Gear Up III "Creating a Vision" Analysis Report for 2008-2009 Academic Year

Vinh Pham, Ph.D.

The Gear Up III Creating a Vision project is a mathematics education intervention program implemented in 7 small school districts in the rural area surrounding the Laredo, Texas area with primarily Hispanic students. The 2008-2009 academic year was the first year of implementation using the TI-Navigator system with TI-84 graphing calculators. High performing students increased their TAKS scores significantly, while the program had no effect on low performing students. Caution should be used when generalizing these results due to a number of complicating factors including severe student data attrition and a lack of control groups.[1]

Introduction

The Gear Up III project is an educational intervention program designed to improve student success in mathematics in the 8th grade. During the 2008-2009 academic year, it was introduced in eight rural school districts near Laredo, Texas. All participating school districts had greater than 90% Hispanic student enrollment according to the available demographic data (Student Assessment Division, 2009b). Student success was defined in this study as increased mastery and knowledge in mathematics as evidenced by an increase in the scores on the mathematics section of the statewide standardized test, the Texas Assessment of Knowledge and Skills (TAKS) exam. An integral component of the intervention program was the use of the Texas Instruments TI-Navigator™ classroom network system with TI-84 graphing calculators to implement generative pedagogy. These technologies have been shown to improve student scores in previous studies (Stroup, Pham, Alexander, 2007; Owens, et al, 2008). Professional development (PD) was provided to teachers in the participating school districts prior to the start of the academic year to instruct the teachers on effective teaching with the TI-Navigator technology in the classroom. The initial PD workshop lasted for 5 days and afterwards 4 more periodic 2 day PD workshops were provided during the year. A TI-Navigator coach also scheduled regular classroom visits to observe the implementation of the devices in classroom practice**.**

[1] This study was funded in part by Texas Instruments, Inc.

Study Design and Analysis

Approximately 1,200 students participated in the Gear Up III intervention program for the academic year 2008-2009. Gear Up III, continues to identify new students but only carries 974 original cohort students in program from 2007. However, only 299 students had complete data due to student data attrition. Furthermore, no control group of similar students was possible since entire districts participated. This limited the types of analysis that could be conducted on the data as well as the conclusions that could be drawn. Fortunately, historical data was available for the two previous years to establish baseline performance prior to the implementation of Gear Up III, so a repeated measures design was used. All math TAKS scores were converted to percentile rankings based on population statistics made public by the Texas Education Agency (Student Assessment Division, 2009a) for the given year and grade level. This score conversion was deemed necessary in order to compare student performance longitudinally against the entire population due to the inherent differences in the TAKS exam across years and grades.

Descriptive Statistics

The descriptive statistics and box plot for the Gear Up III data are shown in Table 1 and Figure 1 respectively. Note that any learning effects due to Gear Up III would only show up in the 2009 scores.

Year	Median	Mean	S.D.
2007	23.60	31.94	26.52
2008	33.18	39.31	27.85
2009	37.23	42.37	29.31

Table 1. Descriptive statistics of students with complete data for all 3 years. Gear Up III was implemented in 2009.

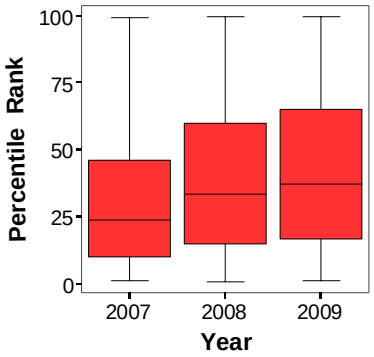

Figure 1. Boxplot of student data with complete data. Gear Up III was introduced in 2009.

The students in the Gear Up III intervention program overall performed below the state mean which is the 50 percentile rank each year based on the descriptive statistics. However, they did improve each year since 2007 and in 2009, when Gear Up was introduced, were approaching the state mean.

Results

Friedman Test

Exploratory data analysis indicated significant departures from normality due to both skewness and kurtosis across all years. This would be a violation of the basic assumptions of parametric statistical analysis. As such, nonparametric statistical analysis was deemed more appropriate for this particular dataset than parametric statistical analysis. The nonparametric statistical equivalent of a repeated measures ANOVA, the Friedman test, was performed on the dataset of complete students. The results of the test are shown in Table 2. Since the test indicated significant differences, Wilcoxon pairwise post hoc testing with Bonferroni correction was done to determine which of the years were different from each other as shown in Table 3.

Year	Mean Rank
2007	1.56
2008	2.09
2009	2.35

N	299
χ^2	97.07
df	2
Asymp. Sig.	0.0000

Table 2. Results for Friedman Test for all students with complete data. Gear Up III was introduced in 2009.

	2008-2007	2009-2008	2009-2007
Z	-7.46	-3.47	-10.01
Asymp. Sig. (2-tailed)	0.0000	0.0005	0.0000

Table 3. Wilcoxon pairwise post hoc testing of Friedman Test. Individual pairwise test is $\alpha = 0.0167$ for Bonferroni correction and overall $\alpha = 0.05$.

The post hoc testing indicated that each year was significantly different from the other years statistically with the scores increasing each year.

Regression Analysis

All regression analysis was done using Reduced Major Axis (RMA) regression as opposed to Ordinary Least Squares (OLS) regression. RMA regression was done to account for the error component inherent in the measurement of a latent trait in both the independent and dependent variables. The latent trait in this case was achievement in mathematics on the TAKS exam for respective years. RMA regression should yield less biased estimates than OLS regression.

The RMA regression lines for respective years are shown in Figures 2 and 3. A large proportion of the variance could be explained by the regression fit (~70%). The regression fits for 2007-2008 and 2008-2009 showed very little deviation from each other. Note that both slopes were slightly greater than 1 and that both intercepts were positive. Taken together these results indicated that the students were improving each year. The graph of the difference between 2009-2008 as a function of the difference between 2008-2007 is shown in Figure 4. Only about 18% of the variance could be explained by the regression fit.

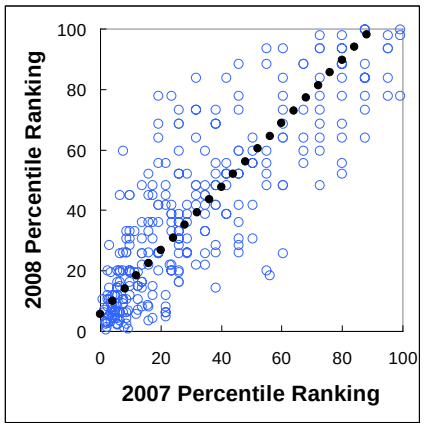

Figure 2. RMA regression of 2007-08 data for students with complete data (y = 1.050 x + 5.769, R^2 =0.666).

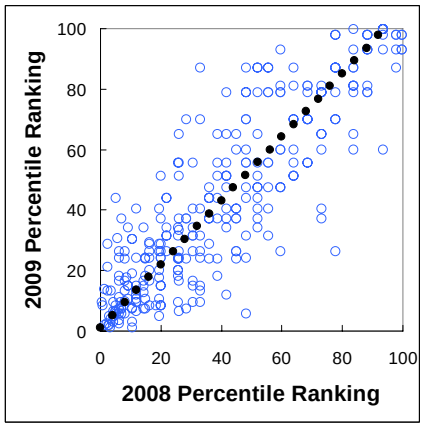

Figure 3. RMA regression of 2008-09 data for students with complete data (y = 1.053x + 0.988, R^2 = 0.749).

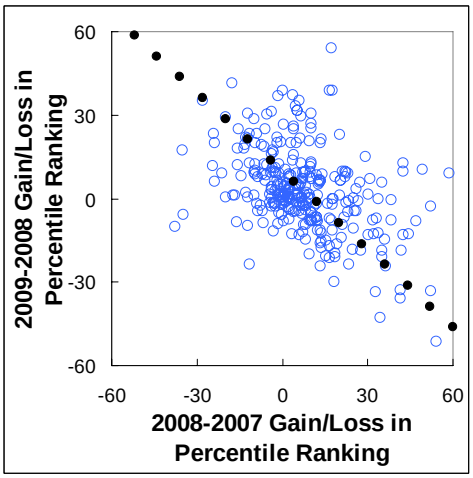

Figure 4. Gain/loss (Difference between 2009 and 2008 scores) as a function of the difference between 2008 and 2007 (y = -0.936 + 9.958, R^2 = 0.179).

Cutoff Analysis

Cutoff analysis was done for statistical reasons that will be detailed later in the report. In cutoff analysis, all students who scored less than the state median in 2007 (scale score < 2266) were separated from those students who scored at or above the state median in 2007 (scale score >= 2266). To facilitate discussion, the students who scored less than the state median in 2007 were labeled as Type 1 students (N = 216) and those students who score at or above the state median in 2007 were labeled as Type 2 students (N = 83). The median was used instead of the mean because the student distribution was non-normal and thus the median would be a better measure of central tendency.

The descriptive statistics and boxplot for Type 1 students are shown in Table 4 and Figure 5, respectively. While there was a rather large increase from 2007 to 2008, the Type 1 students exhibited similar values for both 2008 and 2009.

Year	Median	Mean	S.D.
2007	15.86	17.71	11.76
2008	25.97	27.50	19.89
2009	26.33	29.59	20.87

Table 4. Descriptive statistics of Type 1 (below median) students for cutoff analysis.

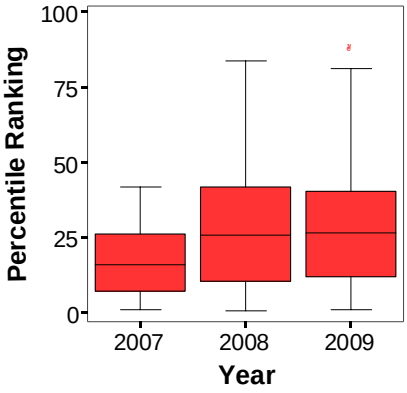

Figure 5. Boxplot of Type 1 (below median) students for cutoff analysis.

A Friedman test with pairwise post hoc testing was also done as shown in Tables 5 and 6 respectively. Only 2007 exhibited significant difference from the other

years. The 2008 and 2009 scores were not significantly different. Mathematically, this can be represented as 2007 < 2008 = 2009 scores.

Year	Mean Rank
2007	1.46
2008	2.16
2009	2.38

N	216
χ^2	100.18
df	2
Asymp. Sig.	0.0000

Table 5. Results for Friedman Test of Type 1 student for cutoff analysis.

	2008-2007	2009-2008	2009-2007
Z	-8.56	-2.21	-9.77
Asymp. Sig. (2-tailed)	0.0000	0.0273	0.0000

Table 6. Wilcoxon pairwise post hoc testing of Friedman Test. Individual pairwise test is $\alpha = 0.0167$ for Bonferroni correction and overall $\alpha = 0.05$.

Similar analyses were done for the Type 2 students as shown in Table 7 and Figure 6, which are the descriptive statistics and boxplot respectively. Friedman test results for the Type 2 students indicated that 2007 and 2008 scores were not significantly different, but 2008 and 2009 were significantly different. As expected, 2009 was also different from 2007. Mathematically this can be represented as 2007 = 2008 < 2009 scores with the gain between the latter two years close to 5.5 percentile ranking change. This change in percentile ranking has been found to be relatively standard in other studies using the TI-Navigator classroom networks (Stroup et al, 2007)

Year	Median	Mean	S.D.
2007	67.09	68.98	16.48
2008	73.10	70.05	21.37
2009	78.89	75.64	20.74

Table 7. Descriptive statistics of Type 2 (above median) students for cutoff analysis

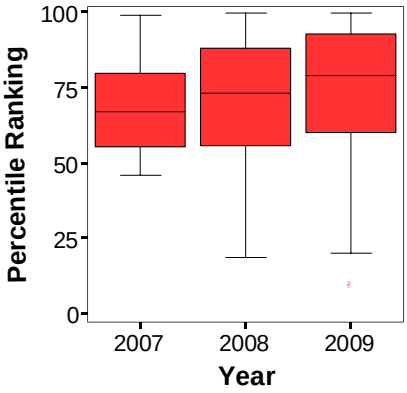

Figure 6. Boxplot of Type 2 (above median) students for cutoff analysis

Year	Mean Rank
2007	1.82
2008	1.92
2009	2.27

N	83
χ^2	9.13
df	2
Asymp. Sig.	0.0104

Table 8. Results of Friedman Test of Type 2 students for cutoff analysis

	2008-2007	2009-2008	2009-2007
Z	-0.55	-2.93	-3.36
Asymp. Sig. (2-tailed)	0.5827	0.0034	0.0008

Table 9. Wilcoxon pairwise post hoc testing of Friedman Test. Individual pairwise test is $\alpha = 0.0167$ for Bonferroni correction and overall $\alpha = 0.05$.

Regression analysis could not be done on cut off data because exploratory data analysis indicated that the residuals were not normally distributed.

Descriptive statistics and the Friedman test indicated that the students as a whole group were increasing their math scores significantly every year, even prior to the implementation of Gear Up III. This was problematic since no performance baseline could be established. These increases were likely the result of a statistical artifact known as regression to the mean (RTM). Under RTM, the further away from the mean students scored, the more likely they would score closer to the mean on the following attempt. Since the students started with a low mean, the probability that they would improve was greater than if they had a higher mean. The presence of a control group would have allowed

for better comparisons to be made about student performance and the learning effects due to Gear Up III.

Regression analysis indicated that students performed similarly each year from the year before; that is, students maintained a relatively stable rank order. High R^2 values of around 70% indicated that the regression fit was very reasonable. The positive intercepts and slopes a little greater than 1 reflected the fact that students improved slightly each year. However, the rates of improvement made were similar before and after the implementation of Gear Up III. Once again, RTM could be obfuscating learning effects due to the intervention. An examination of Figure 4 indicated that the rates of change were close to -1. Students who showed improvement tended to decrease their scores the following year and vice versa. This was a very strong indicator that RTM was taking place.

In order to eliminate RTM as the rationale for student improvement each year, cutoff analysis was done. Cutoff analysis allows for the separation of students who would most likely benefit from RTM (Type 1 students) from those who would most likely be harmed (Type 2 students). By considering these two types of students separately, the effects of RTM could be reduced to draw more meaningful conclusions.

Findings

Cutoff analysis indicated that the Type 1 (below median) students improved between 2007 and 2008 before the implementation of Gear Up III, but did not significantly change between 2008 -2009 when Gear Up III was implemented. The initial improvement was most likely due to RTM. Afterwards no improvement was observed, indicating that Gear Up III did not affect Type 1 students very much, if at all. On the other hand, Type 2 students should have exhibited a decrease in scores due to RTM. Type 2 student scores increased slightly but not significantly between 2007 and 208. However, Type 2 students showed statistically significant increases between 2008 and 2009 during the implementation of Gear Up III. This would indicate that Gear Up III had a positive

effect on students based on their prior performance, which was not due to chance alone. This increase amounted to about 5.5 percentile ranking gain.

Approximately 75% of the population or about 800 students were missing data due to attrition. There was no way to determine that attrition was random and thus the dataset for students with complete data may be suffering from systematic bias. This would make drawing generalized conclusions about the population difficult. For the purposes of discussion, it was assumed that attrition was random and therefore, the dataset was not systematically biased.

Conclusions

Gear Up III had a positive effect on high performing students, with low performing students not benefiting as much. However, even though the low performing students did not seem to benefit from the intervention, they were also not hurt by it. Thus Gear Up III showed promising results in its first year: it was beneficial with high performing students, and it was at least neutral for the low performing students. These results are remarkable since improving the scores of high performing students is extremely difficult, and goes against RTM. Due to the limited nature of the available data for the Gear Up III project, conclusions should be regarded with caution and care taken when making generalizations. A lack of control groups made definitive conclusions difficult. We also should mention that Ernest Educational Concepts was contracted an external consultant to further assist in math curriculum in one of the district in the study.

Further Study

Implementation data from the TI coach for regular classroom visits were not available for analysis. Such data would provide qualitative snapshots of implementation fidelity in the classroom allowing for a more nuanced analysis of how Gear Up III may have affected learning gains. In the future, data for control groups collected from similar nearby districts would allow stronger conclusions

about program impact. Data regarding student groupings into classrooms would also be useful in determining the effect of Gear Up III on learning gains. Lastly, more stringent collection of student data needs to be implemented in order to prevent severe data attrition.

References

Owens, D. T., S. J. Pape, et al. (2008). The Connected Algebra Classroom: A Randomized Control Trial. International Conference on Math Education Research (ICME-11). Monterrey, Mexico: 7

Stroup, W., Pham, V., & Alexander, C. (2007). *Richardson MathForward Project second year final report: Math TAKS results*. Austin, TX: University of Texas.

Student Assessment Division. (2009a). Technical Digests and Reports. In *Texas Education Agency*. Retrieved December 1, 2009, from http://www.tea.state.tx.us/index3.aspx?id=3654&menu_id=793.

Student Assessment Division. (2009b). TAKS™ Region, District, and Campus Level Data Files . In *Texas Education Agency*. Retrieved December 1, 2009, from http://ritter.tea.state.tx.us/student.assessment/reporting/taksagg/dnload.html.