Leveraging Technology to Improve Test Scores: A Case Study of Low-income Hispanic Students

David A. Stevens
Harvard University

Paper presented at the International Conference on Learning with Technology
Temple University – March 8, 2000

INTRODUCTION

Federal Title 1 programs implemented to improve the academic development of disadvantaged children have often been ineffective in affecting students with limited English proficiency (LEP). These students are particularly difficult to reach because they often come from poor families and live in urban communities with high concentrations of poverty. They can have parents with particularly low levels of education who may be unemployed, and who have a limited proficiency in English and rarely speak English at home. (Ross & Puma, 1995) This paper is a case study of one elementary school that appears to have succeeded in using educational technology to improve academic performance for LEP students.

Phonemic awareness is the ability to analyze and manipulate sounds within syllables. It is a strong predictor of reading progress and is crucial for the efficient decoding of print into sounds. Phonologically based interventions are effective techniques for addressing reading difficulties (Bus & van Ijzendoorn, 1999; Snow et. al., 1998). However, these interventions require extensive teacher training and staff time. Computer based reading programs reduce the need for intensive and costly staff training and make replication of the intervention easier to achieve for most schools. These applications of technology to reading difficulties have demonstrated benefits in clinical trials (Merzenich & Jenkins, 1996; Snow et. al. 1998; Wise & Olson, 1995). This study considers the question; "can phonemically based reading software address the specific needs of LEP students?"

Context

Tornillo Elementary School is a public school located 40 miles South-East of El Paso, Texas, near the U.S. – Mexico boarder. The school, which is three miles from the boarder crossing, serves students of many families who have moved from Mexico. Over 98% of the students are Hispanic and more than 96% are economically disadvantaged. The combination of low socio-economic status and limited English proficiency interact to create a high-risk situation for school failure.

Technology

Tornillo Elementary purchased two reading software packages after researching the options and following the suggestion of a local consultant specializing in computer-based school curricula. *Lexia Phonics Based Reading* (Lexia Learning Systems, Inc.)[1] is based on the Orton Gillingham system of reading skills development. The

[1] www.lexialearning.com

software contains three levels, each level offering five learning activities with between 7 and 27 units. Each of the fifteen activities uses extensive logic trees to automatically branch to easier or harder challenges based on student performance. *Level One* provides practice in discriminating "b/d/p", develops letter/sound correspondence for short vowels and consonants and works on the decoding of simple short vowel words. *Level Two* introduces long vowel words and provides practice in decoding silent-e words and more complex short vowel words. In *Level Three*, students decode and construct one- and two-syllable words containing short and long vowels, vowel-r combinations, and vowel combinations. The software provides extensive reports for the teacher to track student performance and identify areas for instruction.

The *Guided Reading* software (Taylor Associates[2]) is designed to foster improvements in basic reading abilities such as visual coordination, visual memory, and concentration. Students first read in a self-paced manner with a reading selection a year below their level. After immersing themselves in the story, the students read a second portion of the selection in a timed and left-to-right guided manner using text moving through a small window on the screen. After reading each passage, students take on-line tests regarding the content. Rate advancement is based on comprehension test performance.

Intervention

Beginning in 1996, 2nd through 5th grade students began using these software packages at Tornillo Elementary. All students started with the *Lexia* software, and when they completed it, they moved on to the *Guided Reading* program. The students worked in the software with a teacher supervising five days a week for 40 minute sessions.

The total yearly contact time that the students interacted with the intervention software is more than 120 hours. This time was a portion of the time previously allotted for traditional reading instruction. Both the total contact time and the frequency of contact with the intervention is considerable in relation to other educational interventions that seek to improve underlying learning abilities. Successful interventions usually maintain near daily contact with the students and have sustained this contact for long periods (Stevens, 1999).

Several of the teachers I spoke with reported that they believed the reading software to be effective in improving the students reading ability. One teacher commented that since the English language has certain sound structures that do not exist in Spanish, it is particularly difficult for the children to develop proficiency with the language. "In Spanish we don't have some of the same sounds." When asked why the software was so effective another teacher replied,

> One of the reasons that it works is that it does so much repetition. It never gets tired like a teacher might. Let's say a student can't hear the difference between the 'e' and 'i' sounds. By about the third time they have gone over it in class the teacher is about to quit. But the computer just keeps saying good job, try again, and in a voice that doesn't get rattled.

The teachers and principal cite the increase of standardized test scores as evidence to support their observation that the technology

[2] www.ptcsi.com/taylor.htm

is making a positive difference. Figure 1 represents the change over time in the estimated percentage of students who take the *TAAS* at Tornillo Elementary and the percentage of students who master all of the test objectives. This graphic demonstrates the increase in test scores the year following the introduction of the software-based reading intervention (1996-1997) and the subsequent increase in the percentage of students who are able to take the test.

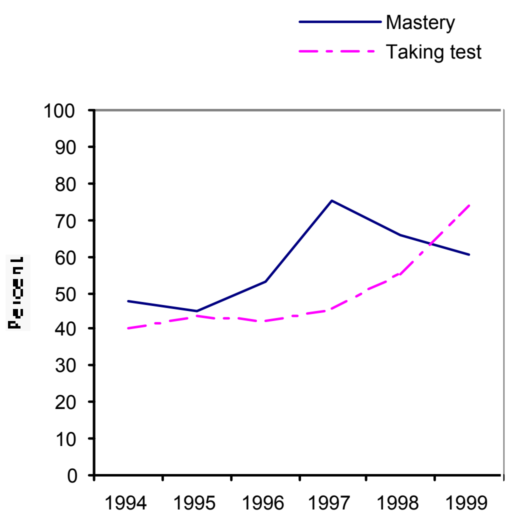

Figure 3. Percentage of students mastering all objectives of the *TAAS* compared to the percentage of students taking it.

METHODOLOGY

Pogrow (1999) suggests that the best measure of the effectiveness of a program is by evaluating how consistently it can produce surprising amounts of learning relative to both gains and achievement levels. He advocates a comparison to historic patterns and reasonable standards. Consistently large gains and achievement is a more relevant measure for practitioners and policymakers than rigorous control group analysis. "It better enables a practitioner to project whether a program is likely to produce large gains in his or her schools (Pogrow, 1999)."

The context of this study presented several potentially confounding variables that made an analysis solely relying on this method problematic. The *Texas Assessment of Academic Skills* affords limited English proficient students an exemption. About half of the students were granted these exemptions before the introduction of the reading software. In the last few years new leadership at the school has minimized the use of exemptions as reading abilities improve. This issue complicates a straightforward analysis of achievement levels over time for several reasons. First, in the first half of the 1990's only the top students (in English) were taking the *TAAS*. Towards the second half of the 1990's, more students of a lower English proficiency started to take the test. As the number of test takers increases (especially the number with a low English proficiency) the composite of ability changes and one would expect English test scores to fall.

Second, even before the move to include more students in the *TAAS* evaluation process, finding stability in the annual test scores was complicated by the fact that only a small number of the students would take the test and they would often represent a wide range in ability.

Third, it is difficult to evaluate the effectiveness of the intervention because the students who theoretically stand to benefit the most do not take the test. In other words, those students who have the most difficulty in acquiring proficiency in English do not take the test and these are the high-risk students we are most interested in learning about.

Fourth, given the migrant nature of many of the families who come to Tornillo, the administration reports a high incidence of students arriving or leaving during the course of the school year. This means that some students who participate in the intervention are not present to take the *TAAS* and some students who take the *TAAS* were not present for a good portion of the school year.

The strategy I outline here represents an effort to overcome these challenges and to determine, as best as possible, the value of the use of new technologies in Tornillo.

Analytic sample

In order to best deal with some of the complex issues that this research context presents I constructed control and experimental groups from different years of school classes. The control groups consist of those students who were in fourth grade in 1994 and those students who were in fourth grade in 1995. I call this the Traditional Group because they were using traditional reading methods exclusively. The experimental group consists of students who were in fourth grade in 1996 and 1997. The 1996 group used the computer-based reading technology in fifth grade and the 1997 group used the reading software in fourth and fifth grade. I call this the Computer Group because they used both the traditional learning methods and the computer-based ones as well.

I excluded all students from each of these groups who did not take the *TAAS* in fourth and fifth grade. Accordingly, this constitutes a convenience sample. In no way does this sample form a representative group of the larger student population. In fact, since we know that the students with less English proficiency were exempt from taking the test, this sample consists of students who were relatively more proficient in English but still qualify as LEP.

In order to address the difficulty posed by the non-representative nature of the sample I used gain scores (difference between fifth grade *TAAS* scores and fourth grade *TAAS* scores) for each individual student. This allows me to identify the specific improvements and regressions of each student as opposed to comparing the overall mean performance of one group to the other.

Outcome measure

The *Texas Assessment of Academic Skills* serves as the outcome measure. The test has two components: reading comprehension and mathematics. The reading comprehension component focuses on word meaning, supporting ideas, relationship and outcomes, inferences and generalizations, and point of view. The math section emphasizes concepts, operations and problem solving and includes extensive word problems.

I have included the math component of the *TAAS* into my analysis because the teachers indicate that the software programs help math performance in addition to reading performance. Here is a sample question from the mathematics section of the test.

> The largest fifth grade class in Pueblo Viejo School has 24 students. The smallest fifth-grade class has 20 students. If there are six fifth grade classes, which is a reasonable total for the number of fifth graders attending Pueblo Viejo School? (The student chooses from four possible answers).

The reading comprehension section uses reading passages followed by multiple choice questions regarding topics such as the main idea of the passage, what a character might have felt, or a specific instruction or detail.

The research question

The research question for this study is, 'are there any differences in gain scores of the traditional group and the computer group?' Differences in favor of the traditional group would lend evidence to support the hypothesis that the reading software intervention has not supported the students' performance on the *TAAS.* It could even suggest that the computer based reading intervention has hindered the students' performance by taking away time from their classroom preparation for the *TAAS*. A finding of no difference between the gain scores of the two groups would suggest that the reading software intervention neither improves nor undermines the computer group's performance on the *TAAS*. A finding that the *TAAS* gain scores are higher for the computer group would lend evidence to support the notion that the computer based reading software provides a different kind of reading skill development and that this supplemental technology is helping the students improve their reading skills.

Model building strategy

In order to answer the research question I constructed a series of multiple regression models to determine the associated change in gain score with the software intervention. I employ the flexible data analytic strategy outlined by Wampold & Freund (1987). I begin with a baseline model using gender as a control to predict change in scores. I then add a variable for the Traditional/Computer group. I add a third variable for a group-gender interaction that would indicate if the groups' performance differed by sex. It is possible that the answer to the research question will be different depending on the students' gender. In a review of education intervention outcome studies (Stevens, 1999) I concluded that researchers should allow for the possibility that interventions influence different groups of students in different ways.

PRIMARY FINDINGS

The *TAAS* scores I use in this study are 'national curve equivalent' which facilitate analyses looking at cross test comparisons and comparisons of changes in scores. The mean NCE score on the *TAAS* is 50 with a standard deviation of 21. In other words, the national average for both fourth and fifth grade tests is 50, and any gains or losses is relative to this 'national curve equivalent' metric. There are 37 students in the traditional group and 33 students in the computer group.

Reading Comprehension

Table 1 represents the multiple regression models for reading comprehension. Based on the delta chi-square test I determined that model number three, with the gender-group interaction is the best fitting model. The significance of the interaction is that the difference between the group's gain scores varies by gender. Not only is the difference between the traditional and computer groups statistically significant, but these differences vary by gender. Accordingly, I present the predicted scores for each gender separately.

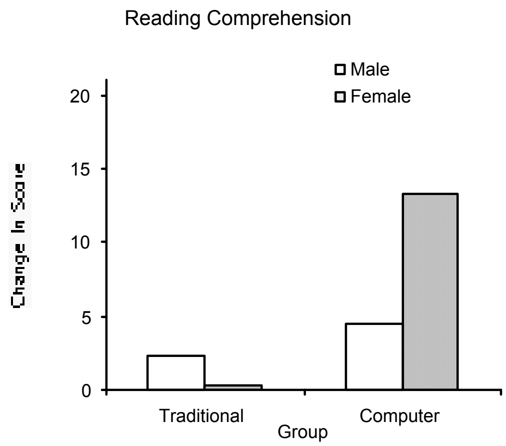

Figure 2. Comparison of reading change scores from fourth to fifth grade (n=70).

	Parameter Estimate (Standard Error)		
Predictor	*Model 1*	*Model 2*	*Model 3*
Intercept	4.496 (2.55)	2.411 (2.72)	0.024 (3.00)
Control Predictors			
GENDER	-0.679 (3.33)	-2.324 (3.37)	2.270 (4.36)
Question Predictor			
GROUP		6.487~ (3.32)	13.236* (5.28)
Interactions			
GROUP_GENDER			-10.965~ (6.74)
Model Comparison			
R^2 Statistic	0.000	0.056	0.093
Error df	66	65	64
Comparison Model		model 1	model 2
Change in R^2		0.056*	0.037*
Change in df		1	2

~p<.10 *p<.05 **p<.01 ***p<.001

Table 1. Nested multiple regression models predicting change of reading comprehension (n=70).

On average, the traditional group boys' performance is associated with a 2.29 difference between fourth and fifth grade test scores while their computer group counterparts are associated with a 4.57 change. On average, the female traditional group is associated with a 0.30 difference in fourth and fifth grade scores while their computer group counterparts are associated with a 13.26 difference.

Mathematics

Table 2 represents the multiple regression models for mathematics performance on the *TAAS*. I conclude that model number two is the best fitting model as determined by the delta chi-square test. There is a statistically significant difference between the math gain scores of the traditional and computer groups. However, there is no gender interactionl.

On average, the boys in the traditional group are associated with a –1.40 difference in math performance from fourth grade to fifth grade. Their counterparts in the computer group are associated with a 14.63 change from fourth grade to fifth grade. On average the girls traditional group is associated with a 0.37 change from fourth to fifth grade while their computer group counterparts are associated with a 16.41 change. Figure 3 compares these gain scores.

	Parameter Estimate (Standard Error)		
Predictor	*Model 1*	*Model 2*	*Model 3*
Intercept	4.782 (2.39	-0.372 (2.05)	-0.516 (2.31)
Control Predictors			
GENDER	2.290 (3.12)	-1.776 (2.54)	-1.472 (3.36)
Question Predictor			
GROUP		16.036*** (2.50)	16.482*** (4.08)
Interactions			

GROUP_GENDER			-0.725 (5.19)
Model Comparison			
R^2 Statistic	0.008	0.392	0.392
Error df	66	65	64
Comparison Model		model 1	model 2
Change in R^2		0.386***	0.00
Change in df		1	2

~p<.10 *p<.05 **p<.01 ***p<.001

Table 2. Nested multiple regression models predicting change of reading comprehension (n=70).

DISCUSSION

Interpretation of findings

The multiple regression analysis supports the hypothesis that the software based reading intervention improves the students' reading ability. In both the math and reading portions of the *TAAS*, the students who participated in the intervention curriculum demonstrated statistically significant higher gain scores than the students who participated in the traditional curriculum. The rate at which the computer group improved differed according to gender in the case of reading comprehension. While on average the boys score improved by 4.57 points the girls score improved by an average of 13.26 points.

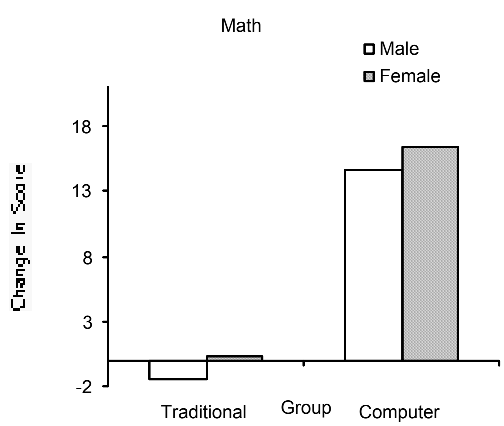

Figure 3. Comparison of math change scores from 4th to 5th grade (n = 70).

The multiple regression analysis of math scores reveals a similar story, although gender appears to be less of a factor. The computer group registered average gain scores of 14.63 (boys) and 16.41 (girls) compared to a decrease of 1.40 (boys) and an increase of 0.37 (girls) for the traditional group.

This study provides strong evidence that the software based reading intervention chosen in Tornillo is associated with strong student gains on the *Texas Assessment of Academic Skills*. Increases in scores range from a quarter to three-quarters of a standard deviation.

Reading technology and math gains?

It is interesting that the gain scores for the computer group are higher for math achievement than for reading. Yet, the intervention software in no way addresses mathematics ability. An analysis of the *TAAS* test revealed that the math portion relies heavily on word-based math problems. I offer the hypothesis that in the past, the students' math scores were depressed relative to their true math ability because of their lack of proficiency with English. According to this scenario, the students would score higher on a math test in their native language. Accordingly, their striking improvements on the math portion of the *TAAS* reflects an improved ability to navigate the verbal text of the math problems (see earlier example) and express more of their true math knowledge.

Strengths and weaknesses of the study

The strength of this study is that it provides a rare critical evaluation of the use of new technologies in a public school that supports a low income and predominantly Hispanic community. These on-site evaluations are important because while an innovative and exciting technology might demonstrate impressive gains in one context, it could prove to be inappropriate or ineffective in another. The weakness of the study is that a representative sample of the student population was not achieved, possibly under estimating the strength of the intervention as those who stood to benefit the most could not be included.

Implications for LEP students

The lesson for research on LEP students is that interventions at the level of phonological understanding of English appear to be a promising avenue. Clearly, we need more research to confirm these findings. Reviews of the literature on academic achievement for LEP students (e.g. Collier, 1989) and statements on reading development in general (e.g. Snow et. al., 1998) make no mention of phonemically based reading interventions for LEP students. Technology based interventions such as those used in this study greatly simplify the task of delivering the same intervention to different groups in different locations.

Implications for the use of technology

Some researchers in the emerging field of educational technology call for a move away from the drill and practice use of technology to more 'cognitively challenging' implementations such as web pages or newsletters. Some have been particularly critical of minority and low income students use of computers for drill and practice while higher SES students are more likely to use them for simulations or applications (Warschauer, 2000). Hoping the era of 'drill and kill may fade away', he echoes a persistent theme in the literature that this sort of extensive task specific use of technology is a lower form of application on the hierarchy of technological innovation.

The *Lexia* and *Taylor* programs clearly fall in the drill and practice category. Based on the results of this study, the teachers and administrators of Tornillo Elementary appear to have found an appropriate methodology and technology to address the specific learning and developmental needs of their students. The concept of cognitive challenge is relative to the developmental level of the user. While to an outsider it may appear that a fifth grade student working through a first grade level of a phonemic awareness program is not being cognitively challenged – I offer that if this is their true level of ability they are indeed being cognitively challenged. We can provide more refined recommendations for specific uses of technology in specific contexts supported by empirical study instead of categorical recommendations for the use of one technology over another regardless of context.

Does technology make a difference?

For the students of Tornillo Elementary School the reading software technology is making a positive difference as the students demonstrate improved test scores in reading and math. The more interesting question is in what ways does certain kinds of technology make a difference in specific contexts?

Why does the technology make a difference?

1) The technology is a good fit for the context. The students in this context have a specific difficulty with school learning related to their limited English

proficiency. The choice of technology is appropriate for the needs of these students.

2) The technology is theory driven. The *Lexia* reading software is based on a well-established and sophisticated theory of addressing reading difficulties. The *Reading Plus* software serves as an effective complement by allowing the students the opportunity to habituate their newly developed abilities and foster other important reading skills.

3) The technology spans many developmental levels. *Lexia Phonics Based Reading* can drop low enough to accommodate students at a first grade level and move high enough to sustain their development over a several year period of time. Teachers indicate that indeed, some 5th grade students did need to start the *Lexia* software at the first grade level.

4) The application of the technology is appropriate to address the needs of the students. Five days a week at forty minutes a day is a considerable commitment to a new technology. However, it is consistent with the amount of contact time of other successful interventions that build underlying learning abilities (Stevens, 1999). Both the frequency and duration of the use of the technology is appropriate to achieve the level of cognitive changes needed.

5) The school was patient enough in its evaluation of the technology to allow for extended use before making a decision of its value. Interventions of this sort often do not yield immediate returns as they work at the level of underlying learning abilities. In some cases interventions do not yield positive results for up to a year after completion of the intervention (e.g. Adey and Shayer, 1994).

6) The educational technology does not rely on intricate plot situations or cute animated characters to engage the students. The design principles of both software packages are streamlined relative to the current trend toward "edutainment" to allow the student to focus on the challenge at hand. The emphasis is on the reading theory behind the software not the media content. As one teacher noted, "It seems so boring to us we can't imagine why it works. But their reading speed improves and so does their comprehension, and they never get bored."

7) A computer lab manager, well trained in the use of both programs, was available at all times when students were using the software. The principal of the school attributes much of their success with technology to her efforts to keep both teachers and students on track with their use of software.

CONCLUSION

Because of decisions made at the district level to introduce other educational technologies into the curriculum, the school has had to reduce the frequency of the use of the reading software reported in this study. Initial teacher interviews reveal that classroom test scores have dropped and the students do not appear to be making the same level of steady progress as when they used the reading software five times a week. One teacher comments, "Reading has just come down, period. Comprehension levels have come down. There have been quite a lot of stories about decreasing test results."

While this anecdotal evidence is hardly conclusive, I present it to illustrate the inherent complexity and messiness of installing and evaluating new educational technologies. There are so many variables in this fluid complex system that our questions and methods need to be appropriate to the reality of the situation. Technology can make a positive and measurable difference, if it is based on sound educational theory, is applied in an appropriate manner, and is a good fit for the context at hand.

David Stevens (David@post.harvard.edu) is a doctoral candidate at the Harvard Graduate School of Education. He specializes in human development and psychology and his research interests are cognition, educational interventions, and technology.

REFERENCES

Adey, P., & Shayer, M. (1994). *Really raising standards: Cognitive intervention and academic achievement.* New York: Routledge.

Bus, A. G. & van Ijzendoorn, M. H. (1999). Phonological awareness and early reading; A meta-analysis of experimental training studies. *Journal of Educational Psychology, 91*(3), p. 403-414.

Chall, J. S. (1983). *Stages of Reading Development*. New York: McGraw Hill.

Collier, V. (1989). How long? A synthesis of research on academic achievement in a second language. *TESOL Quarterly, 23*(3), 509-531.

Merzenich, M. M., & Jenkins, W. M. (1996). Temporal processing deficits of language-learning impaired children ameliorated by training. *Science, 271*.

Pogrow, S. (1999). Rejoinder: Consistent large gains and high levels of achievement are the best measures of program quality: Pogrow responds to Slavin. *Educational Researcher, 28*(8), 24-31.

Ross, M., & Puma, M. (1995). *Prospects: The Congressionally mandated study of educational growth and opportunity. First year report on language minority and limited English proficient students.* (EDD00006). Cambridge, MA: Abt Associates.

Stevens, D. A. (1999). *Transfer revisited: A critical review.* Unpublished qualifying paper. Harvard University, Cambridge, MA.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Wise, B.W., & Olson, R.K. (1995). Computer-based phonological awareness and reading instruciton. *Annals of Dyslexia (45)* 99-122.

Wampold, B. E., & Freund, R. D. (1987). Use of Multiple Regression in Counseling Psychology Research: A flexible data-analytic strategy. *Journal of Counseling Psychology, 34*(4), 372-382.

Warschauer, M. (2000). Technology and school reform: A view from both sides of the tracks. *Education Policy Analysis Archives. 8*(4).