Addressing the Summer Learning Loss:

An Evaluation of the Voyager Summer Reading Intervention Program*

Greg Roberts, Ph.D.
The University of Texas at Austin

Jeri Nowakowski, Ed.D.
Voyager Expanded Learning

*A chapter from
Summer Learning: Research, Policies, and Programs, forthcoming
with preface by Geoffrey Borman, Johns Hopkins University
and Matthew Boulay, Teachers College, Columbia University

Table of Contents From the Forthcoming Book:

SUMMER LEARNING: RESEARCH, POLICIES, AND PROGRAMS

Preface by
Geoffrey Borman, Johns Hopkins University,
and Matthew Boulay, Teachers College, Columbia University

Decades of research have documented the considerable gap in achievement between children attending low-income urban schools and children from more prosperous communities. More recent research has considered possible causes of this gap and potential remedies, including summer school and extended-year programs. The opening to this report briefly summarizes this research and explores its potential impact for urban schools and the children who attend them. The remainder of the chapter considers one application of research in this area. The extended-year program offered by Voyager Expanded Learning is described and the results of recent program evaluation are presented.

SUMMER SCHOOL AND SUMMER LEARNING LOSS

The work of Alexander & Entwistle (1996) and Cooper and his colleagues (Cooper, Charlton, Valentine, & Muhlenbruck, 2000; Cooper, et al., 1996) suggests a need for school-based summer programs that target the summer learning loss of disadvantaged urban school children. Summer school as a policy recommendation may be justified and gathering strength (Borman, 2000), but debate remains over what constitutes an effective summer school program for children attending urban schools. Most agree that any reform strategy must be designed with an urban context in mind, it must be research-based, and it should reflect a systems' orientation (Darling-Hammond, 1993; Fullan, 1991). On the details of such an undertaking, however, there is little consensus and even less research (Borman, 2000; Cooper, et al., 2000). The question remains, "What types of programs are most effective for preventing summer learning loss and closing the achievement gap?"

BACKGROUND OF VOYAGER

Voyager Expanded Learning was founded in 1994 by Randy Best, a corporate business leader; Admiral Tom Hayward, chief naval officer under Presidents Jimmy Carter and Ronald Reagan; and Barbara Nichols, an educator and training professional. Voyager has expanded from elementary after-school and summer school programs to include K–8 programs designed for use during school, after school, between sessions, and during summer school. Recently, several new programs have been launched, including a comprehensive balanced reading program for kindergarten and first-grade students, as well as an after-school and summer intervention reading series for grades 1–8. Voyager's extended-day programs add 144 hours to the school year, while the summer programs add 80 hours for a total of 62% more reading instruction a year.

In 1995, Voyager worked with 11 pilot schools in Dallas. By 1998, it had expanded to include partnerships with 700 sites and 40 states serving approximately 60,000 students. As of June 2000, Voyager summer programs were serving 400,000 students in more than 1,000 districts in 44 states. In fall of 2001, Voyager will launch a Universal Literacy System that combines extended-day and extended-year reading interventions with its K–3 in-school reading program.

Voyager has partnerships and collaborations with the Smithsonian Institution, Discovery Channel, Polaroid, and the Institute for Research on Learning at Stanford University. Additionally, NASA and the University of Oregon provide Voyager with research and technical assistance in the development of new curricula.

Voyager has partnerships and collaborations with the Smithsonian Institution, Discovery Channel, Polaroid, and the Institute for Research on Learning at Stanford University. Additionally, NASA and the University of Oregon provide Voyager with research and technical assistance in the development of new curricula.

THE VOYAGER *TIMEWARP* PROGRAM

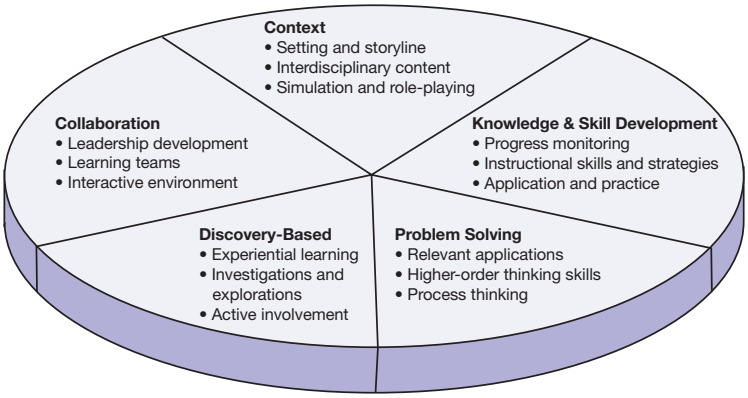
TimeWarp™ is a research-based, strategic summer reading intervention series developed by a team of reading specialists for elementary and middle-school students who have fallen behind. The series is designed to close the achievement gap and prevent summer learning loss as part of an integrated learning system from Voyager Expanded Learning.

The main goal of *TimeWarp*™ is to use available summer hours to improve student performance and increase teacher effectiveness. The series is based on the theory that experiential learning with high interest and academically challenging content will motivate all students to learn and will improve their reading skills and strategies.

The main goal of *TimeWarp* is to use available summer hours to improve student performance and increase teacher effectiveness. The series is based on the theory that experiential learning with high interest and academically challenging content will motivate all students to learn and will improve their reading skills and strategies. *TimeWarp* also actively involves students in leadership roles and cooperative learning through participation as "Team Leaders" and "Pathfinders" (paired partners and allies). Voyager's comprehensive learning framework is shown above. Appendix A on page 12 gives an example of Voyager's Learning Framework applied to the *TimeWarp Egyp*t curriculum.

Chart 1

The Voyager Learning Framework



The TimeWarp model includes four hours of learning activities a day for four weeks in the summer, totaling 80 hours of instruction. Classes of no more than 18 students support the program's collaborative learning and learner-centered curriculum and instruction. Voyager trains teachers on curriculum and effective classroom organizational strategies to ensure faithful program implementation.

Featured components of *TimeWarp* are:

- A restructured classroom that promotes collaborative learning.
- Redefined roles for teachers as co-learners, guides, coaches, and mentors.
- Assessment and evaluation used as a diagnostic tool to facilitate a personalized approach to learning.
- Continuous staff development focusing on standards of authentic instruction.
- Current, interdisciplinary, research-based curricula that are relevant, discovery-based, and focused on critical thinking.
- Curricula aligned with state standards and targeted to reinforce skills based on the results of state and national assessments. (See Table 1 on the following page.)
- Programs that develop students' leadership skills and collaboration that prepares them for citizenship and the real world.

Voyager's training is designed to build instructional capacity at the district and campus levels. Districts select their teachers by expertise, tenure, availability, or other criteria.

Teachers, site directors and district implementation teams attend several days of required training provided by Voyager and delivered by professional educators. This prepares district personnel to implement and support the program with fidelity. Training emphasizes current reading research, Voyager methodology, reading strategies associated with *TimeWarp*, and group management. Throughout the summer program, site directors and district implementation teams provide ongoing training and support for teachers with Voyager's help.

Table 1

Targeted Skills and Strategies

	TimeWarp™ *Egypt* (grades 2–3)	*TimeWarp*™ *Greece* (grades 4–5)	*TimeWarp*™ *The Americas* (grades 6–7)
Concepts of Books and Print			
Conventions of Print	◆		
Book Awareness	◆		
Phonological Awareness			
Sound and Word Discrimination	◆		
Rhyming	◆	◆	
Blending Sounds	◆	◆	
Sound Segmentation	◆	◆	
Letter Awareness and Identification			
Letter Recognition	◆		
Letter Identification and Production	◆		
Alphabetic Principle			
Letter-Sound Knowledge	◆		
Decoding and Word Recognition	◆	◆	◆
Sight-Word Reading	◆	◆	◆
Reading Connected Text	◆	◆	◆
Spelling	◆	◆	◆
Vocabulary Development and Concept Awareness			
Comprehending What Is Read	◆	◆	◆
Literacy Response	◆	◆	◆
Shared and Independent Reading			
Reading for Pleasure and Information	◆	◆	◆
Listening and Speaking			
Listening to Spoken Language	◆	◆	◆
Oral Language	◆	◆	◆
Writing			
Forming Letters and Words	◆		
Mechanics for Writing	◆	◆	◆
Expressing Ideas in Writing	◆	◆	◆

Voyager's training is designed to build instructional capacity at the district and campus levels. Training emphasizes Voyager methodology, reading strategies associated with *TimeWarp*, and group management. Throughout the summer program, site directors and district implementation teams provide ongoing training and support for teachers with Voyager's help.

Voyager Parent Guides for each Voyager student's parents or guardian supply ideas on how to reinforce a child's learning success through supplemental home activities. Students are urged to share what they learn with parents, and parents are asked to make time to hear about their children's daily activities and to attend culminating activities where students demonstrate what they have learned.

A typical TimeWarp day includes:

- Teacher-directed instruction for the whole group to provide instructional focus and encourage reading participation while developing skills and strategies.
- Student learning teams grouped by reading levels. Each team rotates through three learning stations. A student team leader facilitates, allowing teachers to lead challenge lessons to small groups of students who need additional support.
- Response and practice activities to give students a chance to reflect on what they have read and express themselves creatively by writing and creating theme-related projects connecting what they have read to real-world activities.
- A short debriefing period to review what has been learned.

Voyager Parent Guides for each Voyager student's parents or guardian supply ideas on how to reinforce a child's learning success through supplemental home activities. Students are urged to share what they learn with parents, and parents are asked to make time to hear about their children's daily activities and to attend culminating activities where students demonstrate what they have learned.

EVALUATION OF THE VOYAGER SUMMER READING INTERVENTION PROGRAM 2000

Results of the evaluation of the Voyager Summer Reading Intervention Program are meaningful on several levels. First, the findings are of interest to developers and users of the Voyager curricula, and questions related to a program's impact and usability are of importance to its stakeholders. Beyond this, the findings of the Voyager evaluation may help broaden the understanding of what constitutes effective summer school programming.

The meta-analytic work of Cooper and his colleagues (Cooper et al., 2000) highlights the need for considering this issue across a range of populations and settings. A single study may be too limited in scope to address the diversity of questions on this point, and even Cooper's collection of studies left several important questions unanswered. Ongoing and future work in this area should move forward with this larger purpose in mind. A given evaluation can consider the program in question while also contributing to the knowledge about the class of programs to which it belongs. The Voyager evaluation was conducted with this dual purpose in mind.

Evaluation Methods

The primary question of Voyager stakeholders was this: "Is the Voyager summer program, as implemented in schools across the United States, associated with an increase in students' reading skill levels?" Voyager has collected data on program effect at several points in its history, but data have generally been gathered on relatively small samples using Voyager-developed measures. The 2000 project was the second evaluation of Voyager's extended-year program with a national scope, and the first to use a standardized, nationally normed instrument—the Stanford Diagnostic Reading Test-IV (SDRT-IV)—to measure changes in students' reading ability. The 1999 evaluation of Voyager's extended-year program used data collected with a Voyager-developed measure of reading skill level (Voyager, 1999). The instrument was reviewed by reading specialists at the North Central Regional Educational Laboratory for readability, adequacy of format and directions, and quality of test items. Psychometric attributes of the instrument were reviewed by the Measurement and Evaluation

Center at the University of Texas at Austin. The 1999 instrument was used in the 2000 evaluation with approximately 6,500 *TimeWarp* students in Washington, D.C., schools.

Measurement

The SDRT-IV is designed to diagnose students' strengths and weaknesses in the major components of reading. Sections of the SDRT-IV include Phonics, Reading Vocabulary, Reading Comprehension, and Scanning. Results from the SDRT-IV can be used to identify trends in the reading levels of students at the school or district level and to evaluate the effectiveness of instructional programs. Three levels of the SDRT-IV were used in the evaluation. Students in Voyager's *TimeWarp Egypt* program completed the Orange level (grade levels 2.5–3.5). The Purple level (grade levels 4.5–6.5) was used with students in *TimeWarp Greece*, and *TimeWarp The Americas* students took the Blue level (grade levels 9.0–13.0) of the SDRT-IV.

Samples Used in the Evaluation

Participating schools were selected based on two criteria: their anticipated willingness to participate, and their expected level of program implementation. A primary objective was to recruit schools that were willing to participate and that were using the Voyager program according to its design. A second objective of the sample selection process was to represent the diversity of socioeconomic strata (SES) using Voyager programs. Urban schools and suburban schools were selected. A final objective was to sample schools across Voyager's intervention series, including *TimeWarp Greece*, *TimeWarp Egypt*, and *TimeWarp The Americas*. Students participating in the evaluation were characterized as low achieving (the lower quartile). Many were in danger of repeating a grade level.

Identified schools were contacted about their participation. A total of 13 schools agreed to participate. Of these, 7 were located in suburban, rural, or small urban communities and 6 were in large urban districts. There were four middle schools and nine elementary schools in the sample. Five schools were located in the eastern and southeastern United States, five were in the Midwest, and three were in California. Eight of the schools had greater than average levels of economic disadvantage (more than 35% of the school population receiving free or reduced lunch). Seven of the schools had a nonwhite majority student population (greater than 50% nonwhite) and another four had a nonwhite plurality (no ethnic group accounting for more than 50% of the student population).

A total of 400 students from the 13 participating schools completed an SDRT-IV pretest, and 383 completed the post-test. Complete data, pre- and post-test scores on the SDRT-IV, were available for 325 Voyager students. This represents an attrition rate of about 19% (percent of complete cases compared to pretest cases), which is not unexpected given that the Voyager program was delivered during the summer months and a sizable number of participants were lower-SES children living in urban settings. Children leaving the program prior to its conclusion tended to score less well on the pretest SDRT-IV than children with complete data (total scale score of 586 versus 611).

The findings outlined in the following sections are based on the 325 cases for which complete data were available. Attendance was not tracked, so the evaluation findings include the scores of children who attended all Voyager sessions and children who may have attended as few as 2 sessions (the day of pretest and the day of post-test).

A total of 13 schools agreed to participate. Of these, 7 were located in suburban, rural, or small urban communities and 6 were in large urban districts. Eight of the schools had greater than average levels of economic disadvantage.

RESULTS AND FINDINGS

Results from the SDRT-IV are presented using three types of score: effect size, normal curve equivalent, and progress scores. Each of these is addressed in the sections that follow, primarily in terms of Voyager stakeholders' question concerning program impact. Other evaluation questions, particularly those suggested by Cooper's meta-analysis (Cooper et al., 2000), are considered in a subsequent section.

Estimates of Program Effect

Effect-size data. Change in test scores can be expressed in terms of effect size. An effect size (ES) is a value that reflects the standardized amount of change from pre- to post-test. Estimates of effect can be compared across studies and even across programs. By convention, an ES of .20 is considered a small-sized effect, an ES of around .50 indicates a medium-sized effect, and an ES of .80 or greater suggests a large-sized effect. For educational and social programs, effect sizes in the medium-sized range are relatively uncommon. More typical are programs with small-sized effects or no effect.

The overall effect size for the Voyager program (i.e., average effect for all students on the total score for the SDRT-IV) was .42 (40% of one standard deviation), not inconsiderable if one remembers that disadvantaged students typically lose up to a month or more during the summer months (Cooper et al., 1996).

The ES data for the Voyager program suggest considerable change in SDRT-IV post-test scores when compared to pretest scores. The ES were .41, .26, and .55 for *TimeWarp Greece*, *TimeWarp Egypt*, and *TimeWarp The Americas*, respectively. The overall effect size for the Voyager program (average effect for all students on the total score for the SDRT-IV) was .42 (40% of one standard deviation), not inconsiderable if one remembers that disadvantaged students typically lose up to a month or more during the summer months (Cooper et al., 1996). The ES for the subsections of the SDRT-IV ranged from .29 for the Comprehension subtest to .45 for Scanning (Table 2). (As an aside, the Voyager-developed reading test used in Washington, D.C., yielded similar sized effects—about .40—suggesting that the effect may be stable across measures, across implementation, or both.)

Table 2
Traditional Effect Sizes

Voyager Program	Phonics	Vocabulary	Comp.	Scan	Total
TimeWarp Egypt	.32	.46	.27	NA	.41
TimeWarp Greece	NA	.29	.11	.33	.26
TimeWarp The Americas	NA	.18	.42	.55	.55
Totals	.32	.30	.29	.45	.42

Normal curve equivalent scores. The SDRT-IV data were also considered in terms of normal curve equivalent (NCE) scores. NCE scores were originally developed to analyze and report gains in compensatory programs for educationally disadvantaged students. These scores have a mean of 50 and a standard deviation of approximately 21. This results in a scale with 99 equal interval units. A normal curve equivalent score of 50 represents the national average of any grade level at the time of year the test was normed.

Normal curve equivalent scores are similar in range to percentile scores, but with more sophisticated statistical properties, and they generally prove more useful in an evaluation-related context. Normal curve equivalent scores can be used to compute group statistics, compare the performance of students who take different levels of the same test, and compare across different subject matter for the same student.

NCE scores can also be used to evaluate gains over time. NCE is scored in such a way that a year-to-year gain of zero (0) indicates a year's academic growth in the skills measured. A positive gain indicates more than one year's growth; a negative gain indicates less than one year's growth. The average gains for Voyager students were roughly 8 points, 5 points, and 7 points for *TimeWarp Greece*, *TimeWarp Egypt*, and *TimeWarp The Americas*, respectively. Gains of 5 to 8 NCE points are not entirely uncommon, though generally associated with change across the school year. These data suggest that the four-week Voyager program yields gains comparable to those of successful nine-month programs.

Average pre- and post-test scores are indicated for all Voyager programs and for all sections of the SDRT-IV in Table 3. The column labeled % NCE Change suggests the amount of improvement Voyager students experienced when compared to a representative group of same-aged peers; gains are expressed in terms of the percent of one standard deviation. Using NCE data in this way is comparable to calculating an effect size. The primary difference in the ES and the NCE-based ES is the estimate of variance used to standardize the mean difference. In this case, the standard deviation of the NCE scores, approximately 21, was used as the standard unit. NCE scores increased .32 of a standard deviation for Voyager students when compared to the group of students in the SDRT-IV normative sample. Table 3 displays gain scores and effect sizes for groups of Voyager students and sections of the SDRT-IV.

Table 3

NCE Effect Sizes

Voyager Program Level	Avg. Pretest	Avg. Post-test	% NCE Change	# Cases
TimeWarp Egypt				
Phonics	40.0	45.9	28%	106
Vocabulary	34.6	44.5	47%	99
Comprehension	35.7	41.2	26%	102
Scanning	NA	NA	NA	NA
Total	34.8	42.6	37%	96
TimeWarp Greece				
Phonics	NA	NA	NA	NA
Vocabulary	27.5	34.4	33%	85
Comprehension	31.7	33.7	10%	84
Scanning	29.8	35.2	26%	84
Total	26.1	31.3	25%	83
TimeWarp The Americas				
Phonics	NA	NA	NA	NA
Vocabulary	28.3	30.1	9%	124
Comprehension	27.5	33.1	27%	125
Scanning	29.0	36.9	38%	96
Total	27.3	34.0	32%	95
Total Across All Cases				
Phonics	40.0	45.9	28%	106
Vocabulary	30.1	36.1	29%	308
Comprehension	31.3	35.9	22%	311
Scanning	29.4	36.1	32%	180
Total	29.5	36.2	32%	274

NCE is scored in such a way that a year-to-year gain of zero (0) indicates a year's academic growth in the skills measured. A positive gain indicates more than one year's growth; a negative gain indicates less than one year's growth. The average gains for Voyager students were roughly 8 points, 5 points, and 7 points for *TimeWarp Greece*, *TimeWarp Egypt*, and *TimeWarp The Americas*, respectively.

A large percentage of students taking the Orange level of the SDRT-IV (*TimeWarp Egypt* students) moved from below average standing at pretest to at or above average at post-test on the Phonics portion of the test.

Progress indicator scores. The SDRT-IV provides a progress indicator score for the Orange and Purple levels of the test. According to the Teacher's Manual for Interpretation, a progress indicator score can be used to "identify those students who have demonstrated sufficient competence on specific areas to make satisfactory progress in the regular developmental reading program." The cut scores reflect the developmental aspects of learning to read as well as the relative importance of specific areas to the reading process.

A large percentage of students taking the Orange level of the SDRT-IV (*TimeWarp Egypt* students) moved from below average standing at pretest to at or above average at post-test on the Phonics portion of the test. Change was especially notable for test items addressing short vowels (+19%) and long vowels (+9%). On the Vocabulary section of the test, the percentage of students moving to at or above average ranged from 7 to 14 percentage points on the Orange level of the test and 1 to 7 percentage points on Purple (*TimeWarp Greece*). In Comprehension, as many as 16% of students moved from below average to at or above average (paragraphs with questions, textual reading, and interpretation).

Table 4 presents the percentage of students at or above average (at or above the cut score) at pre- and at post-test on the different skill elements included in the comprehension subtest of the SDRT-IV. The progress indicator scores support the conclusions suggested by the effect-size data and the NCE data; the Voyager program had a considerable impact on students at participating schools.

Table 4
Progress Scores for Comprehension Subtest

Comprehension	*TimeWarp Egypt*		*TimeWarp Greece*	
	Pretest	Post-test	Pretest	Post-test
Cloze	76%	81%	NA	NA
Paragraphs W/Questions	43%	59%	NA	NA
Recreational Reading	62%	74%	51%	52%
Textual Reading	38%	54%	61%	63%
Functional Reading	50%	60%	35%	43%
Initial Understanding	48%	50%	51%	58%
Interpretation	46%	62%	44%	52%
Critical Analysis & Process Strategies	61%	75%	NA	NA
Critical Analysis	NA	NA	47%	50%
Process Strategies	NA	NA	62%	67%

Differences in evaluation method can confuse findings related to more substantive factors, like program, student, and outcome characteristics. When the focus is across a number of related studies, as in meta-analysis, it is often possible to control for differences in method and analyze the effect of these more substantive elements. This was the case in Cooper's data; he was able to adjust the effect size of studies in his sample for differences due to method, thereby eliminating the confound of methodological variables and substantive

variables. (This does not eliminate the confound due to intercorrelation among substantive variables.) That said, it should be noted that the Voyager effect-size estimates for the set of moderators in Table 4 are not adjusted in this way (nor could they be without including the Voyager data in a re-analysis of the Cooper studies). If apples are to be compared to apples, the unadjusted Cooper estimates should be used. The oranges (the adjusted effects) are included as an additional point of reference.

Comparison with Results of Cooper et al.'s Meta-Analysis

Cooper et al.'s meta-analysis of summer school effects provides an insightful summary of the collective knowledge in this area. Their findings also offer a meaningful backdrop against which to consider the results of the Voyager program evaluation. The remainder of this chapter is devoted to this task.

The mean program effect across the 99 independent samples in Cooper's study was .26 (assumes a fixed-effects model), and the median effect was .19. The values reflect the impact of all extended-year programs in Cooper's sample; regardless of type, summer school increases student outcomes by about .25 of a standard deviation when compared to students' average pre-program score or the average score of students not attending summer school.

The Voyager program effect across all curricula and for the Total score on the SDRT was .42 when post-test scores were compared to pretest scores and .32 when gain scores were considered in terms of NCE units. The Voyager pre- to post-test ES estimate (.42) is 62% greater than the Cooper average, and it ranks in the upper 25% of Cooper's distribution of effect-size values. Cooper's effect sizes range from -.22 to 2.7, and most cluster around the median (.19) with almost one-half (47%) falling between .01 and .29.

Variables moderating effect sizes. Cooper's meta-analysis considers a number of moderating variables, which he organizes according to several categories including student characteristics, program context, methodological features, and program features. Table 5 displays the first two of these categories and, within each, several moderators that are relevant to a discussion of the Voyager results. Methodological features are not included in Table 5, though they are relevant to the present topic and warrant brief mention as well.

Table 5

Progress Scores for Comprehension Subtest

	Cooper		Voyager
	Unadjusted	Adjusted	
STUDENT CHARACTERISTICS			
Grade Level			
K–3	.24	.19	.41
4–6	.19	.17	.26
7–12	.29	.35	.55
PROGRAM CONTEXT			
Type of Community			
Large Urban	.28	.29	.35
Other	.38	.34	.42

The Voyager program effect across all curricula and for the Total score on the SDRT was .42 when post-test scores were compared to pretest scores and .32 when gain scores were considered in terms of NCE units. The Voyager pre- to post-test effect-size estimate (.42) is 62% greater than the Cooper average, and it ranks in the upper 25% of Cooper's distribution of effect-size values.

The percent difference (difference in Other and Large Urban) in the Voyager data differed somewhat from Cooper's data. In Cooper's sample, Other was 36% larger than Large Urban, while in the Voyager sample, Other was 20% larger. The gap between large urban schools and schools in suburban and rural districts was less prominent in the Voyager data than in Cooper's study.

Methodological features. Cooper's ES estimate for programs evaluated using a one-group pre- to post-test design was .24. When Voyager post-test scores were compared to pretest scores, the estimated effect size was .42, as indicated earlier. For programs in Cooper's sample using a two-group design, the effect size was .07. Of these, studies using randomized treatment and control groups had an average effect of .14, while nonequivalent-group studies had an estimated effect size of .05 (these values do not significantly differ from each other). The Voyager NCE-based effect size involved comparing Voyager students to the normative sample of the SDRT. This two-group design is similar in respect to studies in Cooper's set of nonequivalent-group evaluations, suggesting that the .32 Voyager estimate can be evaluated in relation to Cooper's value of .07.

Student characteristics and program context. The effect sizes for different-aged Voyager students were .41, .26, and .55 for primary, intermediate, and middle school students, respectively. This compares to effect sizes of .24, .19, and .29 for similar-aged students in Cooper's sample. Cooper's data suggest that age and the effect of summer school may be U-shaped in nature. Summer programs tend to be more helpful for younger and older students and less effective for students falling between the two. Cooper speculates that the curvilinear relationship of age and summer school effect may be due largely to differences in the instructional approaches used with the different age groups. The Voyager data do not support this assertion. Voyager's program is very similar in philosophy and instructional focus across its different levels, suggesting that the U-shaped nature of the Voyager data may not be entirely due to grade-related instructional differences. This question deserves further attention.

Program Context, the other category in Table 5, considers attributes of the community in which the program was delivered. In Cooper's sample, the effect sizes were .28 and .38 for Large Urban and Other categories, respectively. The effect sizes for the Voyager data are .35 and .42, for Large Urban and Other, respectively, where Other includes suburban and rural school districts. The trend in the Voyager data is in the same direction as the Cooper data; effects of summer school were greater in schools other than those in large urban districts. The percent difference (difference in Other and Large Urban) in the Voyager data differed somewhat from Cooper's data, however. In Cooper's sample, Other was 36% larger than Large Urban, while in the Voyager sample, Other was 20% larger. The gap between large urban schools and schools in suburban and rural districts was less prominent in the Voyager data than in Cooper's study.

Many of the methodological, student, and programmatic attributes discussed as moderators in Cooper's meta-analysis did not vary in the Voyager evaluation because they are parameters of the program model. For instance, Voyager is generally delivered to classes of 20 students or fewer for 80 instructional hours. While it is possible to compare the total program effect for Voyager to Cooper's moderated effects, such comparisons are perhaps less meaningful than ones that allow analysis within and across Voyager and Cooper (the comparisons outlined in the foregoing section). Readers interested in comparisons not presented in this chapter are referred to the Cooper et al. (2000) study.

DISCUSSION

The Voyager program had an effect on the reading ability of students across the U.S. during the summer of 2000. The size of the Voyager effect can be described as considerable; it is in

the upper quartile of program effects reported by Cooper et al. (2000) and appears to be relatively stable across Voyager program sites and/or instruments used to measure skill gains.

Several questions are suggested by the evaluation findings for Voyager. Notable in this respect is the question related to the internal validity of studies in this area of research. A randomized-groups design was not used in the Voyager evaluation, and the results should be considered with this in mind. At the same time, alternative explanations for the effects evident in these data are less than compelling. Participants were not attending school while participating in Voyager, thus minimizing the possibility of school effects as a causal factor. Maturation effects seem unlikely because the Voyager program lasted about a month, a relatively brief span of time. The Voyager program is the most likely candidate for explaining the effect.

In Cooper's sample, a one-group design is associated with considerably greater effects than a two-group study, suggesting this question, "Aside from maturation and school effects, what factors other than Voyager may have contributed to the difference in pre- and post-test scores?" and the more general question, "What threatens the internal validity of summer school studies with nonrandomized groups designs?" Addressing these questions will require further meta-analysis using program data coded according to type of validity threat. Again, this highlights the earlier-mentioned importance of considering program evaluation at various levels. The program in question is but one focus for evaluators working in this area.

Future and ongoing evaluation of the Voyager program may also shed light on this point. Multi-group studies of Voyager's long-term effect are presently underway in Washington, D.C., and Tulsa, Oklahoma, school districts. These evaluations will consider several issues. Along with the question related to internal validity threats, the data will address the sustainability of the Voyager program effect, the effect of combining Voyager summer school with the Voyager after-school program, and the effect of attending Voyager over consecutive summers. The goal is to gather solid data that will help to improve Voyager programs while also providing input to the wider community of researchers and service providers.

The notion of applying an isolated set of evaluation findings beyond the boundaries of a program is a hallmark of theory-driven models of program evaluation, an approach that will continue to inform the evaluation of Voyager programs. Theory-driven approaches assume that evaluation of an effective program built upon well-tested psychological, social, and learning theories is enhanced when those same theoretical frameworks are embedded in the evaluation design. The research-based Voyager program is built upon recent work on effective instruction, student learning, and the essentials of beginning reading. Ongoing evaluation will consider ways of further enhancing these research-based elements in the Voyager programs.

The summer implementation of Voyager embodies many of the factors that recent research suggests are critical to effective use of and positive results for extended-year programs (see Borman, in press). Summer programs are most effective when they involve parents, remain small in size, undergo careful scrutiny for treatment fidelity, contain substantial academic components aimed at teaching reading and math, and coordinate summer school

The Voyager program had an effect on the reading ability of students across the U.S. during the summer of 2000. The size of the Voyager effect can be described as considerable; it is in the upper quartile of program effects reported by Cooper et al. (2000) and appears to be relatively stable across Voyager program sites and/or instruments used to measure skill gains.

experiences with those that occur during the regular school year. Voyager programs include all of these identified components. Ongoing evaluation of Voyager's summer program will model the inter-relationships of these critical components and evaluate their relative effect in improving student achievement and narrowing the gap in achievement.

Summer programs are most effective when they involve parents, remain small in size, undergo careful scrutiny for treatment fidelity, contain substantial academic components aimed at teaching reading and math, and coordinate summer school experiences with those that occur during the regular school year. Voyager programs include all of these identified components.

Appendix A

VOYAGER LEARNING FRAMEWORK EXAMPLE USING *TIMEWARP EGYPT* CURRICULUM

TimeWarp Egypt is an example of how the curricula use Voyager's framework of learning in context, knowledge and skill development, problem-solving, discovery-based instruction, and opportunities for collaboration among students.

Learning Context. In the *TimeWarp Egypt* curriculum, Voyager students ("Voyagers") search for their Time Tracker counterparts—fictional students who disappeared through a time warp to ancient Egypt. In order to bring the Time Trackers back home, students study archaeology, history, geography, and fine arts, as well as literacy skills and strategies. Voyagers learn about Egyptian hieroglyphs, ancient etiquette and fashion, and mummies and the mummification process.

Knowledge and Skill Development. Teachers monitor the progress of Voyagers using the pre-assessment tool, periodic performance assessments, observations, student reflections, and the post-assessment tool. Instructional skills and strategies are targeted to help students comprehend text, analyze words, utilize graphic sources, and write to communicate with others. The curriculum correlates to the highest state and national grade-level standards.

Multiple opportunities for application and practice are supplied through the curriculum's instructional framework, including Instructional Focus, Reading Experiences, Learning Stations, teacher-directed Challenge Lessons, Word Making and Word Sorting, and Response sections.

Problem Solving. Voyagers learn problem-solving skills by investigating how the Egyptians may have moved massive stones to construct pyramids and which items were considered essential for a pharaoh in the afterlife. Higher-order thinking skills are strengthened by practicing metacognitive reasoning in "Think-Alouds," formulating advice for the Time Trackers, and evaluating Egyptian toys and games.

Students learn process thinking as they record the details of the Time Trackers' mysterious disappearance, chart the steps in mummy making, and follow written steps to simulate an authentic archaeological dig.

Discovery-Based Learning. Experiential learning is integrated throughout *TimeWarp Egypt* as Voyagers take on the roles of archaeologists and Egyptologists. Students get their hands wet constructing a model of the River Nile, an Ancient Splendors videotape takes them on a virtual field trip to Egypt, and Ancient Egyptian Bingo helps them learn to use context clues. Voyagers are actively involved in every segment of a lesson, often working in small learning teams and engaging in interactive discussions.

Collaborative Learning. Voyager's format allows students opportunities to develop collaborative and leadership skills throughout the curriculum. The Voyager teacher's role is to act as a guide or facilitator, encouraging leadership and collaboration among students. Students work in pairs and act as "Pathfinder" partners for each other. Small-group learning teams help students learn to work effectively with others. Voyagers are grouped and regrouped to gain experience, and they may become "Team Leaders" who direct their groups through an activity.

BIBLIOGRAPHY

Alexander, K. (2000, July). Paper presented at the First National Conference: Summer Learning and the Achievement Gap, Baltimore, MD.

Alexander, K.L., & Entwistle, D. (1996). Early schooling and educational inequality: *Socioeconomic disparities in children's learning.* In J. Clark (Ed.), James S. Colemen. London: Falmer Press.

Borman, G.D. (in press). Summers are for learning. *Principal Magazine.*

Borman, G. (2000). The effects of summer school: Questions answered, questions raised. *Monographs of the Society for Research in Child Development, 260, 65*, 1–117.

Cooper, H., Charlton, K., Valentine, J., & Muhlenbruck, L. (2000). Making the most of summer school: Meta-analytic and narrative review. *Monographs of the Society for Research in Child Development, 260, 65*, 1–117.

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research, 66*, 227–268.

Darling-Hammond, L. (1993). Reframing the school reform agenda: Developing capacity for school transformation. *Phi Delta Kappan, 74*, 52–61.

Feldman, S., & Elliott, G. (1990). *At the threshold: the developing adolescent.* Cambridge, Massachusetts: Harvard University Press.

Fullan, M. (1993). Getting Reform Right: What Works and What Doesn't. Phi Delta Kappan, 73, 744–752.

Jencks, C., & Phillips, M. (1998). *The Black-White test score gap.* Washington D.C.: Brookings Institution Press.

National Center for Educational Statistics. (1997).

Paris, S. (2000, July). Paper presented at the First National Conference: Summer Learning and the Achievement Gap, Baltimore, MD.

Phillips, M., Crouse, J., & Ralph, J. (1998). Does the black-white test score gap widen after children enter school? In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap.* Washington D.C.: Brookings Institution Press.

Roderick, S. (2000, July). Paper presented at the First National Conference: Summer Learning and the Achievement Gap, Baltimore, MD.

Voyager Expanded Learning (1999). *Solutions that impact performance: Voyager research base and results.* Dallas, Texas: Author.

Voyager's format gives students opportunities to develop collaborative and leadership skills throughout the curriculum. The Voyager teacher's role is to act as a guide or facilitator, encouraging leadership and collaboration among students. Students work in pairs and act as "Pathfinder" partners for each other. Small-group learning teams help students learn to work effectively with others.