An Orchestra's Guide To The Universe
Pilot Program Evaluation

A Renovation In Music Education (RIME) program
www.rimeusic.org

March 29, 2008

An Orchestra's Guide To The Universe pilot program was funded by an Initiative to Develop Education through Astronomy and Space Science (IDEAS) grant administered by the Space Telescope Science Institute, NASA, the National Endowment for the Arts, and other sources; Arthur Bloom, Principal Investigator (RIME), Dr. Ilana Harrus, Co-Investigator (NASA/GFSC), Dr. Hirokazu Yoshikawa, Co-Investigator (Harvard/NYU).

a) EVALUATION: SUMMARY OF RESULTS

In addition to the success of the implementation of *An Orchestra's Guide To The Universe*, the results of a controlled evaluation of the program, using a rigorous pre- / post-test design with intervention and comparison groups, showed an impressive pattern of positive effects on school achievement. In the evaluation, we assessed multiple domains of school achievement, engagement, and well-being that were related to targeted skills of the program. Students in the intervention school were assessed both before and after the intervention (at the beginning of the school year and at the end). Changes in their skills were compared to those of students at a matched comparison school, who were also assessed at the beginning and end of the school year. The matching appeared to have been successful; that is, students in the intervention and comparison schools did not differ significantly on the assessed pre-test measures.

The evaluation results show that the program significantly increased science, math, and overall grades over the course of the intervention year. In addition, the program had a positive effect on overall academic competence as reported by teachers, and on student-reported grades. The observation of effects on school performance across three types of measures – student-reported grades; teacher-reported academic competence on a teacher survey; and actual school grades across multiple subjects – strengthens our confidence in these results. The positive effects on science and math grades, in particular, reflected central skills targeted in the *Orchestra's Guide To The Universe* program – scientific and critical analysis skills; computational, abstract operational, and memorization skills involved in learning and performing music in professional rehearsal and performance. Finally, the program resulted in higher levels of student-reported liking of African music, likely a reflection of the prominent role an African drummer played in the music curriculum, rehearsals, and final performance with the students.

b) EVALUATION: INTRODUCTION

The evaluation of *An Orchestra's Guide To The Universe* was conducted during the school year of the intervention (2005-2006). The evaluation was directed by Co-I Hirokazu Yoshikawa (currently Professor of Education at the Harvard Graduate School of Education) and Erin Brooke Godfrey (currently a doctoral candidate in community psychology at New York University, and Co-Director of the evaluation), and involved the assistance of several additional doctoral students in psychology at New York University (Amanda L. Roy, Maria Ramos, Anna Gassman-Pines, Nia Ebon West-Bey), NYU undergraduates Ashley Firth and Francesca Luongo, as well as Elisa Rosman, Ph.D., a developmental and community psychologist who received her Ph.D. at NYU.

The aim of the evaluation of *An Orchestra's Guide To The Universe* was to detect its effects on targeted skills in domains of science and music, and more generally in areas of academics and school engagement and performance. The "theory of change" underlying the program encompasses positive effects on multiple domains that are important to school performance – increased efficacy to excel in science and music and more generally in academics; increases in liking of a variety of music genres; increases in ability to think scientifically; positive feelings towards careers in science and music; feelings of attachment to school; increases in self-esteem and psychological well-being; educational expectations and aspirations; and ultimately improvements in school performance itself.

c) EVALUATION: METHODS

The research design for the evaluation was a pre- / post design, with comparison group. The school in which the intervention was implemented was compared to a comparison-group school, which was matched to the degree possible within the school district on a variety of demographic characteristics. A variety of data sources were used, including a 45-minute student survey, teacher reports, school grades, and school test score assessments. This study was approved by the human subjects committee at New York University, as well as by the research and evaluation department of the Prince Georges County Public Schools.

All assessments were conducted twice during the school year, once before the intervention began (the *pre-test*) and once after it had ended (the *post-test*). The pre - and post- test comparison design allows for a more rigorous evaluation design than a post-test-only comparison design. This approach adjusts end-of-year outcomes by the beginning-of-year levels of the same outcomes, and helps to rule out unobserved confounding factors that might otherwise explain the measured changes in student attitudes, behaviors, and school performance.

The intervention school was Berwyn Heights Elementary School. The name of the comparison school is not reported here for confidentiality reasons; it was chosen because of its similarity to Berwyn Heights on aspects of school size; grade structure; fourth- and fifth-grade standardized test scores in reading and math; racial/ethnic composition; and proportion of students eligible for state/federal reduced or free lunch programs. In addition, a school was chosen that was in the same school district, in order to control for district-level policies and procedures that might affect school characteristics. Information from the Prince Georges County Public Schools administrative data (publicly available for the prior academic year) were used to conduct this matching. Once the school with the closest match on these multiple characteristics was chosen, the principal of that school was contacted for approval of the research. That principal agreed; therefore, we did not have to resort to recruiting any of the less closely matched schools in the Prince Georges County Public Schools.

The evaluation included 85 students at Berwyn Heights Elementary School and 115 students from the comparison school. Active parental consent for participation in the evaluation study was obtained, with all research procedures approved by the Prince Georges County Public Schools and the University Committee on Activities Involving Human Subjects of NYU. Recruitment rates for student-reported outcomes were 77% at Berwyn Heights Elementary and 49% at the comparison school. Retention rates were 95% in Berwyn Heights Elementary and 92% at the comparison school. Demographic characteristics of the samples at each school are listed in Table 1.

Table 1: Sample Demographics by School

	Berwyn Heights Elementary	*Comparison School*
	Mean (SD)	
Number of cities lived in in past 3 years	2.00 (1.20)	2.17 (1.34)
Number of study materials at home*	5.46 (1.49)	5.17 (1.52)
Summary index of family resources	2.45 (.62)	1.91 (.72)
Number of years instrument lessons	0.40 (.74)	0.63 (1.27)
Number of years singing lessons	0.91 (1.75)	1.1 (1.75)
	Percentage	
Grade		
5th	65%	70%
6th	35%	30%
Gender		
Girl	56%	57%
Boy	44%	43%
Race/Ethnicity		
White	16%	7%
Black	38%	58%
Latino	31%	30%
Asian/Other	15%	5%
Generation of Immigration		
1st generation	14%	24%
2nd generation	47%	40%
3rd generation	39%	36%
Speak language other than English at home	51%	53%
Ever participated in Colours?†	33%	28%
Ever participated in Instrumental?‡	56%	23%

* Child report measure of SES taken from the World Health Organization

† Colours is an after-school dance and arts program offered to students in the school district

‡ Instrumental is an in-school musical instrument program offered to students in the school district

d) EVALUATION: MEASURES

We collected data on the following constructs. All (with the exception of demographic information, which was not expected to change over the course of the year and was therefore only collected at pre-test) were collected twice, prior to the beginning of the intervention, in October 2005, and after its completion, in May 2006. Student-reported constructs were generally multi-item scales, with strong psychometric properties and good reliability (Cohen's alpha above .70).

An analysis of pre-test differences across the schools found no evidence of statistically significant differences above chance. This analysis was important in demonstrating that the students in the two schools did not differ prior to the intervention in their levels of the outcome measures.

Student-reported measures:

- Self-efficacy in science / astronomy (e.g., "I think I can be good at science")
- Self-efficacy in music (e.g., "when I listen to music, I think "I could do that"")
- Self-efficacy in overall academics (e.g., reporting that one can do school work quickly)
- Intrinsic achievement motivation (e.g., "I try hard at school because I am interested in what I am learning")
- Liking of music genres (popular, jazz, hiphop / rap; Spanish; opera; classical; country; and African), and a count variable of the overall number of music genres the student reported liking
- Attachment to science / astronomy (e.g., "I love to do science projects in class")
- Attachment to and efficacy concerning music (e.g., "I want to be a musician or singer")
- Attachment to school (e.g., "I feel like a real part of my school;" "I love [name of school] School")
- Thinking like a scientist (responses to vignettes asking how scientific different approaches are to finding out particular questions about the world)
- Grades (average across math, reading, science, social studies, and music)
- Self-esteem (e.g., "I am satisfied with myself")
- Depressive symptoms (e.g., "I am sad [once in a while, many times, all the time]")
- Academic expectations (e.g., "How likely do you think it is that you will go to college?")

Teacher-reported measures:

- School engagement (homework, attention)
- School performance (average of writing and math skills)
- Aggressive, antisocial, hyperactive (externalizing) behaviors
- Withdrawn, depressive, anxious (internalizing) behaviors
- Official school grades across the year (quarter 1, prior to the beginning of the intervention in Berwyn Heights; quarter 4, after the end of the intervention), in oral / written, reading, math, science, and music, as well as the overall average of these

Standardized state assessments of school achievement:

- Standardized assessments in reading and math

e) EVALUATION: ANALYTIC APPROACH

We conducted multiple linear regression analyses to estimate the effects of the *Orchestra's Guide To The Universe* program, using a difference-in-difference method to analyze the results. This approach, as mentioned above, adjusts for pre-test scores when analyzing post-test scores. Effects for each construct are interpreted as differences between the intervention and comparison schools in change in student outcomes across the school year. All analyses also adjusted for a set of demographic characteristics, including the grade of the child (5^{th} vs. 6^{th}); gender of the child; ethnicity; generation of immigration; language spoken at home; prior music education involvement (instrumental lessons, chorus, other special music programs in the participating schools); residential mobility; and an index of family resources (across a variety of assets in the home, including # bedrooms, bathrooms, cars, and computers). We used .05 as the level of statistical significance, using a two-tailed test. We report results at the .10 level of significance as associations of marginal or trend significance.

f) EVALUATION: RESULTS

Table 2 below summarizes the significant effects found in our analyses. We report in each row the construct for which a significant association was found, the unstandardized (b) coefficient with significance level, its standard error, and standardized (β) coefficient.

Table 2. Summary of Statistically Significant Effects

Outcome	B (unstandardized coefficient) and significance level	Standard error	B (standardized coefficient)
Student-Reported Grades	.26┼	.14	.17
Liking of African music	.39*	.19	.19
Teacher-reported overall school performance	.36*	.15	.23
School grades: Science	.73***	.24	.32
School grades: Math	.44*	.21	.22
School grades: Overall (average across math, music, oral/written, reading, and science)	.26*	.11	.24

Note: ┼ represents p (statistical significance level) <.10; * p <.05; ** p < .01; *** p < .001.

Overall, the number of constructs for which statistically significant effects were found (5, with an additional one – student-reported grades – at a marginal or trend level of significance) was above chance. That is, positive effects observed were above the levels at which one would expect to find significant effects purely by chance. Moreover, all of the effects found were in the positive direction. This consistent pattern indicates that we can have relative confidence in the results.

The results show that students at Berwyn Heights showed greater improvements in their school performance and grades than students at the comparison school. Improvements were found across three types of measures of school performance – student-reported, a 2-item measure we collected from teachers combining writing and math skills, and the students' actual grades (science, math, and overall grades, which were an average of oral / written, reading, math, science, and music grades). Finding similar patterns across different types of measures and sources of data (students and teachers) strengthens our confidence in these results. These data show that *An Orchestra's Guide To The Universe* appears to have been successful in its ultimate aim of improving children's school performance.

The areas of greatest change in school performance reflect central skills targeted by the music and science components of the *Orchestra's Guide To The Universe* curriculum. The science curriculum emphasized theories of empirical discovery, the historical construction of science, comparison of different theories regarding the universe (such as the Big Bang and other theories of origin), and students' active roles in coming up with new theories and approaches to empirical discovery. Improvements in these targeted skills may have contributed to the improvements in science and math grades observed. The integrated music portion of the curriculum also may have played a role in observed improvements in math skills. The extensive memorization, rehearsals, and performance required a variety of skills linked to mathematics performance (computation, particularly division and multiplication involved in memorizing rhythmic patterns across time; verbal facility; the coordination of action with computational skills, which ultimately occurred under the pressure of professional-level rehearsals and performances).

In addition, students at Berwyn Heights reported greater increases in liking African music. This is likely because the final performance involved an African drummer who gave workshops at the school as part of the music curriculum and was a featured musician with whom the students rehearsed and performed.

The study has several limitations. First, our research design, though rigorous in including a matched comparison school, cannot rule out the role of unobserved factors that may have differed across the two schools. We did, however, conduct an initial analysis comparing the two groups of students on all outcome measures at the pre-test, and found no evidence of significant differences above chance. Second, we cannot tease apart the effects of the music and science components of the curriculum, since they were offered together. *An Orchestra's Guide To The Universe* was designed as an academically challenging program with a close integration between its music and science curricula and preparation for the ultimate performance at the end of the year. Finally, we conducted a post-test, but were not able with limited funds to conduct a long-term follow-up in the evaluation. We therefore cannot answer the question of how long effects observed might have been sustained after the end of the school year. Finally, we should note that our recruitment rate was lower for the comparison school. Although this may have biased the results, if more "prepared" students are those who are more likely to obtain active parental consent, this would have biased the estimates of our effects downward, rather than inflated them.

Despite these limitations, this is the first controlled evaluation of a whole-grade approach to combining music and science education that demonstrates effects in increasing levels of school performance. The innovations of the *Orchestra's Guide To The Universe* program – its targeting of entire grades in a school; its close integration of music and science education in a single curriculum; its culmination in a demanding professional-level rehearsal and performance process – are reflected in the strong and consistent impacts we observed on multiple measures of achievement.