ_i, and can be expressed as:

$$P_{ij} = \frac{e^{(\theta_j - \delta_i)}}{1 + e^{(\theta_j - \delta_i)}} \qquad (1)$$

A benefit of the use of an IRT model is that values of achievement levels, θ_j, and the values of item difficulties, δ_i, reside on the same scale. The scale is equal interval in the sense that the ratio of the log odds of success on dichotomously scored items of the same difficulty for any two individuals of differing achievement levels is maintained throughout the scale.

The RIT Scales

The value of the achievement levels and item difficulties in Equation (1) is on the logit metric. The logit metric is an arbitrary scale that is commonly used for academic studies of the Rasch model. To allow the measurement scale to be easily used in educational settings, a linear transformation of the logit scale is performed to place it onto the RIT (Rasch unIT) scale. This transformation is:

$$RIT = (\theta_j * 10) + 200 \qquad (2)$$

This scale has positive scores for all practical measurement applications and is not easily mistaken for other common educational measurement scales. The RIT scale was developed by NWEA for use in all MAP tests and MAP for Primary Grades survey with goals tests.

The RIT scales, like other IRT measurement scales, have a number of useful characteristics when applied and maintained properly. The most important characteristics to the development the measurement scales and item banks are:

- Item difficulty calibration is sample free. This means that if different sets of students who have had an opportunity to learn the material answer the same set of questions, the resulting difficulty estimates for any particular item are estimates of the same parameter that differ only in the accuracy of the estimate's value. The accuracy will differ due to the sample size and the relative achievement of the students compared to the difficulty of the items.
- Achievement level estimation is sample free. This means that if different sets of questions are given to a student who has had an opportunity to learn the material, the scores obtained are estimates of the same student achievement level. Again, accuracy may differ due to the number of items administered and the relative difficulty of the items compared to the student's level of achievement.
- The item difficulty values define the test characteristics. This means that once the difficulty estimates for the items to be used in a test are known, the precision and the measurement range of the test are determined.

These properties of IRT have been empirically verified for the RIT scales (Ingebo, 1997), and can be used in a variety of test development and delivery applications. Since IRT enables one to administer different items to different students and obtain comparable results, the development of targeted tests becomes practical. Targeted testing is the cornerstone for the development of computerized adaptive testing as well as level testing systems. These IRT characteristics also facilitate the building of item banks with item content that extends beyond a single grade level or school district. This enables the development of measurement scales and item banks that extend from elementary school to high school. By combining these properties of IRT with appropriate scale development procedures, NWEA has developed scales and item banks that endure across time and generalize to a wide variety of educational settings.

PITSCO EDUCATION
www.pitsco.com | 800-828-5787