AN ANALYSIS OF THE EFFECTS OF AN ACADEMIC SUMMER PROGRAM FOR MIDDLE SCHOOL STUDENTS

Marie-Andrée Somers
Rashida Welbeck
Jean B. Grossman
Susan Gooden

mdrc
BUILDING KNOWLEDGE TO IMPROVE SOCIAL POLICY

March 2015

An Analysis of the Effects of an Academic Summer Program for Middle School Students

Marie-Andrée Somers
Rashida Welbeck
Jean B. Grossman
Susan Gooden

March 2015

Funders

This report is funded by the Edna McConnell Clark Foundation's Social Innovation Fund (SIF), a key White House initiative and program of the Corporation for National and Community Service (CNCS). The Edna McConnell Clark Foundation's Social Innovation Fund includes support from CNCS and 15 private co-investors: The Edna McConnell Clark Foundation, The Annie E. Casey Foundation, The Duke Endowment, The William and Flora Hewlett Foundation, The JPB Foundation, George Kaiser Family Foundation, The Kresge Foundation, Open Society Foundations, Penzance Foundation, The Samberg Family Foundation, The Charles and Lynn Schusterman Family Foundation, The Starr Foundation, Tipping Point Community, The Wallace Foundation, and Weingart Foundation.

The broader dissemination of MDRC publications is supported by the following funders that help finance MDRC's public policy outreach and expanding efforts to communicate the results and implications of our work to policymakers, practitioners, and others: The Annie E. Casey Foundation, The Harry and Jeanette Weinberg Foundation, Inc., The Kresge Foundation, Laura and John Arnold Foundation, Sandler Foundation, and The Starr Foundation.

In addition, earnings from the MDRC Endowment help sustain our dissemination efforts. Contributors to the MDRC Endowment include Alcoa Foundation, The Ambrose Monell Foundation, Anheuser-Busch Foundation, Bristol-Myers Squibb Foundation, Charles Stewart Mott Foundation, Ford Foundation, The George Gund Foundation, The Grable Foundation, The Lizabeth and Frank Newman Charitable Foundation, The New York Times Company Foundation, Jan Nicholson, Paul H. O'Neill Charitable Foundation, John S. Reed, Sandler Foundation, and The Stupski Family Fund, as well as other individual contributors.

Opinions or points of view expressed in this document are those of the authors and do not necessarily reflect the official position of — or a position that is endorsed by — the CNCS, the Social Innovation Fund program, or other MDRC funders.

For information about MDRC and copies of our publications, see our website: www.mdrc.org.

Overview

This report examines the implementation and effects of the academic summer program for middle school students offered by Building Educated Leaders for Life (BELL). BELL's middle school program serves rising sixth- through eighth-grade students who are performing one to two years below grade level. The goals of the program are to increase students' literacy and math skills and to enhance their social development. To achieve these goals, BELL provides students with 6.5 hours of daily programming for approximately five weeks, five days per week. Several types of activities are provided: academic instruction in math and English Language Arts; social and academic enrichment activities; and field trips, guest speakers, and community service. BELL's contributions to summer learning began with its now well-established program for elementary school students. More recently, growing demand for programs serving older students has led BELL to expand into middle school.

In this study, which is funded by the Edna McConnell Clark Foundation's Social Innovation Fund, the impact of BELL's middle school program was evaluated using a random assignment research design — a lottery-like process used to assign eligible students either to a program group that was invited to participate in the BELL program or to a control group that was not. The study was conducted in summer 2012 in three school districts that were new partnerships for BELL. Due to various challenges related to student recruitment, the study's sample size is smaller than planned, and the margin of error around the impact findings is quite large. Even so, the results in this report can still be useful for generating suggestive or preliminary evidence about the potential effects of a full-day, academically oriented summer program model for middle school students.

Overall, the findings from this study indicate that BELL mounted a fairly well-run and well-staffed five-week summer program in summer 2012 and that students attended at a high rate even though the program was voluntary. The pattern of impact estimates suggests that, on returning to school in fall 2012, BELL students may have had stronger math skills than they would have had otherwise — equivalent to a little over one month of learning, which is the effect that one would expect from a five-week program during the regular school year. Though the magnitude of this effect is not statistically significant, it is similar in size to what has been found in prior evaluations of voluntary summer programs at the elementary school level. On assessments of reading skills, however, there is no indication that the BELL students outscored their counterparts in the non-BELL group.

Taken together, the findings provide suggestive preliminary evidence that voluntary academic summer programs can have positive effects on middle school students' math achievement but that improving their reading achievement is a more challenging task because it is harder to keep students in this age group engaged. While additional research would be required to confirm these preliminary findings, if true, this suggests that strategies for teaching reading skills to middle school students may need to be different than the approaches used with elementary school students.

Contents

List of Exhibits

Table

Table

Table

Figure

Box

Preface

Far too many children living in economically disadvantaged households are below grade level academically. While economically advantaged families can step in and provide needed academic support to their struggling children, this type of support is far less available to children in underserved neighborhoods. As a result, many school districts turn to programs, such as Building Educated Leaders for Life (BELL), to offer free summer services to students. The summer — a time when students have many free hours to fill — offers a perfect opportunity for schools to provide more instruction and, hopefully, improve students' academic outcomes.

Founded in 1992, BELL has been a pioneer in providing rigorous academic services during the summer to children living in low-income urban communities. It has also been a pioneer with respect to unflinchingly using data to examine and improve its program. For example, during summer 2005, BELL's elementary school summer program was evaluated using the most rigorous methodology: a randomized controlled trial. Because the elementary school program emphasizes reading, only reading (not math) was assessed. The evaluation found that BELL had a positive effect on elementary school students' reading ability.

Buoyed by these findings, and given the growing demand for middle school programs, BELL began to expand into middle school. Although it had good evidence indicating that the elementary school program was effective, it did not know whether its middle school program would be equally successful. Thus, in summer 2012, BELL embarked on a randomized controlled trial to evaluate its middle school program. To date, there has been very little evidence on the effectiveness of summer academic programs for middle school students, especially programs in which participation is voluntary. Thus, the present study is important not only to BELL but also to leaders of other middle school summer programs. The report concludes by offering lessons about implementing academic summer programs for middle school students and by making recommendations for further study.

Gordon L. Berlin
President

Acknowledgments

This report is based on work supported by the Social Innovation Fund (SIF), a key White House initiative and program of the Corporation for National and Community Service (CNCS). The Social Innovation Fund combines public and private resources with the goal of increasing the impact of innovative, community-based solutions that have compelling evidence of improving the lives of people in low-income communities throughout the United States. The Edna McConnell Clark Foundation's SIF includes support from CNCS and 15 private co-investors: The Edna McConnell Clark Foundation, The Annie E. Casey Foundation, The Duke Endowment, The William and Flora Hewlett Foundation, The JPB Foundation, George Kaiser Family Foundation, The Kresge Foundation, Open Society Foundations, Penzance Foundation, The Samberg Family Foundation, The Charles and Lynn Schusterman Family Foundation, The Starr Foundation, Tipping Point Community, The Wallace Foundation, and Weingart Foundation. This report would not have been possible without these organizations' support and commitment to the well-being of youth in low-income communities in the United States.

We owe special thanks to BELL's national and local staff, whose dedication to helping children and to continuous program improvement allowed us to implement the evaluation with rigor and integrity. We are especially indebted to Tiffany Cooper Gueye, Lauren Gilbert, Bryan Hall, and the rest of BELL's national leadership team for their ongoing support and perceptive insights about the results of the study.

We thank the following people at CNCS, The Edna McConnell Clark Foundation, The Wallace Foundation, MDRC, and other organizations for their thoughtful comments about the study's findings and this report: Rob Ivry, Fred Doolittle, John Hutchins, Pei Zhu, Gabriel Rhoads, Albert Chung, Elizabeth Reisner, Bob Granger, Lily Zandniapour, Nicole Vicinanza, Duncan Chaplin, Leslie Goodyear, Priscilla Little, and Beth Miller. At MDRC, we also thank Monica Cuevas, Alyssa Ratledge, Zachary Pinto, and Kateryna Lashko for their contributions to the project. Robert Weber edited the report, and Stephanie Cowell and Carolyn Thomas prepared it for publication.

The Authors

Executive Summary

The middle school years are a critical turning point for youth educationally. Numerous studies have shown that students' success in grades 6 to 8 has profound implications for their future.[1] Attendance, grades, and test scores during the middle school years all predict students' odds of graduating from high school, which, in turn, predicts future earnings.[2] Yet teaching middle school students who are behind is notoriously difficult because of the developmental changes that occur during this period.[3] After years of relatively stable growth, middle school students begin to experience dramatic changes cognitively, physically, socially, and emotionally. Finally, in conjunction with all these struggles, middle school students are also striving to have more autonomy in their relationships with adults, especially with their parents. It is no wonder that middle school has been called the "Bermuda Triangle of education."[4]

Despite the difficulty of the task, strong pressure to perform well on standardized tests has led more school districts to respond to the struggles of their middle school students by providing them with extra help over the summer to enable them to start the new year with stronger basic skills. Some superintendents have made summer school mandatory for students who score particularly poorly on critical tests. Others, worried about discipline and the engagement of mandated students, strongly encourage struggling students to attend voluntarily. While there are many studies of elementary summer school programs (including some that have found positive impacts), there are few studies of the impact of summer programs on middle school students. This is problematic, given that summer school is a costly endeavor and districts are operating with increasingly tight budgets.

This report presents the findings from a study of the middle school academic summer program offered by Building Educated Leaders for Life (BELL). BELL's middle school program serves rising sixth- through eighth-graders who are identified by their school as performing one to two years below grade level, on average. The program operates five days a week for approximately five weeks during the summer. The program day is a traditional "full day" (6.5 hours), in which the morning is devoted to math and reading instruction and the afternoon provides enrichment through instruction in science, physical education, the arts, and other creative subjects — except on Fridays, when there are guest speakers or field trips.

[1] See, for example, Reyes, Gillock, Kobus, and Sanchez (2000); Roderick (1995); Balfanz (2009); Balfanz, Herzog, and Mac Iver (2007); and Kieffer and Marinell (2012).

[2] Levin, Belfield, Muenning, and Rouse (2007).

[3] Eccles (1999), p. 36.

[4] Juvonen et al. (2004), p. xv.

BELL also operates an elementary school summer model, and an earlier randomized controlled trial of this program concluded that it had improved younger students' reading achievement by the equivalent of about one month of learning.[5] Given the growing demand for middle school programs, BELL has more recently expanded into serving middle school students. As a mission-driven learning organization, BELL decided to rigorously investigate whether its middle school model was as effective as its elementary school model, by participating in another randomized controlled trial. This study — which is funded by the Edna McConnell Clark Foundation's Social Innovation Fund (SIF) — provides a unique opportunity to gain a better understanding of the potential effects of full-day academic summer programs for middle school students.

In this study, the impact of BELL's middle school model is evaluated using a random assignment research design, which is the most rigorous type of design for evaluating program effects. A lottery-like process was used to determine which eligible students would be invited to participate in the BELL middle school program (the BELL group) and which students would not be invited to participate in BELL (the non-BELL group). Importantly, because admission to the program was determined using random assignment, students in the BELL and the non-BELL groups were comparable with respect to their motivation and ability at the start of the program. This means that any subsequent difference between the two groups with respect to academic outcomes in the fall after participating in the program can be attributed to the impact of the BELL program.

Despite its rigorous research design, this study has three important limitations that affect the generalizability and statistical power of its findings:

- **The study is underpowered.** Due to various challenges related to student recruitment, the margin of error around the impact findings from this study is quite large. Therefore, even though random assignment ensures that the study provides an unbiased picture of how BELL and non-BELL students differed at the end of the summer, these effects are unlikely to be statistically different from zero unless they are large in magnitude — much larger than would be expected from a five-week summer program. (For its effects to be statistically significant, BELL's five-week program would have to be three times more effective than five weeks of regular schooling or three times more effective than previously evaluated academic summer programs.)

- **The study districts may not be representative of BELL's other middle school sites.** Given the eligibility criteria for the study, the school districts

[5]Chaplin and Capizzano (2006).

that are included in this evaluation ended up being new partnerships for BELL in summer 2012, and they operated voluntary (rather than mandatory) summer programs. It is difficult to determine how these two district features affect the generalizability of the study's findings to BELL's more experienced middle school sites and/or to sites where student participation was mandatory.

- **The program has evolved since the evaluation.** This study is an evaluation of BELL's middle school model as it existed in summer 2012. As an organization that embraces continuous improvement, BELL has made changes to its middle school model since then, most notably with respect to staff training and the math and reading curricula that are used for instruction. Thus, the findings presented in this report may not generalize to the impact of BELL's middle school model in its present form.

Given these limitations, the present study of BELL's middle school program cannot provide a definitive or generalizable answer about the impact of summer programs for middle school students. Because of random assignment, however, the study's findings are unbiased; therefore, the results in this report can still be useful for generating *preliminary* evidence about the potential effects of middle school summer programs and for understanding the environment in which such programs operate. One goal of this report is to look for consistent patterns in the direction and magnitude of BELL's effect on students' summer activities and their academic outcomes in the fall. The report also analyzes impacts and program implementation by school district, to explore whether particular features of implementation might be associated with more positive effects. Such analyses can be useful in generating strategies for building stronger summer learning programs for middle school students.

Overall, the findings from this study indicate that BELL mounted a fairly well-run and well-staffed five-week summer program in summer 2012 and that students attended at a high rate, even though the program was voluntary. The pattern of impact estimates suggests that, on returning to school in fall 2012, BELL students may have had stronger math skills than they would have had otherwise — equivalent to a little over one month of learning beyond what was achieved by students in the non-BELL group. Although this effect is not statistically significant, its size is what one would expect from a five-week program during the regular school year. Its size is also similar to what has been found in prior evaluations of voluntary summer programs at the elementary school level. On assessments of reading skills, however, there is no indication that the BELL students outscored their counterparts in the non-BELL group.

Taken together, the findings provide suggestive preliminary evidence that BELL's voluntary academic summer programs could have positive effects on middle school students' math

achievement but that improving their reading achievement is a more challenging task because it is harder to keep students in this age group engaged. While additional research would be required to confirm these preliminary findings, if true, this suggests that strategies for teaching reading skills to middle school students may need to be different than the approaches BELL used with elementary school students. For instance, the content of reading materials may need to be tailored explicitly to the needs and interests of young adolescents, to keep them engaged.

The BELL Middle School Model

The goals of the BELL middle school program are to increase children's literacy and math skills by providing them with engaging and age-appropriate instruction and to enhance their social development by giving them opportunities to be successful and to experience the broader community.

To achieve these goals, BELL provides middle school students with 6.5 hours of daily programming for approximately five weeks, five days per week. During this time, several types of activities are provided to students: academic instruction in math and English Language Arts (ELA); social and academic enrichment activities; community time; and field trips, guest speakers, and community service.

Instruction occurs Monday through Thursday mornings and is provided by a certified English Language Arts (ELA) or math teacher and an assistant (called a "mentor"). BELL academic teachers are certified teachers, and they receive training prior to the beginning of the program. In summer 2012, teachers received one full day of in-person training and were expected to complete nine hours of online training before the start of the program.

In any given week, students receive six hours of ELA and math instruction (twelve hours total). Monday through Thursday mornings, students receive an hour of literacy instruction and an hour of math instruction each day. During the week, students also participate in two hours of project-based literacy activities, anchored by a novel or writing assignment, and two hours of project-based math activities. In total, across all five weeks of the program, students are offered 30 hours of ELA instruction and 30 hours of math instruction.[6] In summer 2012, the literacy curriculum was Houghton Mifflin Harcourt *Summer Success,* and the math curriculum was *On Core,* a new Common Core State Standards (CCSS) curriculum.

Because the program is remedial and is intended to help students catch up if they are below grade level, teachers cover material from the prior school year. To help each class stay on

[6]Each week for five weeks, students receive six hours of instruction per subject area, for a total of 30 hours per subject area.

track with the learning objectives, teachers are given a pacing guide that shows them the material that they should be covering each week. Students' reading and math skills are also tested at the beginning of the five-week program, to help teachers assess the strengths and weaknesses of each student, and then are tested again at the end of the program so that changes in students' test scores can be measured and reported to the district. In summer 2012, BELL used the Stanford Diagnostic Reading Test and the Stanford Diagnostic Math Test for diagnostic assessments.

Monday through Thursday afternoons, students participate in two hours of fun and engaging social or academic enrichment activities to broaden their interests, develop positive teamwork and leadership skills, and allow them to discover and demonstrate their strengths in different ways. The enrichment activities are either designed by teachers (such as playing steel drums, cooking, or journalism), are requested by the district, or are grade-specific thematic enrichment curricula offered by BELL. On Fridays, students participate in field trips and community service projects — and, in some sites, attend guest lectures by community leaders — to broaden their interests and extend their learning beyond the classroom.

To achieve its goals, BELL aims to hire staff who will be strong positive adult role models. At each school, the operation of the BELL program is overseen by a program manager (who is typically a principal or assistant principal in the district during the regular school year), an assistant program manager, and a lead teacher who acts as a resource for teachers and their teaching assistants.

As noted above, the BELL middle school model has evolved since the time of this evaluation. The structure of the program and the amount of instruction provided remain the same, but some of the features related to instructional quality — most notably, the curriculum and the way in which teacher training is provided — have changed since summer 2012. A description of how the model has changed is provided at the end of this Executive Summary.

Overview of the Study's Design

The present study of BELL's middle school program was conducted in summer 2012 in three school districts. (Box ES.1 presents an overview of the study's key features.) Districts that partner with BELL usually have more eligible students than BELL has the capacity to serve. In a typical summer, BELL fills its limited program slots on a "first-come, first-served" basis. For the purposes of this study, however, random assignment was used to select which students would be admitted to BELL. To make this possible, schools in the study continued to identify students who were below grade level and to encourage applications from these students until shortly before the start of the program. A lottery-like process was then used to determine which students would be invited to participate in the BELL middle school program (the BELL group)

Box ES.1

Overview of the BELL Evaluation

Intervention. The Building Educated Leaders for Life (BELL) middle school program is a full-day academically oriented summer program that serves rising sixth- through eighth-graders who are identified by their school as performing one to two years below grade level, on average. The program operates five days a week for approximately five weeks during the summer. Its day is a traditional "full day" (6.5 hours), in which the morning is devoted to math and reading instruction and the afternoon provides enrichment through activities in science, physical education, the arts, and other creative subjects — except on Fridays, when there are guest speakers or field trips.

Study sample. Three of BELL's partner districts were eligible for the study in summer 2012 and agreed to participate. Schools in these districts identified students who were performing below grade level and encouraged them to apply to the program. In total, 1,032 rising sixth-, seventh-, or eight-grade students applied to the middle school program and agreed to be part of the study.

Research design. Random assignment was used to determine which students would be invited to participate in the BELL program (the BELL group) and which students would not be invited to participate in BELL (the non-BELL group). Students and BELL staff were informed of the decision shortly before the program began. Because group membership was determined using random assignment, the impact of the program can be estimated by comparing the outcomes of students in the BELL group and those in the non-BELL group. Because non-BELL students were free to participate in any other summer activities instead, this is a test of BELL's middle school program relative to the "business as usual" summer activities that students would have experienced otherwise.

Data collection and the analysis sample. Information about students' characteristics at baseline was obtained from the application form for BELL. Schools also provided data on students' scores on state tests in the spring before program participation. Classroom observations, interviews, and focus groups with staff were conducted in the third and fourth weeks of the program. Attendance data were obtained from BELL. In the fall after the summer program, students in both groups took a reading achievement test (Group Reading Assessment and Diagnostic Examination, or GRADE) and a math achievement test (Group Mathematics Assessment and Diagnostic Examination, or GMADE), and they completed a survey. The analysis of impacts is based on 919 students (89 percent of the study sample) for whom fall 2012 achievement and survey data are available.

Outcomes. The study focuses on reading achievement test scores, math achievement test scores, and student engagement in fall 2012, after participating in the program.

Limitations. The study has three main limitations. First, the margin of error around the impact findings is quite large; therefore, though the study does provide an unbiased picture of how BELL and non-BELL students differed at the end of the summer, the differences cannot be confidently attributed to BELL unless the impacts are large in magnitude. Second, the three school districts in the study were new partnerships for BELL in summer 2012, and they operated voluntary (rather than mandatory) programs; therefore, the findings may not be representative of the effect of BELL's program in districts that have more experience with it or in districts where student participation is mandatory. Finally, BELL's middle school model has changed since summer 2012 — especially with respect to teacher training and the math and literacy curricula; thus, the findings may not be representative of the impact of the model as it now exists. Given these limitations, this study cannot provide conclusive evidence of impacts, and its findings may not be generalizable. Because of random assignment, however, the findings can still be useful for generating *preliminary* evidence about the potential effects of middle school summer programs and for understanding the environment in which they operate.

and which students would not be invited to participate (the non-BELL group). Students and BELL staff were informed of the decision shortly before the program began.

In early June 2012, a total of 1,032 rising sixth-, seventh-, or eighth-grade students had applied to the middle school program in the three study districts and had agreed to be part of the study. Of these students, 643 were randomly assigned to the BELL group, and the remaining 389 were placed in the non-BELL group. Non-BELL students were, of course, free to participate in any other summer activities instead. Thus, this study is a test of BELL's middle school program relative to the "business as usual" summer activities that students would have experienced otherwise.

Several types of information were collected to measure impacts on student outcomes and to understand the context in which the program was operated. On their return to school in fall 2012, students in the study were encouraged to take standardized tests in math and reading, as well as to complete a short survey asking them to describe what they had done over the summer and the extent to which they were engaged in school in the fall. To understand program implementation, the evaluation team also visited each district in the third or fourth week of the program to observe several classrooms and to interview teachers, mentors (teaching assistants), program managers, and assistant program managers.

Program Implementation, Student Attendance, and the Summer Activities of Students

To better understand the context in which the BELL program was implemented, this study examined several features of the program's implementation in the three study districts. Prior research has shown that some summer programs produce positive effects but that many do not. Thus, learning more about the conditions that can facilitate or challenge a summer program's success is important for advancing the field of summer learning. The study's key findings are summarized below.

- **How well was the BELL program implemented in the study districts?** Overall, in summer 2012, the program was well implemented relative to the BELL middle school model. In all three districts, program leaders (program managers, assistant program managers, and lead teachers) expressed that teachers were of high quality and were performing strongly in the program. The academic instruction offered by BELL was also strong relative to national quality standards of summer learning programs.

- **Were there any challenges to program implementation?** In summer 2012, there were two main challenges to implementation. First, all the BELL program leaders reported delays in receiving program materials and diagnostic

testing data. This start-up challenge may have been exacerbated by the fact that student recruitment for the study continued until shortly before the start of the program, and the curriculum vendor was experiencing a backlog of orders to fill. Second, BELL teachers — all of whom are certified — reported that they would have benefited more from the staff training if it had focused on the BELL curricula, rather than on instructional practices and pedagogy. (BELL has made several changes to its model since summer 2012, and some of them aim to address these challenges.)

- **How often did students attend the program? How many hours of instruction did they receive?** In the average study district in summer 2012, the attendance rate among students who attended at least one day of the program was 82 percent, which is above BELL's internal monitoring target of 80 percent. Students in the BELL group received, on average, about 23 hours of academic instruction in each subject area.

- **How do the summer activities of BELL students differ from the experience of non-BELL students?** In summer 2012, BELL students in the average study district received about 18 more hours of formal instruction (per subject area) than non-BELL students. Although BELL students did not write poems, letters, or stories more often than non-BELL students, they did report playing math games or doing math problems more often. Also, participating in BELL did not prevent students from engaging in other summer activities: BELL students were not less likely than non-BELL students to play sports, watch TV, go to camp, read a book, or go the library during free time.

In general, the study's findings indicate that BELL implemented a well-run and well-attended program in summer 2012 and that students in the BELL program received more academic instruction than they would have received otherwise. The findings also suggest that the program may have been more effective at changing middle school students' math behaviors than their writing behaviors.

Impacts on Academic Achievement and Engagement

As explained above, this evaluation lacks the ability to statistically detect effects of the magnitude seen in prior evaluations of summer programs. This means that effects on academic achievement must be very large (equivalent to about 14 to 17 weeks of regular schooling) in order to conclude that they are not due simply to chance. However, the impact estimates themselves are still rigorous and unbiased, and thus the results can be used to identify suggestive or

preliminary patterns of effects to inform the field of summer learning. The key findings follow and are summarized in Table ES.1.

- **What was BELL's impact on middle school students' reading achievement when they returned to school in the fall?** In the average study district in fall 2012, BELL students did not have higher reading test scores than non-BELL students (effect size = 0.01; p-value = 0.929). These results are consistent across reading subtests. Thus, it cannot be concluded that BELL had a positive impact on students' reading scores. In one of the three study districts, the effect on reading scores is negative and statistically significant, which further supports the hypothesis that the program did not improve students' reading achievement.

- **What was BELL's impact on middle school students' math achievement when they returned to school in the fall?** In the average study district in fall 2012, BELL students outperformed non-BELL students in math by an effect size of 0.07, which is equivalent to a little over one month of additional learning and is the amount by which students are expected to grow during a five-week period during the regular school year. The magnitude of this effect is also similar in size to what has been found in prior evaluations of voluntary summer programs at the elementary school level. On the one hand, this difference is not statistically significant, which means that this result could simply be due to chance rather than to the effect of BELL. On the other hand, some of the study's ancillary findings support the hypothesis that BELL had a small but positive effect on math achievement. For instance, in one of the study districts, BELL had a statistically positive impact on students' math scores in one subdomain. BELL also had a statistically significant effect on students' participation in math-related activities during the summer, which is an important precursor to impacts on math achievement.[7]

- **What was BELL's impact on middle school students' emotional and behavioral engagement when they returned to school in the fall?** In the average study district in fall 2012, BELL students appear to have been no more (or no less) engaged than non-BELL students when they returned to school (effect size = –0.01; p-value = 0.927). Thus, despite having attended an academically focused program for five weeks during the summer, the BELL group did not "burn out" and return to school with less motivation to learn.

[7]Furthermore, BELL's effect on students' math-related summer activities is largest in the study district that also had statistically positive effects on one of the math subdomains.

The Evaluation of Building Educated Leaders for Life (BELL)

Table ES.1

Impacts on Academic Achievement in the Fall:
Fall 2012 Analysis Sample

Outcome	BELL Group	Non-BELL Group	Estimated Impact	Effect Size	P-Value for Estimated Impact
Reading achievement (standard score)[a]	91.6	91.5	0.1	0.01	0.929
Corresponding grade equivalent	5.2	5.2			
Corresponding percentile	32	32			
Corresponding normal curve equivalent (NCE)	38	38			
Math achievement (standard score)[a]	87.6	86.6	0.9	0.07	0.286
Corresponding grade equivalent	5.1	4.9			
Corresponding percentile	27	25			
Corresponding normal curve equivalent (NCE)	33	32			
Sample size (N = 919)	585	334			

SOURCES: MDRC calculations based on the GRADE and GMADE assessments administered in fall 2012.

NOTES: The analyses reported in this table are based on the sample of students who took the GRADE and GMADE assessments and who responded to the student survey in fall 2012 (Fall 2012 Analysis Sample). Estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012, as well as random differences between the BELL and non-BELL groups with respect to the following variables: a student's score on state reading and math tests taken in spring 2012, whether a student has an individualized education plan (IEP), whether the student has English as a Second Language (ESL), whether a student is eligible for free or reduced-price lunch, parent education, race/ethnicity, and gender. The values in the column labeled "BELL Group" are the observed means for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-BELL group, using the observed mean covariate values for the BELL group as the basis for the adjustment. Each of the three study districts is given an equal weight when estimating the results reported in this table. Rounding may cause slight discrepancies in calculating sums and differences.

Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome measure for students in the Fall 2012 Analysis Sample who are in the non-BELL group.

A two-tailed t-test was applied to differences between BELL and non-BELL groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]Students enrolled in fifth grade in spring 2012 were given Level 5 of the GRADE and GMADE; students in sixth grade were given Level 6; and students in seventh grade were given Level M. The national average for GRADE and GMADE standard scores is 100, and the standard deviation is 15. No statistical tests or arithmetic operations were performed on grade equivalents and percentiles because these are not equal-interval scales of measurement.

In general, these findings provide suggestive preliminary evidence that the BELL middle school program did not have an impact on students' reading skills but that it may have had positive effects on students' math skills.

Discussion and Next Steps

This study provides several encouraging findings with respect to the potential of full-day academically oriented summer programs to engage middle school students. First, it is possible to implement such a program well, relative to the intended model and relative to standards in the field of summer learning. Second, voluntary academic summer programs for middle school students can have high attendance rates, even though these students have more control over their time than when they were younger. Thus, a five-week summer program can substantially increase the amount of academic instruction received by students — in BELL's case, about 18 extra hours per subject area. Third, participating in an academic summer program does not prevent students from doing other, "fun" summer activities, like playing sports or watching TV, nor does it make them less engaged in their schoolwork when they return to school in the fall. Finally, there is suggestive preliminary evidence that BELL's summer program for middle school students may have an impact on students' math achievement, equivalent to a little more than one month of regular schooling. Though not statistically significant, the magnitude of this effect is similar in size to what has been found in prior evaluations of voluntary summer programs at the elementary school level.

Findings from this study of BELL's middle school model also point to several challenges that academic summer programs for this age group may face. First, strong start-up is important for summer programs because they are short in duration; yet it can be difficult to hit the ground running on the first day.[8] The exact number of students is often uncertain until shortly before the program starts, so teachers are sometimes hired and materials are ordered within days of the program's start. Thus, summer program staff should make a concerted effort to be ready to start on Day One of the program. Second, staff training should be tailored to the qualifications of the teaching staff. If teachers are certified, for instance, then the teaching staff may benefit more from a training that focuses on the summer program's curricula, rather than on general pedagogy or instructional practices. Finally, it may be more difficult for summer programs to improve middle school students' reading achievement than their math achievement. Prior research has shown that summer programs for elementary school students (including BELL's elementary program) can have a positive effect on the reading achievement of younger students. The findings for middle school students from this study are not as encouraging. One lesson that may be drawn from these findings is that serving middle school students (especially in the area of reading and writing instruction) may require a different approach. To keep them engaged, for instance, interactive activities and hands-on tasks are recommended.[9]

[8]Beckett et al. (2009).
[9]Beckett et al. (2009).

As a continuous learning organization, BELL has made several changes to its middle school model since summer 2012, with the goal of improving instructional quality. Teacher training has been strengthened and decentralized to allow for greater individualization of the training to the local staff's needs. BELL has also replaced its previous curricula with new ELA and math curricula that are aligned with Common Core standards. These new curricula are structured in a way that provides teachers with opportunities to individualize instruction (through one-on-one and small-group activities), and they include hands-on project-based activities that are more engaging to middle school students. BELL is also using a different diagnostic tool to assess students' math and reading achievement, which allows teachers to identify specific skill deficiencies. Lead teachers are also now expected to serve as an "instructional coach": They observe classrooms each week; they provide advice to teachers about how to improve instruction and better engage students; and they give teachers feedback on their weekly lesson plans. Finally, BELL has made changes to the distribution process for delivering materials to sites, which has resulted in the more timely arrival of key material resources at the start of summer.

These programmatic enhancements are in line with the best practices recommended by the field of summer learning and are a positive step toward strengthening BELL's middle school model. In the coming summers, BELL intends to continue to strengthen and refine its program. With the support of long-standing funders, the organization has embarked on a multi-year process to look for ways to better engage and teach struggling middle school students. As part of this process, BELL has created a Middle School Advisory Board whose membership includes researchers and practitioners with expertise in middle school interventions and summer programs, who will advise BELL on best practices for teaching middle school students. BELL plans to implement further modifications to its program, based on the board's recommendations, and to assess whether these modifications have the potential to improve student outcomes. Given that there are so few examples of effective models for middle school summer programs, these changes to the BELL model — and the evaluation of their implementation and effects — will be of interest to the larger field.

References for the Executive Summary

Balfanz, Robert. 2009. *Putting Middle Grades Students on the Path to Graduation: A Policy and Practice Brief.* Westerville, OH: National Middle School Association.

Balfanz, Robert, Liza Herzog, and Douglas J. Mac Iver. 2007. "Preventing Student Disengagement and Keeping Students on the Graduation Path in Urban Middle-Grades Schools: Early Identification and Effective Interventions." *Educational Psychologist* 42, 4: 223-235.

Beckett, Megan, Geoffrey Borman, Jeffrey Capizzano, Danette Parsley, Steven Ross, Allan Schirm, and Jessica Taylor. 2009. "Structuring Out-of-School Time to Improve Academic Achievement: A Practice Guide." NCEE 2009-012. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Chaplin, Duncan, and Jeffrey Capizzano. 2006. *Impacts of a Summer Learning Program: A Random Assignment Study of Building Educated Leaders for Life (BELL).* Washington, DC: Urban Institute.

Eccles, Jacqueline S. 1999. "The Development of Children Ages 6 to 14." *Future of Children* 9, 2: 30-44.

Juvonen, Jaana, Vi-Nhuan Le, Tessa Kaganoff, Catherine Augustine, and Louay Constant. 2004. *Focus on the Wonder Years: Challenges Facing the American Middle School.* Santa Monica, CA: RAND Education.

Kieffer, Michael J., and William H. Marinell. 2012. *Navigating the Middle Grades: Evidence from New York City.* New York: Research Alliance for NYC Schools.

Levin, Henry, Clive Belfield, Peter Muenning, and Cecilia Rouse. 2007. *The Costs and Benefits of an Excellent Education for All of America's Children.* New York: Columbia University, Teachers College.

Reyes, Olga, Karen L. Gillock, Kimberly Kobus, and Bernadette Sanchez. 2000. "A Longitudinal Examination of the Transition into Senior High School for Adolescents from Urban, Low-Income Status, and Predominantly Minority Backgrounds." *American Journal of Community Psychology* 28, 4: 519-544.

Roderick, Melissa R. 1995. "School Transitions and School Dropout." In K. Wong (ed.), *Advances in Educational Policy.* Connecticut: JAI Press.

Chapter 1

Introduction

The middle school years are a critical turning point for youth educationally. Numerous studies have shown that students' success in grades 6 to 8 has profound implications for their future.[1] Attendance, grades, and test scores during the middle school years all predict students' odds of graduating from high school, which in turn predicts future earnings.[2] Students who are academically behind in middle school are much less likely to overcome the challenges of learning more demanding material in high school, which is a less student-centered environment.[3] Thus, effective supplemental middle school educational services targeted at these students are likely to be a good investment for society.

Yet teaching middle school students who are behind is notoriously difficult because of the developmental changes that occur during this period.[4] After years of relatively stable growth, youth ages 10 to 14 begin experiencing dramatic changes cognitively, physically, socially, and emotionally. Cognitively, their ability to think abstractly increases. Physically, puberty starts — but not at the same time for all youth — leading to large variation in physical maturity. Socially, impressing and fitting in with their peers becomes significantly more important to self-esteem than doing well in areas that their parents and teachers value, such as doing well in school.[5] Indeed, the desire for peer conformity peaks at this age despite the fact that, developmentally, there is much greater diversity among youth in this period than either before or after. Emotionally, middle school students are more self-conscious than ever before and, thus, tend to spend more time engaged in activities in which they are already strong and to suffer more anxiety than before when engaging in activities in which they are weaker.[6] Thus, students who struggle academically often try to avoid activities aimed at helping them with their weaknesses, and their anxieties tend to inhibit the learning process when they do engage in such activities.[7] Finally, in conjunction with all these struggles, middle school students are also striving to have more autonomy in their relationships with adults, especially with their parents. It is no wonder that middle school has been called the "Bermuda Triangle of education."[8]

---

[1] See, for example, Reyes, Gillock, Kobus, and Sanchez (2000); Roderick (1995); Balfanz (2009); Balfanz, Herzog, and Mac Iver (2007); and Kieffer and Marinell (2012).

[2] Levin, Belfield, Muenning, and Rouse (2007).

[3] Holcomb-McCoy (2007).

[4] Eccles (1999), p. 36.

[5] Harter (1998).

[6] Eccles (1999).

[7] Eccles and Wigman (2000).

[8] Juvonen et al. (2004), p. xv.

Despite the difficulty of the task, strong pressure to perform well on standardized tests has led more school districts to respond to the struggles of their middle school students by providing them with extra help over the summer to enable them to start the new year with stronger basic skills. Some superintendents have made summer school mandatory for students who score particularly poorly on critical tests. Others, worried about discipline and the engagement of mandated students, strongly encourage struggling students to attend voluntarily. While there are many studies of elementary school summer programs (including some that have found positive impacts), there are few studies of the impact of summer programs on middle school students. This is problematic, given that summer school is a costly endeavor and districts are operating with increasingly tight budgets.

This report presents the findings from a study of the middle school academic summer program offered by Building Educated Leaders for Life (BELL). BELL's middle school program serves rising sixth- through eighth-graders who are identified by their school as performing one to two years below grade level, on average. The program operates five days a week for approximately five weeks during the summer. The program day is a traditional "full day" (6.5 hours), in which the morning is devoted to math and reading instruction and the afternoon provides enrichment through activities in science, physical education, the arts, and other creative subjects — except on Fridays, when there are guest speakers or field trips. This study — which is funded by the Edna McConnell Clark Foundation's Social Innovation Fund (SIF) — provides a unique opportunity to explore the effects of full-day academic summer programs for middle school students. (See Box 1.1.)

In this study, the impact of BELL is evaluated using a random assignment research design, which is the most rigorous type of design for evaluating program effects. (See Box 1.2.) A lottery-like process was used to determine which eligible students would be invited to participate in the BELL middle school program (the BELL group) and which students would not be invited to participate in BELL (the non-BELL group). Importantly, because admission to the program was determined using random assignment, students in the BELL and the non-BELL groups were comparable with respect to their motivation and ability at the start of the program. This means that any subsequent difference between the two groups with respect to academic outcomes in the fall after participating in the program can be interpreted as the impact of BELL.

Despite its rigorous research design, this study has three important limitations that have implications for the generalizability and statistical power of its findings:

- **The study is underpowered.** Due to various challenges related to student recruitment, the margin of error around the impact findings from this study is quite large. Therefore, even though random assignment ensures that the study provides an unbiased picture of how BELL and non-BELL students differed

Box 1.1

The Edna McConnell Clark Foundation (EMCF) Social Innovation Fund

The Social Innovation Fund (SIF) — an initiative enacted under the Edward M. Kennedy Serve America Act — targets millions of dollars in public-private funds to expand effective solutions across three issue areas: economic opportunity, healthy futures, and youth development and school support. This work seeks to create a catalog of proven approaches that can be replicated in communities across the country. The SIF generates a 3:1 private-public match, sets a high standard for evidence, empowers communities to identify and drive solutions to address social problems, and creates an incentive for grant-making organizations to target funding more effectively to promising programs. Administered by the federal Corporation for National and Community Service (CNCS), the SIF is part of the government's broader agenda to redefine how evidence, innovation, service, and public-private cooperation can be used to tackle urgent social challenges.

The Edna McConnell Clark Foundation, in collaboration with MDRC and The Bridgespan Group, is leading a SIF project that aims to expand the pool of organizations with proven programs that can help low-income young people make the transition to productive adulthood. The project focuses particularly on young people who are at greatest risk of failing or dropping out of school or of not finding work; who are involved or likely to become involved in the foster care or juvenile justice system; or who are engaging in risky behavior, such as criminal activity or teenage pregnancy.

EMCF, with its partners MDRC and Bridgespan, selected an initial cohort of nine programs and a second cohort of three programs to receive SIF grants: BELL (Building Educated Leaders for Life), Center for Employment Opportunities, Children's Aid Society-Carrera Adolescent Pregnancy Prevention Program, Children's Home Society of North Carolina, Communities In Schools, Gateway to College Network, PACE Center for Girls, Reading Partners, The SEED Foundation, WINGS for Kids, Youth Guidance, and Children's Institute, Inc. These organizations were selected through a competitive selection process based on prior evidence of impacts on economically disadvantaged young people, a track record of serving young people in communities of need, strong leadership and a potential for growth, and the financial and operational capabilities necessary to expand to a large scale.

The EMCF Social Innovation Fund initiative is called the "True North Fund" and includes support from CNCS and 15 private co-investors: The Edna McConnell Clark Foundation, The Annie E. Casey Foundation, The Duke Endowment, The William and Flora Hewlett Foundation, The JPB Foundation, George Kaiser Family Foundation, The Kresge Foundation, Open Society Foundations, Penzance Foundation, The Samberg Family Foundation, The Charles and Lynn Schusterman Family Foundation, The Starr Foundation, Tipping Point Community, The Wallace Foundation, and Weingart Foundation.

Box 1.2

Why Is Random Assignment Important?

The BELL evaluation and many of MDRC's other studies use a random assignment research design to measure the effectiveness of programs created to help students succeed. This approach involves a lottery-like process that places students who are eligible and willing to participate into either a program group that receives a specific intervention or a control group that receives regular "business as usual" services. Random assignment ensures that the characteristics of students in the program group and in the control group are not systematically different at baseline, the start of the study, and that any differences between the two groups at the end of the study can be attributed to the program that is being evaluated. By using random assignment and measuring the outcomes of both groups after the end of the program, MDRC is able to estimate the causal impact of the program on specific student outcomes. This rigorous method of evaluation produces results that policymakers and practitioners alike can readily understand and trust.

at the end of the summer, it is difficult to conclude that these effects are statistically different from zero unless they are large in magnitude — much larger than would be expected from a five-week summer program. (For its effects to be statistically significant, BELL's five-week program would have to be three times more effective than five weeks of regular schooling or three times more effective than previously evaluated academic summer programs.)

- **The study districts may not be representative of BELL's other middle school sites.** Given the eligibility criteria for the study, the school districts that are included in this evaluation ended up being new partnerships for BELL in summer 2012, and they operated voluntary (rather than mandatory) summer programs. It is difficult to determine how these two district features affect the generalizability of the study's findings to BELL's more experienced middle school sites and/or to sites where student participation was mandatory.

- **The program has evolved since the evaluation.** This study is an evaluation of BELL's middle school model as it existed in summer 2012. As an organization that embraces continuous improvement, BELL has made changes to its middle school model since then, most notably with respect to staff training and the math and reading curricula that are used for instruction. Thus, the findings presented in this report may not generalize to the impact of BELL's middle school model in its present form.

Given these limitations, the present study of BELL's middle school program cannot provide a definitive or generalizable answer about the impact of summer programs for middle

school students. Because of random assignment, however, the study's findings are unbiased; therefore, the results in this report can still be useful for generating *preliminary* evidence about the potential effects of middle school summer programs and for understanding the environment in which such programs operate. One goal of this report is to look for consistent patterns in the direction and magnitude of BELL's effect on students' summer activities and their academic outcomes in the fall. The report also analyzes impacts and program implementation by school district, to explore whether particular features of implementation might be associated with more positive effects. Such analyses can be useful in generating strategies for building stronger summer learning programs for middle school students.

This chapter provides further context for the current study by discussing the rationale for summer academic programs and what is known about their benefits. This is followed by a description of the features of BELL's middle school summer program, as well as the design of the current evaluation study — including its research questions and the methodology used to evaluate BELL's impacts. The chapter concludes with a preview of the findings and an overview of the rest of the report.

Potential Benefits of Summer Academic Programs

It is common for program and policymakers to motivate the need for summer programming by referring to "summer learning loss," a phenomenon seen in test data from the 1980s and 1990s. This earlier research showed that students from poorer families might actually forget as much as a half a grade of math and reading skills over the summer.[9]

While research strongly supports the hypothesis that the skills gap between students from wealthier and poorer family increases over the summer, a few recent studies are starting to question whether all poor youth suffer summer learning loss.[10] For example, an evaluation of Higher Achievement — a middle school academic after-school and summer program for economically disadvantaged students — found no summer learning loss among either the treatment (program) group or the control group.[11] Similarly, recent unpublished work by von Hippel and Downey finds that while children learn more slowly during the summer, learning loss is not inevitable.

Regardless of whether or not all struggling students lose skills over the summer, middle school students who begin the school year behind — like the students served by BELL — are at

[9]Heyns (1978); Entwisle and Alexander (1992); Cooper, Charlton, Valentine, and Muhlenbruck (2000); Downey, von Hippel, and Broh (2004).

[10]Downey, von Hippel, and Broh (2004).

[11]Herrera, Linden, Arbreton, and Grossman (2011).

greater risk educationally than those who are at grade level. For example, sixth-grade students who fail a course, have poor behavior, or attend school less than 80 percent of the time have only a 10 percent to 20 percent chance of graduating on time.[12]

Remedial summer programs aimed at addressing this problem are premised on the hypothesis that if students receive additional instruction on the material that they have not yet mastered, their math and reading skills will improve. The research into this hypothesis is quite mixed.[13] Although some summer school programs have improved students'' reading and/or math test scores, many have not.

Most evaluations of voluntary remedial summer programs have been conducted at the elementary school level. The most rigorous of these studies suggest that when these programs are effective, they increase test scores by an amount that is approximately equal to one month of regular schooling (which is about the effect that one would expect from a program of four to six weeks), though not necessarily in both math and reading.[14] A study that is particularly relevant to the present one is the random assignment study of BELL's elementary school summer program. It found that children in the BELL treatment group gained a month's worth of reading skills during the summer, relative to their counterparts in the control group.[15] (Math achievement was not tested in this study.)

It is unclear from the literature whether BELL's impact on middle school students would be expected to be larger or smaller than its effect on elementary school students. The only rigorous studies of remedial summer programs for middle school students have been evaluations of *mandatory* programs that enroll students who have failed a test that they must pass in order to progress to the next grade. The impacts of these programs range from having no effect to having effects that are equivalent to three to six months of regular schooling.[16] Given that the material is being delivered in a high-stakes environment, the impact of summer programs on the academic achievement of students who are mandated to attend could be greater than for students who are attending voluntarily.

[12]Balfanz, Herzog, and Mac Iver (2007).

[13]Two excellent summaries of this literature are found in Cooper, Charlton, Valentine, and Muhlenbruck (2000) and in Sloan McCombs et al. (2011).

[14]For reviews of these studies, see Cooper, Charlton, Valentine, and Muhlenbruck (2000); Sloan McCombs et al. (2011); Terzian, Moore, and Hamilton (2009); and Kim and Quinn (2013). These impacts were translated into "a month of regular schooling" by using the data on the average effect-size gains experienced by students in different grades reported in Hill, Bloom, Black, and Lipsey (2007).

[15]Chaplin and Capizzano (2006).

[16]See Matsudaira (2008); Jacob and Lefgren (2004); and Mariano and Martorell (2013).

The BELL Middle School Summer Program

The goals of the BELL middle school program are to increase children's literacy and math skills by providing them with engaging and age-appropriate instruction and to enhance their social development by giving them opportunities to be successful and to experience the broader community.

To achieve these goals, BELL provides middle school students with 6.5 hours of daily programming for approximately five weeks, five days per week. During this time, several types of activities are provided to students: academic instruction in math and English Language Arts (ELA); social and academic enrichment instruction; community time; and field trips, guest speakers, and community service.

The BELL program day typically starts with *community time.* This time is intended to build community and strengthen the bonds among the students and the staff. Part of community time is spent working on a jingle that focuses on positive aspects of being part of BELL; each classroom has its own jingle. The remainder of community time can be used in different ways. In some schools, community time is like a homeroom at the start of the day, whereby the "mentors" (also called "teaching assistants") engage students using activities from a positive social development and health curriculum. In other schools, community time is more like an all-school assembly with a "pep-rally feel," in which guest speakers encourage and inspire the students to strive for success.

Core academic instruction occurs Monday through Thursday mornings and is provided by a certified English Language Arts (ELA) or math teacher. BELL academic teachers are certified teachers, and they receive training prior to the beginning of the program. In summer 2012, teachers received one full day of in-person training and were expected to complete nine hours of online training before the start of the program.

In any given week, students receive six hours of ELA and math instruction (twelve hours total). Monday through Thursday mornings, students receive an hour of literacy instruction and an hour of math instruction each day. During the week, students also participate in two hours of project-based literacy activities anchored by a novel or writing assignment, and two hours of project-based math activities. In total, across all five weeks of the program, students are offered 30 hours of ELA instruction and 30 hours of math instruction.[17] In summer 2012, the literacy curriculum was Houghton Mifflin Harcourt *Summer Success,* and the math curriculum was *On Core,* a new Common Core curriculum.

[17]Each week for five weeks, students receive six hours of instruction per subject area per week, for a total of 30 hours per subject area.

Because the program is remedial and is intended to help students catch up if they are below grade level, teachers cover material from the prior school year. To help each class stay on track with the learning objectives, teachers are given a pacing guide that shows them the material that they should be covering each week. Students' reading and math skills are also tested at the beginning of the five-week program, to help teachers assess the strengths and weaknesses of each student, and then are tested again at the end of the program so that results can be measured and reported to the district. In summer 2012, BELL used the Stanford Diagnostic Reading Test and the Stanford Diagnostic Math Test for diagnostic assessments.

BELL teachers are assisted by a mentor. The teacher plans the lessons for each day and informs the mentor of the plans briefly before the start of each class. The teacher leads the instruction of the lessons, while mentors assist with student learning by working with individual groups and taking the lead on behavioral management. The academic teachers are with the students only during the morning, and the math and reading teachers rotate into the classroom when it is their turn to teach. In contrast, the mentors stay in the same classroom all morning, and they also follow students into their afternoon activities.

Monday through Thursday afternoons, students participate in two hours of fun and engaging social or academic enrichment activities to broaden their interests, develop positive teamwork and leadership skills, and allow them to discover and demonstrate their strengths in different ways. The enrichment activities are either designed by teachers (such as playing steel drums, cooking, or journalism), are requested by the district, or are grade-specific thematic enrichment curricula offered by BELL.[18] In some schools, students stay in the same type of enrichment during the entire program; in other schools, students rotate to a different type of enrichment class halfway through the program.

On Fridays, students participate in field trips and community service projects — and, in some sites, attend guest lectures by community leaders — to broaden their interests and extend their learning beyond the classroom. Field trips include going to museums, plays, the zoo, science centers, and other interesting local attractions.

To achieve its goals, BELL aims to hire staff who will be strong positive adult role models. At each school, the operation of the BELL program is overseen by a program manager (who is typically a principal or assistant principal in the district during the regular school year), an assistant program manager, and a lead teacher who acts as a resource for teachers and their

[18]These curricula emphasize (1) social-emotional skills, goal-setting, and positive choices; (2) project-based thematic units to research and explore such community issues as global health and homelessness: (3) gender-based focuses on impulse control, anger management, academic achievement, and decision-making for boys and on self-image, womanhood, anger management, academic achievement, and community advocacy for girls; and (4) hands-on science activities.

teaching assistants. At a higher level, a regional leader oversees the management of all centers in the different regions where the BELL program is offered.

As noted above, BELL also operates a summer program for elementary school students. The elementary school model and the middle school model are similar in several ways. At both levels, students are given three hours of reading or math instruction four mornings a week, and this instruction covers material from the prior school year. In summer 2012, teachers in both the elementary school and the middle school program used the same Houghton Mifflin Harcourt research-based curriculum, *Summer Success.* But while the middle school program offers six hours per week of reading instruction and the same amount per week of math, the elementary school models offers eight hours per week of reading instruction and four hours per week of math. In addition, in BELL's elementary school model, students receive academic instruction in both reading and math from the same teacher; in the middle school model, the math and reading teachers rotate into the classroom when it is their turn to teach. (This latter approach reflects the middle school practice of having teachers be content area experts.)[19] Although students at both levels participate in enrichment activities in the afternoon, middle school participants can have a choice of their afternoon activities. Finally, the field trips, guest lectures, and community service on Fridays are tailored toward middle school students.

As noted, the BELL middle school model has evolved since the time of this evaluation. The structure of the program and the amount of instruction provided remain the same, but some of the features related to instructional quality — most notably, the curriculum and the way in which teacher training is provided — have changed since summer 2012. Chapter 4 describes how the model has changed since then.

Overview of the Evaluation

The primary purpose of this study is to determine how an academically oriented summer program — with math and reading instruction in the morning and enrichment activities in the afternoon — affects the academic outcomes of struggling middle school students.[20] The study addresses this question by examining the academic benefits experienced by middle school students who voluntarily participate in BELL's middle school program:

- **Reading achievement.** What is BELL's impact on middle school students' reading achievement when they return to school in the fall?

[19]In the models for both levels, the mentors stay in the same classroom all morning and then work with the same group of students during the afternoon enrichment activities.

[20]A copy of the evaluation plan that was written at the start of the study is available on request.

- **Math achievement.** What is BELL's impact on middle school students' math achievement when they return to school in the fall?

In addition to looking at impacts on academic achievement — which are the primary outcomes targeted by the program — the study also examines whether the BELL program had an effect on students' engagement in the fall. On the one hand, by helping students improve their skills during the summer, formal academic summer programs may also help students to be more engaged in their schoolwork when they return to school in the fall. On the other hand, a concern that parents may have about academic summer programs is that their child will be "burned out" in the fall and possibly less engaged. Thus, the study also examines the following secondary question:

- **Attitudes and behaviors.** What is BELL's impact on middle students' emotional and behavioral engagement when they return to school in the fall?

Beyond examining the impact of BELL on student outcomes, it is also important to understand the context in which these impacts are fostered. Prior research has shown that some summer programs produce positive effects but that many do not. Learning more about the conditions that can facilitate or challenge a summer program's success is important for advancing the field of summer learning. Thus, the study also examines several questions related to the program's implementation:

- **Program implementation.** How well was the BELL program implemented in the study districts relative to the intended model and to standards in the field of summer learning? Were there any challenges to implementation?

- **Dosage.** How often do students attend the BELL program? How many hours of instruction do they receive?

- **Service contrast.** How do the summer activities of students in the BELL program differ from the summer experience of similar students who do not participate in the program?

These research questions are examined for all three school districts in the study pooled together and for each of the districts separately. The setting in which an academic summer program is implemented can greatly affect the program's success, for several reasons. First, summer programs like BELL must rely on the resources and infrastructure of the school district (staff, space, and equipment) to operate the program. The extent to which districts make these resources available to summer programs can have an important bearing on the strength of program implementation. Second, the impact of an academic summer program depends not only on the quality of the program itself but also on the extent to which the program improves on the summer services that are otherwise available to students. In a district that is already rich in

summer programs, the incremental effect of a program like BELL would be smaller than in a district where summer services are scarcer. For these reasons, program impacts may vary across school districts and local contexts.

The Study's Design

This evaluation of BELL's middle school program uses a random assignment research design to examine BELL's effects on student outcomes. Some of the districts that partner with BELL operate voluntary summer programs where there are more eligible students than BELL has the capacity to serve. In these oversubscribed voluntary programs, random assignment was used to determine which students would be invited to attend the BELL middle school program (the BELL group) and which students would participate in "business as usual" summer activities (the non-BELL group). The following two sections describe the process by which BELL sites were recruited into the study and the process used for randomly assigning students to the two study groups.

Site Eligibility and Recruitment

In summer 2012, three of the ten districts that partnered with BELL to serve middle school students had oversubscribed voluntary programs and were willing to participate in the evaluation. Of the seven study districts that did not participate in the evaluation, two were operating the BELL program on a mandatory basis (making random assignment infeasible); four districts operated voluntary programs but were unlikely to be oversubscribed (also making random assignment infeasible); and the seventh district did not participate because it would not have been possible for the study team to obtain research approval from the district in a timely fashion.

The three districts in the evaluation are diverse in terms of their geographic location and the range of grade levels served. One district is located in the West (District A), and two are located in the Southeast (Districts B and C). Districts A and B offered the BELL program in one middle school each; District C offered the program in three schools. The schools in Districts A and C served only rising seventh- and eighth-grade students, whereas the middle school in District B served students in all three middle school grades.[21]

The three study districts are unique among some of BELL's other middle school sites in two ways: The study districts were *new partnerships* for BELL in summer 2012, and they operated programs that students were attending *voluntarily.* It is difficult to determine with certainty how these two programmatic features played a role in the magnitude of BELL's impact on stu-

[21]The schools in District C also served rising sixth-grade students, but they received the BELL elementary school model rather than the middle school model. Therefore, sixth-grade students in District C were excluded from the study.

dent outcomes in these districts in summer 2012 and, by extension, whether the findings from this study are generalizable to BELL's other middle school sites.

On the one hand, the three study districts appear to be similar to the nonstudy districts in various ways. First, as discussed in Chapter 2, the three study districts implemented the components of the BELL middle school program with fidelity relative to the intended model. Second, the test scores of students in the three study districts in summer 2012 changed by a similar amount during the program as the scores of students in BELL middle school sites that were more experienced and/or that operated mandatory programs (based on the Stanford diagnostic assessments that BELL administered to students at the start and end of the program). Third, like all school districts that partner with BELL, the three study districts are primarily urban, and their middle schools serve a large proportion of economically disadvantaged and minority students. Almost 60 percent of middle school students in the average study district are eligible for free or reduced-price lunch, and almost all schools in these districts (93 percent) receive Title I funding. Approximately 58 percent of students are black or Hispanic.[22]

On the other hand, this does not guarantee that the findings from this study are generalizable, because the two groups of sites could differ in unobserved ways that affect program impacts. New district partnerships present unique challenges that may have affected the strength of program implementation in unobserved ways. (That is, new district-level relationships must be developed; new program leaders and instructional staff must be hired and trained; and so on.) In this respect, the study's findings may underestimate the impact of the BELL middle school model in districts that have greater experience with the program.

Student Eligibility, Random Assignment, and Sample Size

As noted, BELL aims to serve students who are struggling academically, and so eligibility for the study was limited to students in the three study districts who were performing below grade level academically. In order to make random assignment possible, a further requirement was that students had to be attending the program voluntarily to be eligible for the study. In Districts A and C, the BELL middle school programs were entirely voluntary, and so all students in these two districts had made the decision to attend the summer program. In District B, however, BELL also served students who were mandated to attend the program due to low scores on the state assessment. In this district, only students who participated voluntarily were eligible for the study (though the program did still serve students who were required to attend).

In a typical summer, BELL would have filled the voluntary program slots in these districts on a "first-come, first-served" basis. To make random assignment possible, however,

[22] Appendix F compares the characteristics of the study districts and the nonstudy districts.

schools in the study continued to identify students who were performing below grade level and to encourage applications from these students until shortly before the start of the program. To be included in the study, students and their parents also had to complete the BELL application form and sign the informed consent form. In total, 1,032 rising sixth-, seventh-, or eighth-grade students applied to the middle school program in the three study districts and agreed to be part of the study. Of these 1,032 students, 385 students are from District A; 127 students are from District B; and the remaining half (520 students) are from District C.

Random assignment was then used to determine which of these students would be invited to participate in the BELL middle school program (the BELL group) and which students would not be invited to participate in BELL (the non-BELL group).[23] In order to ensure that each grade-level classroom in the BELL study sites would have 20 students, the research team conducted a separate random assignment lottery-like process for each grade level, as well as for each school that students attended in the spring before the summer program.[24] In total, 643 students (62 percent of study participants) were randomly assigned to the BELL group, while the remaining 389 students (38 percent of study participants) were placed in the non-BELL group.[25] Non-BELL students were, of course, free to participate in any other summer activities instead. Thus, this study is a test of BELL's middle school program relative to the "business as usual" summer activities that they would have experienced otherwise.

As noted above, student recruitment proved to be more challenging than expected, and so schools continued to recruit students into the study until shortly before the start of the program. By extension, in some study districts, randomization occurred very close to the program start date. In District A, randomization occurred four workdays before the start of the program; in District B, students were randomized one workday before the start of the program; and, in District C, randomization was conducted 13 days (two weeks) before the program start date.

[23]To mimic how the program typically operates, a small number of students were also assigned to a nonresearch waiting list. Students on this waitlist were used to backfill the slots of BELL students who did not show up or who left the program. Waitlist students are not included in the study sample or the analysis.

[24]There are 44 grade-by-school random assignment blocks in the full study sample. These blocks represent different combinations of students' grade level and their school in spring 2012. It is important to note that the blocks are defined based on students' school during the previous school year, not on the school where the summer program was held (each of which serves students from many feeder schools). This was done to ensure that the BELL and non-BELL groups would be similar in terms of the distribution of schools that they attended during the school year before the program.

[25]Depending on the extent of oversubscription in a given school and grade level, the percentage of students who were invited to participate in BELL varies across random assignment blocks — from a minimum of 15 percent to a maximum of 88 percent for the sample of students used in the impact evaluation. These differences in the random assignment ratio (and the probability of being invited to attend BELL) must be accounted for to obtain an unbiased estimate of impacts. This was accomplished by including an indicator for each random assignment block in the statistical model. For further information about the statistical analysis, see Appendix A.

Thus, in two of the study districts, students and BELL staff were informed of who would be invited to participate in BELL only as the program was close to kicking off.

Yet, despite the extended recruitment period, the number of students in the study sample is still smaller than anticipated. For this reason, the study is underpowered, and the margin of error around the impact estimates from this study is quite large.[26] This means that only very large impacts can be statistically distinguished from zero. For its effects to be statistically significant, BELL's five-week program would have to be three times more effective than five weeks of regular schooling and also three times more effective than previously evaluated academic summer programs at the elementary school level. Thus, although random assignment ensures that study does provide an unbiased picture of how the outcomes of BELL and non-BELL students differed at the end of the summer, the study will not be able to reliably attribute these differences to BELL unless they are very large in magnitude.[27]

Because the study is underpowered — and because its findings may not be generalizable to all BELL middle school sites — the results presented in this report are *preliminary* and do not provide a definitive answer about the impact of middle school academic summer programs. Rather, the findings should be used to help formulate hypotheses about the potential effectiveness of such programs, to better understand the context in which they are implemented, and to formulate strategies for how such programs might be further strengthened.

Data Sources

Table 1.1 summarizes the types of data that were collected and the timing of data collection activities. These data sources can be grouped into two categories: (1) data about student outcomes and characteristics and (2) data about program implementation. The nature and purpose of these data sources are described below.

[26] Across the three study districts 1,032 students applied to the middle school program and were enrolled in the study. However, because the distribution of the sample was heavily skewed toward one district (District C), the impact of BELL had to be calculated separately by district then averaged across the three districts, so that each district would have an equal weight. This reweighting widened the confidence intervals and lowered the power of the study to detect true impacts of the size seen in other studies. The minimum detectable effect size (MDES) is 0.15 for reading and 0.17 for math. (Appendix A provides a detailed discussion of the MDES.) This means that, in order for effects to be statistically significant, BELL's five-week program would have to have an effect on reading that is equivalent to 17 weeks of regular schooling and an effect on math that is equivalent to 14 weeks of regular schooling. These effect sizes are translated into weeks of regular school-year instruction based on the benchmarks in Hill, Bloom, Black, and Lipsey (2007).

[27] When a study is underpowered, there are two possible reasons for a nonsignificant impact: Either (1) the impact of the program is truly zero or (2) the impact is not truly zero, but the study does not have enough statistical power to confirm that the impact is *not* zero (Murnane and Willett, 2011). It is not possible to disentangle these two explanations, which is why findings from underpowered studies do not provide definitive evidence of effects or no effects.

The Evaluation of Building Educated Leaders for Life (BELL)

Table 1.1

Data Sources for the BELL Evaluation

Data Source	Measure	Purpose	Collection Period
End-of-summer student outcomes and background characteristics			
GRADE assessment	Reading achievement scores (total score, reading vocabulary and reading comprehension)	Fall impacts	Fall 2012
GMADE assessment	Math achievement scores (total score, math operations, math concepts, and math processes)	Fall impacts	Fall 2012
Student survey	Student engagement scales (overall engagement, behavioral engagement, emotional engagement); activities during the summer (library, reading, watching TV, sports, summer programs, etc.); reasons for not attending summer program	Fall impacts; service contrast in summer activities	Fall 2012
Baseline intake form	Race/ethnicity, parent education	Descriptive analyses and covariates in the impact analysis	Spring 2012
School records	State test scores (reading and math), individualized education plan (IEP), free or reduced-price lunch status, English as a Second Language (ESL)	Descriptive analyses and covariates in the impact analysis	Spring 2012
BELL internal data	Characteristics of middle school students served by BELL nationally	Descriptive analysis of students typically served by BELL	Summer 2012

(continued)

Table 1.1 (continued)

Data Source	Measure	Purpose	Collection Period
Program implementation			
Attendance data	Number of BELL days attended	Descriptive analysis of dosage	Summer 2012
Teacher survey	Teacher characteristics (education, experience, grade level taught, role, etc.) and teacher perceptions of the program (materials, training, leadership, etc.)	Descriptive analysis of BELL teacher characteristics and teachers' perceptions about the program	Summer 2012
Program leader interview (program managers, assistant program managers, and lead teachers)	NA	Learn about experience and preparation of leadership staff; local program context; implementation of program elements	Summer 2012
Regional leader interviews	NA	As above	Summer 2012
School district liaison interviews	NA	Learn about local context and nature or partnership between BELL and district	Summer 2012
Teacher and mentor focus groups	NA	Learn about background and training of the teaching staff and perspectives on implementation of the program elements	Summer 2012
Classroom observations	NA	Describe the elements of the BELL model	Summer 2012

Student Data and Analysis Sample

DATA SOURCES AND OUTCOMES

As described below, several types of data were collected about students' characteristics, their summer activities, and their outcomes in the fall after the BELL program ended.

- **Spring (baseline) characteristics and test scores.** Various pieces of information were collected to describe the sample of students in the study. First, during the application process, parents provided information about their child's socioeconomic characteristics (racial or ethnic group, parents' education, and so on). In addition, schools provided information about whether students in the study were eligible for free or reduced-price lunch, whether they had an individualized education plan (IEP), and whether English was their second language. Schools also provided students' scores on the spring 2012 math and reading assessments administered by their state. These test scores were used to determine whether students were proficient, based on local cutoff scores on their state test.[28]

- **Fall testing.** To assess program impacts on academic achievement, students in the study were encouraged to take standardized tests in math and reading in fall 2012. Students' reading achievement was assessed using the Group Reading Assessment and Diagnostic Examination (GRADE), and their math achievement was assessed with its math counterpart, the Group Mathematics Assessment and Diagnostic Examination (GMADE).[29] As diagnostic tests, the GRADE and GMADE are especially useful for measuring the skill levels of students with weak academic skills, such as the students served by BELL.

[28]State test scores were also used as a covariate in the impact model as a way to increase the precision of the estimated impacts. Interaction terms between state test scores and the grade or district of the assessment were used to deal with the different scales of the tests across states. See Appendix A.

[29]The GRADE and GMADE are norm-referenced, research-based assessments that can be administered to groups. They are meant to be diagnostic tools to assess what reading and math skills students have and what skills need to be taught. Level 5 of the GRADE and GMADE was administered to students rising to sixth grade; Level 6 was given to students rising to seventh grade; and Level M was administered to students rising to eighth grade. The GRADE includes 84 test items, and the GMADE includes 82 test items. None of the students in the sample had a zero score or the maximum score. For further technical information about the GRADE and GMADE, see Pearson Education (2001, 2004).

The GRADE contains two subtests (reading comprehension and vocabulary), and the GMADE contains three (concepts, operations, and processes).[30]

- **Fall student survey.** In the same session as when the GRADE and GMADE were administered, students also completed a short survey asking about the extent to which they were engaged in various aspects of instruction when they returned to school in the fall (for example, whether they paid attention in class and whether they completed their homework on time). Students' responses to these items were used to examine BELL's effect on student engagement in the fall after the program.

The GRADE and GMADE achievement tests and the student survey were administered in the fall in order to make it possible to assess the outcomes of BELL and non-BELL students at the same time. In the average study district, students took the test and survey six weeks after the end of the program or one week after the start of the next school year.[31]

For purposes of gauging the effectiveness of BELL's middle school program, this evaluation focuses on two primary outcomes: GRADE total reading scores and GMADE total math scores. Impacts on these two primary outcomes are used as the benchmark for determining BELL's effectiveness. In contrast, BELL's effect on other student outcomes — students' summer activities, their engagement in the fall, and their scores on GRADE and GMADE subtests (reading comprehension, reading vocabulary, math concepts, math operations, and math processes) — are secondary outcomes in this evaluation. Impacts on these outcomes are presented only for the purposes of contextualizing or explaining the pattern of effects on the two primary outcomes. Similarly, impacts on student achievement by study district are also considered secondary; these findings are presented as a means of exploring the consistency (or variability) of effects across different contexts.[32]

[30]In addition to the raw score (total number of items answered correctly), the GRADE and GMADE also provide standardized scale scores, normal curve equivalent scores, grade equivalent scores, percentile scores, and stanine scores.

[31]The follow-up testing was conducted at the beginning of the school year rather than at the end of the BELL program to maximize the likelihood that the response rates for the treatment and control groups would be similar and that the testing environments would be the same. Testing was done over the weekend at several schools, and students in both groups were comingled. In District A, testing occurred an average of 33 days after the program ended; in District B, the average was 46 days; in District C, it was 40 days. Students in Districts A and B had attended five days of school, on average, when testing happened, while students in District C had attended an average of 10 days of school. For more information about fall testing and surveys, see Appendix B.

[32]Because there are only two primary outcomes — and each one is a measure of a different achievement domain (reading or math) — it is not necessary to make adjustments to p-values for multiple hypothesis testing, based on standards used in education research (What Works Clearinghouse, 2014).

One limitation of the data collection effort for this study is that it is not possible to measure the gain (or loss) in students' skills during the summer. The content and scale of the GRADE and GMADE are different from the content and scale of the state assessments that students took in the spring; thus, it is not possible to look at spring-to-fall test score gains by comparing students' scores on spring state tests with their scores on the GRADE or GMADE in the fall.[33] Nor is it possible to use BELL's diagnostic assessment (which it administers to students at the beginning and end of the program) to measure student gains: The same form of the Stanford diagnostic tests was administered to BELL students in both test sittings. Therefore, the change in Stanford test scores may overestimate true growth in student achievement; that is, students may have performed better on the posttest because they remembered questions from the pretest.

The Analysis Sample

The impact analyses presented in this chapter are based on students who completed the GRADE and GMADE assessments as well as the fall student survey: the "Fall 2012 Analysis Sample." Of the 1,032 students recruited into the study, 919 students (89 percent) meet these criteria and are included in the analysis sample.[34] Of these 919 students, 585 are in the BELL group, and 334 are in the non-BELL group.

As noted above, the number of students from District C is larger than the number of students from Districts A and B. But because the three study districts are weighted equally in the pooled findings, District C does not have a larger weight than the other two districts in the overall findings. Thus, the pooled results in this report should be interpreted as the findings *for the average study district.*

Table 1.2 presents the characteristics of students in the Fall 2012 Analysis Sample, for the average study district. (Box 1.3 explains how to interpret the findings presented in this report's tables.) In the average study district, the characteristics of students in the BELL and the non-BELL groups are similar, which demonstrates that random assignment was successful in creating two equivalent research groups at baseline.[35] Both groups of students are high-needs academically: Only about 40 percent were "proficient" on their state's assessment, and almost

[33]State tests are normed based on local (not national) populations, so it is not possible to convert students' scores on these tests to a metric (such as normal curve equivalents or percentiles) that would make them comparable to students' nationally normed GRADE or GMADE scores in the fall. Nor is it possible to obtain scores on state tests in the fall, because state tests are administered only in the spring of each school year.

[34]Response rates did not differ by a statistically significant amount across the BELL and the non-BELL groups. For more information about response rates, see Appendix C.

[35]An omnibus test confirms that, overall and by study district, students in the BELL group and the non-BELL group were not systematically (or statistically) different from each other at baseline. For details, see Appendix C.

20 percent had an individualized education plan (IEP).[36] The majority of students (about 80 percent) were rising into seventh or eighth grade, while 20 percent were rising into sixth grade. More than 75 percent of students in the average study district are black or Hispanic, and almost 90 percent are eligible for free or reduced-price lunch. Demographically, these students are representative of the population typically served by BELL.[37]

Table 1.2

Baseline Characteristics of Students in the Fall 2012 Analysis Sample, by Treatment Group

Characteristic in Spring 2012 (%)	BELL Group	Non-BELL Group	Estimated Difference	P-Value for Estimated Difference
Grade level				NA
Rising into grade 6	19.6	19.6	0.0	
Rising into grade 7	41.6	41.6	0.0	
Rising into grade 8	38.8	38.8	0.0	
Race/ethnicity				1.000
Hispanic	33.9	34.3	-0.4	
Black, non-Hispanic	44.1	45.4	-1.4	
White, non-Hispanic	6.2	4.7	1.6	
Asian	8.6	9.3	-0.7	
Other	7.2	6.3	0.9	
Female	43.0	46.2	-3.2	0.492
Eligible for free/reduced-price lunch	89.1	90.1	-1.0	0.720
English as a Second Language	8.4	11.0	-2.6	0.319
Parent education level[a]				0.636
Did not finish high school	17.7	15.5	2.2	
Has high school diploma or GED certificate	34.8	27.6	7.3	
Has some postsecondary education	27.0	33.1	-6.1	
Has bachelor's degree or higher	12.5	14.6	-2.1	
Other	7.9	9.2	-1.3	
Has an individualized education plan (IEP)	18.1	19.5	-1.4	0.667
Proficient on state test in spring 2012[b]				
Reading	39.5	37.1	2.4	0.568
Math	42.3	40.6	1.6	0.715
Joint test of difference between groups[c] ($\chi 2 = 12.3$)				0.950
Sample size[d] (N = 919)	585	334		

(continued)

[36]For more information about the characteristics of students in the study, see Appendix C.

[37]See Appendix F. Information provided by BELL indicates that, nationally, about 73 percent of middle school students served by BELL are black or Hispanic, which is similar to their proportion in the Fall 2012 Analysis Sample.

Table 1.2 (continued)

SOURCES: MDRC calculations based on the BELL baseline intake form administered in spring 2012 and student records obtained from school districts.

NOTES: The analyses reported in this table are based on the sample of students who took the GRADE and GMADE assessments and who responded to the student survey in fall 2012 (Fall 2012 Analysis Sample). The estimated differences between the BELL group and the non-BELL group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012. The values in the column labeled "BELL Group" are the observed means for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-BELL group, using the observed distribution of the BELL group across random assignment blocks as the basis for the adjustment. Each of the three study districts is given an equal weight when estimating the results reported in this table. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to differences between BELL and non-BELL groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]For students with two guardians, this is the maximum education level of the two guardians.

[b]A student's proficiency is based on the standards in the state where he or she is attending school.

[c]A chi-square test was used to determine whether there is a systematic difference between the BELL group and the non-BELL group at baseline, based on the characteristics included in this table as well as indicators of missing data for all relevant student characteristics.

[d]Due to missing values, the number of students included varies by characteristic. The sample size reported here is for the full Fall 2012 Analysis Sample. The percentage of missing data on any given characteristic does not exceed 10 percent.

Implementation Data

Besides collecting information about students, data were collected to learn about various features of the BELL program and the context in which it was implemented. These data were collected with three goals in mind: (1) to understand how the BELL model as implemented in the study districts compared with the intended BELL model and with objective standards from the field of summer learning, (2) to measure the amount of instruction received by BELL students (the dosage), and (3) to gauge the extent to which the summer activities of BELL students differed from the activities of non-BELL students (the "service contrast").

These aspects of implementation were assessed through several data sources. First, in summer 2012, the evaluation team visited the program schools that served students in the study sample during the third and fourth weeks of the BELL program. (Five schools were visited: one in District A, one in District B, and three in District C.) During these site visits, the following data collection activities were conducted.

- **Interviews with school program leaders.** During the site visits, the study team conducted interviews with all BELL program leaders (defined as the program manager, the assistant program manager, and the lead teacher at each school: 13 program leaders in total). The purposes of these interviews were to learn about the experience and preparation of these key staff, to understand the context of the local program, and to learn about how the model's

Box 1.3

How to Interpret the Findings in This Report's Tables

Many tables in this report show the characteristics, summer activities, or fall outcomes of students in the BELL group and the non-BELL group — and the difference between them — in the average study district. The values in the tables are derived as described below.

"Estimated Impact" or "Estimated Difference" column. This column shows the difference between BELL and non-BELL students with respect to their baseline characteristics or summer activities ("Estimated Difference") or their outcomes in the fall after the summer program ("Estimated Impact"). To calculate the values in this column, the difference between BELL and non-BELL students is estimated for each study district, and these district-specific findings are then averaged across districts. Thus, these values represent the impact (or difference) *for the average study district.* The statistical significance of the estimated difference or impact is indicated by asterisks (*) when the p-value is less than or equal to 10 percent, based on a two-tailed test. Estimated impacts are regression-adjusted to account for random differences in the baseline characteristics of BELL and non-BELL students. All impact findings represent "intent-to-treat" estimates because 8 percent of students in the BELL group did not attend the program at all. Appendix A presents further information about the statistical analysis.

"BELL Group" column. This column shows the observed mean fall outcomes (or baseline characteristics or summer activities) of students randomly assigned to the BELL group. When calculating these outcome levels, each school district is weighted equally. Thus, this column reflects the mean outcomes of BELL students in the average study district.

"Non-BELL Group" column. This column shows the *counterfactual;* that is, it provides an estimate of what the mean outcomes of BELL students would have been had they not been randomly assigned to participate in the program. These values are regression-adjusted based on the observed characteristics of students in the BELL group in the average study district. In practice, they are obtained by subtracting the values in the "Estimated Difference" or Estimated Impact" column from the values in the "BELL Group" column.

"Effect Size" column. This column shows the estimated impact (or difference) scaled as an *effect size* — a metric that is widely used for gauging whether the magnitude of a program's impact is large or small. An effect size is defined as the estimated effect of a program (or the difference in outcomes between BELL and non-BELL students) divided by the standard deviation of the outcome of interest. For example, an effect size of 0.20 represents an improvement in student outcomes that is equal to 20 percent of the standard deviation of the student-level distribution for that particular outcome. The effect size, therefore, indicates how much the BELL program improves a student's outcomes relative to where the student would have been in the outcome distribution for students in the program's target population. As context for interpreting effect sizes, it is useful to keep in mind that, during the regular 36-week school year, the achievement of middle school students is expected to grow by an effect size of 0.32 in reading and by an effect size of 0.42 in math.[*] Thus, five weeks of regular schooling (the duration of the BELL program) is expected to improve student achievement by an effect size of 0.04 in reading and 0.06 in math. In this report, effect sizes are calculated based on the standard deviation of the outcome of interest *for students in the non-BELL group.* The standard deviation for the non-BELL group reflects the expected variability in the outcome that one would find in the absence of the BELL program. Appendix A lists the standard deviations used to calculate effect sizes in this report.

[*]Hill, Bloom, Black, and Lipsey (2007).

elements were being implemented at the time of the interviews and during the first two weeks of the program.

- **Interviews with regional leaders.** During the site visits, researchers also interviewed the BELL regional leader in each of the three study districts. These interviews focused on similar topics as the interviews with program leaders and had similar objectives.

- **Interviews with school district liaison.** During site visits to two of the three school districts, researchers interviewed a school district liaison for summer learning. These interviews aimed to gain a deeper understanding of the local program's context and the nature of the partnership between BELL and the school district.

- **Focus groups of teachers and mentors.** During the site visits, the research team led separate focus groups with about half the BELL teachers and mentors who taught rising sixth- to eighth-grade students. At each school, focus groups were held with teachers (academic and enrichment teachers), and a focus group was held with mentors; the average focus group had five participants.[38] The goal of these focus groups was to collect data on the background of teachers and mentors, the preparation they received for their roles, and their perspective on the implementation of the program elements with which they worked directly. All the focus groups were voluntary, and participants were offered $50 for their time.

- **Observations of classrooms and activities.** During the site visits, researchers observed four to six classrooms in each of the study schools. These observations were conducted for the purpose of being able to accurately describe the components of BELL's program model.

In addition to data from the site visits, the implementation of the BELL program was also evaluated using internal data collected by BELL as part of its regular program monitoring activities, along with data from the fall student survey:

- **BELL teacher survey.** BELL provided the evaluation team with data from the teacher survey that it administers each summer. The responses of aca-

[38]In the study school in District A, the study team held one focus group with teachers; in Districts B and C, the team held two focus groups per school (one with academic teachers and one with enrichment teachers). For more information about the number of teachers and mentors interviewed in each study school and about the protocols for the interviews and focus groups, see Appendix D.

demic teachers who taught students in the study sample were used to measure teachers' experience and satisfaction with various aspects of the BELL program (such as training, materials, and staffing), their own performance in the classroom, and their students' performance and engagement.[39] In summer 2012, the response rate among academic teachers in the average study district who taught students in the study sample was 85 percent.[40] The characteristics of these teachers are discussed in Chapter 2.

- **Attendance records.** BELL also provided the evaluation team with the attendance records of students in the study. These data were used to measure student participation in the BELL program and to understand the amount of academic instruction received by students (the dosage).

- **Student survey.** As noted, the research team also administered a survey to BELL and non-BELL students in fall 2012. The survey includes a set of items asking students to describe their activities during the summer. Students' responses to these questions were used to gauge the extent to which the summer activities of BELL students differed from the activities of non-BELL students (the service contrast).

Appendix B provides additional information about the student and teacher surveys, while Appendix D provides details about the data collected during the site visits, including the number of program leaders, teachers, and mentors who participated in interviews and focus groups.

The Structure of the Report and a Preview of the Findings

This report is structured as follows. Chapter 2 examines the implementation of the BELL middle school program in the three study districts and the context in which the programs operated. Chapter 3 examines whether the BELL program had an impact on students' academic achievement and their engagement in the fall. Chapter 4 concludes by discussing the findings and their implications for the field of summer learning.

Overall, the findings from this study suggest that the BELL middle school model — as implemented in summer 2012 — was strong by several measures. First, in all three study districts, the instructional components of the BELL middle school program were well implemented

[39]For more information about the teacher survey, see Appendix B.

[40]Response rates were 80 percent in District A, 100 percent in District B, and 75 percent in District C.

relative to the intended model and relative to standards in the field of summer learning. Second, BELL was successful at getting middle school students to come to the program; average daily attendance rates among students who attended at least one day of the program exceeded 80 percent, which is notable, given the voluntary nature of the program and the fact that middle school students have more control over their time than when they were younger. Given these attendance patterns, BELL students received about 18 more hours of academic instruction per subject area than non-BELL students during the summer. Third, participating in BELL did not prevent students from doing other "fun" summer activities, like playing sports or watching TV, nor did it make them less engaged in their schoolwork when they returned to school in the fall. Finally, there is suggestive preliminary evidence that BELL may have had small but positive effects on students' math achievement. Specifically, BELL students outperformed non-BELL students by the equivalent of a little over one month of learning, which is the effect that one would expect from a five-week program during the regular school year. Though not statistically significant, the magnitude of this effect is also similar in size to what has been found in prior evaluations of voluntary summer programs at the elementary school level.

Findings from this study also point to several challenges that academic summer programs for middle school students may face. First, strong start-up is important for summer programs because they are short in duration; yet it can be difficult to hit the ground running on the first day.[41] In this study, for instance, the BELL program leaders reported that the programs experienced delays in receiving program materials and diagnostic testing data. These start-up challenges may have been exacerbated by the fact that student recruitment for the study continued until shortly before the start of the program, and the curriculum vendor was experiencing a backlog. However, start-up challenges are likely to always be present, because the exact number of students is often uncertain until shortly before the program starts, so teachers are sometimes hired and materials are ordered within days of the start of the program. Thus, summer programs should make a concerted effort to be ready to start on Day One of the program. Second, staff training should be tailored to the qualifications of the teaching staff. In this study, BELL teachers (all of whom are certified) reported that they would have benefited more from the staff training if it had focused on the BELL curricula, rather than on instructional practices and pedagogy. And, finally, it may be more difficult for summer programs to improve middle school students' reading achievement than their math achievement. In the average study district, BELL's effect on reading scores is numerically close to zero and is not statistically significant. Prior research has shown that summer programs for elementary school students (including BELL's elementary

[41] Beckett et al. (2009).

program) can have a positive effect on the reading achievement of younger students. The findings for middle school students are not as encouraging. One lesson that may be drawn from these findings is that serving middle school students (especially in the area of reading and writing instruction) may require a different approach. To keep them engaged, for instance, interactive activities and hands-on tasks are recommended.[42]

[42]Beckett et al. (2009).

Chapter 2

Program Implementation, Student Attendance, and the Summer Activities of Students

Chapter 2 examines the implementation of the Building Educated Leaders for Life (BELL) middle school summer program in the three study districts in summer 2012 and the broader context in which the program operated. (See the opening pages of Chapter 1.) Several features of program implementation are explored. First, the chapter examines how well the districts implemented the BELL program relative to the intended model and relative to standards in the field of summer learning, and it looks at whether there were any challenges to implementation. Second, the chapter examines whether students' average daily attendance in the program met BELL's internal quality standards and how this affected the amount of academic instruction that students received during the summer. Finally, the chapter explores whether the academic and typical summer activities of students who were admitted to BELL (the BELL group) differed from the experiences of students who were not admitted to the program (the non-BELL group). Exploring these factors is important for learning more about the conditions that can foster or challenge a summer program's success. The study's key findings are summarized below.

- **How well was the BELL program implemented in the study districts?** Overall, in summer 2012, the program was well implemented relative to the BELL middle school model. In all three study districts, program leaders (program managers, assistant program managers, and lead teachers) expressed that teachers were of high quality and were performing strongly in the program. The academic instruction offered by BELL was also strong relative to national quality standards of summer learning programs.

- **Were there any challenges to program implementation?** In summer 2012, there were two main challenges to implementation. First, all the BELL program leaders reported delays in receiving program materials and diagnostic testing data. This start-up challenge may have been exacerbated by the fact that student recruitment for the study continued until shortly before the start of the program, and the curriculum vendor was experiencing a backlog. Second, BELL teachers — all of whom are certified — reported that they would have benefited more from the staff training if it had focused on the BELL curricula, rather than on instructional practices and pedagogy. (BELL has made several changes to its model since summer 2012, and some of them aim to address these challenges.)

- **How often did students attend the program? How many hours of instruction did they receive?** In the average study district in summer 2012, the attendance rate among students who attended at least one day of the program was 82 percent, which is above BELL's internal monitoring target of 80 percent. Students in the BELL group received, on average, about 23 hours of academic instruction per subject area.

- **How do the summer activities of BELL students differ from the experience of non-BELL students?** In summer 2012, BELL students in the average study district received about 18 more hours of formal instruction (per subject area) than non-BELL students. Although BELL students did not write poems, letters, or stories more often than non-BELL students, they did report playing math games or doing math problems more often. Also, participating in BELL did not prevent students from engaging in other summer activities: BELL students were not less likely than non-BELL students to play sports, watch TV, go to camp, read a book, or go the library during free time.

This chapter discusses each of these topics in detail. Because the purpose of this study is to examine the effect of the average BELL program, the three study districts (Districts A, B, and C) are weighted equally when presenting the pooled findings in this chapter. Thus, the pooled results in this chapter are outcomes for the average study district. (Box 1.3 in Chapter 1 explains how to interpret the findings in this report's tables.)

Program Implementation

This section examines the implementation of BELL's middle school program in the three study districts in summer 2012, relative to the intended model and relative to standards for high-quality programs from the field of summer learning. Prior research on academic summer programs suggest that programs should include several key elements if they are to improve student outcomes. A recent study by RAND has synthesized these recommendations into a set of program quality indicators.[1] The National Summer Learning Association (NSLA) has also developed a set of program quality measures that, in some areas, overlap with those identified in the RAND study, as well as some program dimensions that the RAND study does not include.

As shown in Table 2.1, the key elements of the BELL middle schools model — as well as practices that are recommended by the field of summer learning — can be grouped into three

[1]Sloan McCombs et al. (2011).

Table 2.1

Key Dimensions of Academic Summer Programs

Dimension	Key Component of BELL Middle School Model	RAND/NSLA Quality Indicator
Staffing and training	Positive adult role models	Staff empowerment
	Strong site managers	Quality of staff training and development
Academic instruction	Engaging and age-appropriate reading and math instruction	Small class sizes
	Opportunities for success	Differentiated instruction
		High-quality instruction
		Alignment between school year and regular school year
Student attendance	--	Practices for ensuring student participation and attendance

NOTE: Quality indicators are drawn from RAND and National Summer Learning Association (NSLA).

categories, or dimensions: staffing and training, academic instruction, and student attendance.[2] For each dimension, this section examines the implementation of the BELL middle school model relative to the intended program components and relative to the RAND/NSLA quality indicators.

The findings in this section are based primarily on interviews with program leaders (program managers, assistant program managers, and lead teachers) in the five study schools that were visited (one school in District A, one school in District B, and three schools in District C), as well as findings from the BELL teacher survey. Focus groups with academic teachers and mentors were also used to understand program implementation from the perspective of the teaching staff who delivered the instruction to students.[3]

[2]The BELL model and field recommendations also cover elements related to community involvement and parental engagement. In this evaluation, however, these dimensions were not assessed as thoroughly.

[3]At the school in District A, one focus group was conducted with teachers. However, in Districts B and C, two focus groups with teachers were conducted at each school — one with academic teachers and one with enrichment teachers. The findings for Districts B and C in this section are based on the focus groups for academic teachers.

Staffing and Training

Providing positive adult role models and strong site managers are two explicit components of the BELL model. BELL also provides pre-program training, as recommended by the field of summer learning. Thus, this section discusses three main topics related to staffing in summer 2012: the characteristics and training of the teaching staff, the characteristics of site managers, and the extent to which the BELL program provided positive adult role models.

In general, the findings indicate that BELL succeeded in its objective of hiring strong program managers and providing positive role models for students. BELL was also able to hire highly qualified teachers. With respect to staff training, teachers reported that they would have preferred training that acknowledged their level of teaching experience.

Characteristics and Training of the Teaching Staff

As explained in Chapter 1, academic instruction in the BELL program is provided by certified teachers. Each academic teacher is also assisted by a mentor (teaching assistant), who helps the teacher with classroom management and with small-group instruction.

- **BELL's teaching staff is highly qualified; in summer 2012, almost 70 percent of teachers had a master's degree or a doctorate, and 89 percent had at least five years of teaching experience.**

Table 2.2 presents the characteristics of BELL's academic teachers in the three study districts, based on the teacher survey administered by the program in summer 2012.[4] These findings confirm that BELL academic teachers are highly qualified. In the average study district, almost 70 percent of teachers had completed a master's degree or a doctorate, and 89 percent had at least five years of teaching experience. In two Districts A and B, about 60 percent of teachers worked at the same school during the regular school year. In District C, however, most teachers worked at other schools (not the school where the summer program was operating) during the regular school year.

In terms of staff training, BELL provided teachers and mentors with a combination of online and in-person training. BELL teachers took a nine-hour online training, called "BELL University," to be completed before the in-person training. The in-person training was a full day where teachers — as well as mentors — were trained together by national BELL staff and individuals hired to conduct the trainings.

[4]The findings in Tables 2.2 and 2.3 are based on academic teachers and on dual academic-and-enrichment teachers. They exclude enrichment teachers.

The Evaluation of Building Educated Leaders for Life (BELL)

Table 2.2

Characteristics of BELL Academic Teachers, Overall and by District

Characteristic in Spring 2012 (%)	Average Across Districts[a]	By District		
		District A	District B	District C
Grade level taught				
Grade 5	40.00	--	40.00	--
Grade 6	40.00	50.00	20.00	50.00
Grade 7	46.67	50.00	40.00	50.00
Teacher education				
Completed bachelor's degree	16.67	16.67	--	33.33
Some master's coursework	13.89	16.67	--	25.00
Completed master's degree	61.11	41.67	100.00	41.67
Some doctoral coursework	2.78	8.33	--	--
Completed doctorate degree	5.56	16.67	--	--
Teaching experience				
First time teaching	0.00	--	--	--
1 year	2.78	8.33	--	--
2-4 years	8.33	16.67	--	8.33
5-9 years	45.00	33.33	60.00	41.67
10 or more years	43.89	41.67	40.00	50.00
Teacher role at BELL				
Academic teacher - ELA	36.67	--	60.00	50.00
Academic teacher - Math	31.67	25.00	20.00	50.00
Dual teacher - Academic and Enrichment	47.50	75.00	20.00	--
Previous experience with BELL	0.00	0.00	0.00	0.00
Works at the same school during the school year	42.22	58.33	60.00	8.33
Sample size	29	12	5	12

SOURCE: MDRC calculations based on the BELL teacher survey administered in summer 2012.
NOTE: This analysis is based on teachers who responded to the BELL teacher survey and who taught students in the study sample. Rounding may cause slight discrepancies in calculating sums and differences.
[a]Each of the three districts is given an equal weight when calculating the results in the "Average Across Districts" column.

- **In summer 2012, BELL's training was well aligned with the qualifications of the mentors (teaching assistants) but less well aligned with the qualifications of teachers.**

Findings from the BELL teacher survey are shown in Table 2.3 and suggest that teachers' perceptions of the training in summer 2012 tended toward the positive but that there might be room for improvement. In the survey, teachers were asked to rate various aspects of the BELL program on a 5-point scale where 1 represents strong disagreement and 5 represents

The Evaluation of Building Educated Leaders for Life (BELL)

Table 2.3

Teacher Perceptions of BELL Summer Program: BELL Academic Teachers

BELL Program Characteristic	Average Across Districts[a]	By District		
		District A	District B	District C
Staffing and training				
Usefulness and adequacy of				
BELL's preparation and training (1-5)	3.7	3.3	3.7	4.0
BELL's academic resources and materials (1-5)	3.5	3.9	3.2	3.5
Level of support from BELL leadership team (1-5)	3.9	3.5	3.8	4.4
Quality of teacher's relationship with students (1-5)	4.3	4.5	4.2	4.2
Academic instruction				
Quality of teacher's classroom management (1-5)	4.2	4.3	4.1	4.1
Student engagement in the program (1-5)	4.3	4.2	4.2	4.5
Usefulness and adequacy of BELL's behavior management system (1-5)	3.9	3.5	4.2	4.1
Within 5 days of test administation (%)				
Teacher received Stanford results	51.8	91.7	0.0	63.6
Teacher received quiz reports	57.6	81.8	0.0	90.9
Number of weeks needed to determine				
Academic issues of each student in the class	2.0	2.1	2.0	2.0
Behavioral issues of each student in the class	1.4	1.6	1.4	1.3
Learning styles of each student in the class	1.8	2.0	1.4	2.1
Sample size[b]	29	12	5	12

SOURCE: MDRC calculations based on the BELL teacher survey administered in summer 2012.
NOTES: This analysis is based on teachers who responded to the BELL teacher survey and who taught students in the study sample. Measures with a scale of 1 to 5 were constructed from teachers' responses to a set of survey items that have a 5-point agreement scale: 1 = "strongly disagree," 2 = "disagree," 3 = "undecided," 4 = "agree," and 5 = "strongly agree." Rounding may cause slight discrepancies in calculating sums and differences.

[a]Each of the three districts is given an equal weight when calculating the results in the "Average Across Districts" column.

[b]Due to missing values, the number of teachers included varies by characteristic.The sample size reported here is for sample of academic teachers who responded to at least one item on the survey. The percentage of missing data on any given characteristic does not exceed 7 percent.

strong agreement. Composite measures representing teachers' perceptions of different program features were created by averaging teachers' responses across relevant items.[5] BELL academic

[5]1= "Strongly disagree"; 2 = "Disagree"; 3 = "Undecided"; 4 = "Agree"; and 5 = "Strongly agree." The internal consistency reliability (Cronbach's alpha) of scales constructed from the survey and shown in Table 2.3 ranges from 0.80 to 0.95. Appendix B describes the items included in each survey scale.

teachers in the average study district gave BELL's training an average rating of 3.7, which indicates that they were somewhere between being "undecided" and in "agreement" that the training was useful and adequate.

Data collected from interviews with program leaders and focus groups with teachers point to specific areas of the training that garnered more mixed reviews. The most consistent feedback from teachers was about the alignment between the training and teachers' qualifications and experience.[6] Although BELL teachers are highly qualified and experienced, none of them had previous experience with the BELL curriculum (Table 2.2). Yet the focus of the training was not on the curriculum: Teachers reported that the training focused on instructional practices that they had learned prior to becoming certified and did not focus enough on the BELL program's content. The following sentiment is representative of what was heard from many of the teachers who participated in focus groups: "We know how to teach. We were taught how to teach. That's how we got here…. What we really needed was access to the curriculum so that we would be prepared up front." Thus, while most mentors — because of their limited classroom experience — found the training to be very instructive, the focus group teachers were less satisfied with it.

When asked to reflect on the fit of the teacher training, senior staff at BELL headquarters explained that, in previous years, most BELL teachers were more inexperienced, and so the training did not presume that teachers were familiar with best practices in teaching. This senior staff person explained that, more recently, there has been a shift in the composition of BELL's teaching staff and that more seasoned teachers are coming to teach in the BELL program.

Characteristics of Program Managers

Each school operating the BELL program has a program manager who oversees instruction and discipline. In most of the study schools in summer 2012, program managers were principals or assistant principals either at the school where they managed the BELL summer program or at another school in the district. In District A's school, the program manager was a seasoned principal who had been running summer programs for over 15 years. In the school in District B, the program manager was a seasoned teacher who had been in education for over 20 years and was currently the department chair of her discipline. In the three schools in District C, the program managers were also assistant principals.

- **In summer 2012, BELL teachers and mentors had positive perceptions of the program managers.**

[6]In all interviews and focus groups, staff training was cited as an area for improvement.

Data collected for the study indicate that BELL met its goal of hiring strong program managers in summer 2012. Across all three study districts, these managers received high praise from other program leaders as well as from teachers and mentors:

- In District A, most teachers characterized the program manager as consistent, supportive, or easy going,[7] and most mentors praised the program manager for being effective, approachable, accessible, and willing to help.[8] Mentors and teachers alike gave examples of how this manager constructively handled two different serious student disciplinary issues.[9]

- In District B, the program manager's style was described by one program leader as "clear and direct." Three of seven mentors also characterized the program manager as "no nonsense." Several mentors also noted that the program manager was "hands-on," competent, and supportive of their behavioral management role when then needed "backup."[10] All teachers in the focus group appreciated the fact that the program manager was up-front and honest when she did not know something. Teachers further noted that students seemed to respect the program manager.[11]

- In District C, one of the program managers was characterized by a program leader as follows: "You know, she pretty much knows how everything's suppose to line up. She's very efficient. She's very fair." In focus groups, teachers and mentors agreed with these observations. Most academic teachers expressed that they liked her management style and described it as "no nonsense" or as eliciting a strong and respectful response from students.[12] About another District C program manager, all academic teachers,[13] program leaders, and mentors offered praises like "excellent," "awesome," and "inspiring." Teachers noted how this program manager took the time to talk with each child at breakfast, which seemed to be a meaningful gesture to both students and staff alike.[14] Additionally, this program manager communicated with all the teachers and staff via a blog that he updated daily. All the information that staff needed to know was posted on this blog.

[7]This is based on four of seven teachers in the focus group for this school.
[8]This is based on five of six mentors in the focus group for this school.
[9]This is based on one teacher and five mentors in the focus group for this school.
[10]This is based on four of seven mentors in the focus group for this school.
[11]This is based on three of six teachers in the focus group for this school.
[12]This is based on three of four teachers in the focus group for this school.
[13]This is based on four of four teachers in the focus group for this school.
[14]This is based on two of four teachers in the focus group for this school.

Data from the teacher survey (Table 2.3) support these positive reviews. In the average study district, teachers "agreed" (a score of 3.9 out of 5) that they had received strong support from the BELL leadership team.

However, during interviews and focus groups, staff also provided constructive criticism of the managers. In District A, for example, many teachers thought that the program manager's "laid-back" disposition led to the manager's seeming to be disorganized and too easy on the students.[15] In District B, teachers expressed that while students respected the program manager, the manager seemed unapproachable, so students preferred to go to the assistant program manager with any issues or questions.[16] The different management styles of the program managers illustrate that different approaches can lead to strong leadership but that certain qualities may also have drawbacks.

Positive Adult Role Models

Providing positive role models to students is one of the elements of BELL's middle school model. In all three districts, program leaders (managers and lead teachers) noted that teachers and mentors played this role in summer 2012. In District B, for example, a program leader expressed that strong positive role models were provided "with our TAs [teaching assistants], our teachers, and just in our daily communication with them, helping [them] to understand what they need to succeed, how to handle different situations positively." These findings from program leaders are supported by responses to the teacher survey (Table 2.3): Teachers in the average study district "agreed" (a score of 4.3 out of 5) that that they had a strong positive relationship with their students.

Academic Instruction

The two most important goals of the BELL model are to provide engaging and age-appropriate reading and math instruction and to provide opportunities for student success. In addition, the field of summer learning further recommends that class sizes be small and that academic instruction be high quality, differentiated, and aligned with or informed by knowledge of the regular school year's activities. Thus, this section examines the following aspects of academic instruction in the BELL study districts in summer 2012: classroom organization (including class size and management), teacher quality and student engagement, and instructional differentiation and program materials.

In general, the findings suggest that high-quality academic instruction was offered across all three study districts in summer 2012. In the first two weeks of the program, there were

[15]This is based on four of six teachers in the focus group for this school.
[16]This is based on three of six teachers in the focus group for this school.

challenges in getting program materials and diagnostic tests to teachers on time. But these start-up issues had been resolved by the time of the site visits, and program leaders did not feel that they had affected instructional quality.

Classroom Organization and Management

- **In summer 2012, teachers had positive perceptions of classroom management, in part because of the assistance that they received from mentors (teaching assistants).**

Each BELL academic teacher is assisted by a mentor, which means that, at most, there was a student-to-instructor ratio of 10:1.[17] The way in which the teachers and mentors worked together differed across classrooms. In some classrooms, teachers used the mentor as a co-teacher who taught parts of lessons or assisted groups of students with class work while the teacher assisted other students. In other instances, mentors played a less active role in instruction, primarily focusing on behavioral management and such administrative tasks as attendance. The teacher survey results (Table 2.3) corroborate that this system appears to have worked well for them. In the average study district in summer 2012, teachers felt that their classroom management was strong (a score of 4.2 out of 5), and they reported that BELL's behavior management system was useful (a score of 3.9 out of 5).

Teacher Quality and Student Engagement

- **In all three study districts in summer 2012, program leaders expressed that teachers were of high quality and that teachers used various strategies to engage students.**

Based on interviews with program leaders, the instruction provided by BELL teachers seems to have been strong and engaging to the students in summer 2012. When asked about academic and enrichment instruction, all program managers and lead teachers commented on the strength of the teachers or the high quality of instruction that they witnessed in the classrooms. One program leader in District C remarked: "I think these teachers are absolutely wonderful. Some of the things that they have been doing in the classroom have been phenomenal.… When you go to the classrooms, you just see a lot of good things happening, and the kids seem to be excited about what they're doing."

Program leaders gave several examples of the strategies used by teachers to engage students and to give them an opportunity to succeed. A program leader in District A offered the

[17]This is based on the number of students in each classroom (which was 20 students or less) and the number of instructional staff per classroom (two staff, including the teacher and the mentor). Data collected during site visits confirmed that there was always a mentor present in classrooms.

following thoughts on practices that seemed to be especially beneficial to students: "I think they assist the kids at the very beginning, and then they differentiate instruction to make sure every kid is successful." In District B, a program leader noted that teachers were providing students with positive reinforcement: "With the completion of the program, or even just their daily activities, [we offer] positive praise because we want them to realize that, 'Your efforts are appreciated, and it's gonna pay off.'" A program leader in District C school also pointed out that teachers were integrating the academic and enrichment activities: "I think the teachers are even taking the information from the field trips and implementing that into their classrooms." Findings from the teacher survey (Table 2.3) lend support to the claim that these practices were appealing to students. In the average study district, academic teachers felt that students were engaged in the different aspects of the BELL program (an average score of 4.3 out of 5).

Instructional Differentiation and Program Materials

- **In summer 2012, the study districts experienced delays in receiving program materials and diagnostic testing data, due to a backlog that the vendor was experiencing and delays in recruiting and randomly assigning students. These issues had been resolved by the time of the site visits.**

As explained in Chapter 1, BELL administers diagnostic tests to students early in the first week of the program to help teachers identify the unique needs of each student. The teacher survey (Table 2.3) indicates that, across the three study districts in summer 2012, there was considerable variation in how quickly the results of tests were returned to teachers. In District A, 92 percent of teachers received diagnostic testing results within the first five days of the program. In contrast, none of the teachers in District B had received test results within the first five days.

The start-up delays in District B do not appear to have affected teachers' ability to identify students' needs. Based on the teacher survey (Table 2.3), teachers in District B believed that they had a good grasp of the learning styles of each student 1.4 weeks, on average, into the 5-week program; a good idea of each student's academic issues 2.0 weeks into the program; and a good idea of each student's behavioral issues 1.4 weeks into the program. These results are similar to those for District A, where diagnostic test results were received earlier. In addition, as is discussed in Chapter 3, the most promising findings in terms of effects on student outcomes were found in District B.

The three districts also experienced delays in the arrival of the some instructional materials from the vendor. Start-up delays were noted in all the interviews with program leaders, and they had two causes. First, as noted in Chapter 1, randomization did not occur in District B until one business day before the start of the program; in Districts A and C, randomization occurred 4 and 13 business days, respectively, before the start of the program. This created delays in ordering the right number and types of curricular materials from the vendor, especially in District B.

Second, the problems were compounded because the supplier of the program's curriculum was experiencing a backlog.

To meet these challenges, BELL national came up with creative solutions. Program leaders in all study districts reported that BELL provided sufficient funding for the schools to buy the classroom materials that were not provided by BELL's headquarters. Similarly, in District B, a program leader noted that while the formal curricular and supplemental materials did not arrive until the third week of the program, teachers did receive the missing math curriculum in the early weeks, in the form of photocopied packets.

Despite such setbacks, program leaders reported that the teaching quality did not suffer as a result of delays. In Districts A and B, for instance, program leaders noted that teachers were highly effective even when some materials for a lesson were not available. A program leader in District B stated: "I don't think it has affected the quality. It just made them work harder, and therefore they kinda just adapted and added what they needed to in order to make it work."

Student Attendance

- **In summer 2012, to encourage student attendance, program staff in all three study districts used a variety of strategies, including calling parents and offering incentives to students.**

The field of summer learning recommends that strong practices be put in place to maximize student attendance and participation. Although attendance is not an explicit element of the BELL middle school model, the program's efforts to monitor and maximize attendance are strong relative to best practices recommended by the field. In the first instance, BELL enforces an attendance policy of 80 percent, and this policy is understood and monitored by the regional and national BELL staff. As a national staff person explained: "We set an objective of 80 percent on an average daily basis, so we say, 'Okay, we want — if there's 100 kids at your site, we expect you to have 80 there every single day. If you've got less than that, you're doing something wrong. If you've got more than that, you're doing something very right.'"

BELL study schools implemented a number of approaches to encourage high levels of student attendance and participation. In all study districts in summer 2012, local BELL staff actively called parents when a student was absent to report their child's absence and to reiterate the importance of attendance. Mentors also communicated the importance of attendance to parents when they dropped off or picked up their child at the program. In District A, to encourage attendance, students with perfect weekly attendance could be entered into a weekly raffle to earn green bands or water bottles and to have their name listed on an attendance chart. The other study schools did not have such incentives in place. District A also enforced a policy by which students would be removed from the program after three unexcused absences; in Districts B and

C, however, this policy was applied more leniently. (After the following brief summary of program implementation, the next section of this chapter presents detailed findings on attendance rates in summer 2012.)

Summary of Program Implementation

In general, the implementation of the BELL model in the three study districts in summer 2012 was strong relative to the intended program model and relative to recommendations from the field. Two challenges that surfaced during the site visits are related to program start-up and staff training. First, all the BELL program leaders reported delays in receiving program materials and diagnostic testing data. This start-up challenge may have been exacerbated by the fact that student recruitment for the current study continued until shortly before the start of the program, and the curriculum vendor was experiencing a backlog. By the time of the site visits, these issues had been resolved. Second, BELL teachers — all of whom are certified — reported that they would have benefited more from the staff training if it had focused on the BELL curricula, rather than on instructional practices and pedagogy.

Student Attendance and the Amount of Academic Instruction Received

In order for an academically oriented summer program like BELL to have a positive impact on students' academic achievement, students must attend the program to receive the instruction that is provided. Maintaining high attendance rates is especially important for summer programs of short duration, because the amount of instruction offered is limited.[18]

However, getting students to attend a voluntary summer program is inherently challenging. First, there are structural barriers to attendance — such as the lack of transportation, family vacations, responsibilities for siblings, and parents' work schedules. In addition, students face motivational barriers to attendance, because an all-day program like BELL's reduces the amount of time that they could spend with their friends or engage in other extracurricular activities. Motivational barriers like these are likely to be even more formidable for middle school students than for younger children.

Table 2.4 presents average daily attendance rates in the three study districts and across the districts, based on data collected by BELL teachers in summer 2012. Attendance rates are

[18]Beckett et al. (2009).

The Evaluation of Building Educated Leaders for Life (BELL)

Table 2.4

Summer Program Attendance of BELL Students: Fall 2012 Analysis Sample

Attendance in Summer 2012	Average Across Districts[a]	By District		
		District A	District B	District C
Attendance among all students in the BELL group[b]				
Number of days attended	19	20	21	16
Attendance rate (%)	75.9	81.5	80.0	66.3
Attended at least 1 day of the program (%)	92.3	91.8	93.7	91.4
Sample size[b]	585	279	63	243
Attendance among students who attended at least 1 day[c]				
Number of days attended	21	22	22	17
Attendance rate (%)	82.2	88.8	85.4	72.5
Sample size[c]	537	256	59	222

SOURCE: MDRC calculations based on summer attendance records provided by BELL teachers.

NOTES: The analyses reported in this table are based on the sample of students in the BELL group who took the GRADE and GMADE assessments and who responded to the student survey in fall 2012 (Fall 2012 Analysis Sample). The average daily attendance rate is defined as the the average percentage of total program days that students attended. The number of program days offered was 25 in District A, 26 in District B, and 24 in District C.

[a]Each of the three districts is given an equal weight when calculating the results in the "Average Across Districts" column.

[b]This includes students in the BELL group who never attended the program ("no-shows").

[c]This excludes students in the BELL group who never attended the program ("no-shows").

.

shown for two groups of students. The first group includes all students who were invited to participate in BELL (the BELL group), including students who did not participate at all. This attendance rate is most useful for interpreting the impact findings, because it measures the average amount of instruction received by all students in the BELL group. The second group includes the 92 percent of students in the BELL group who *attended at least one day of the program.* This rate is most useful for gauging whether BELL's attendance policies were successful at helping meet the operational target of 80 percent attendance, since BELL monitors attendance for program participants only.[19]

- **In the average study district in summer 2012, the attendance rate among all students invited to participate in BELL (the BELL group) was 76 percent. The attendance rate among students who attended at least one day of the program was 82 percent.**

[19]The average daily attendance rate is defined as the percentage of total program days that students attended, averaged across the relevant group of students.

As explained above, BELL encourages regular attendance using a combination of strategies, including calls to parents and giving incentives for attendance. BELL also actively enforces a policy of 80 percent attendance. The findings in Table 2.4 indicate that the strategies used by BELL to promote attendance were largely successful. In the average study district, the daily attendance rate among students who attended at least one day of the program was 82 percent, which exceeds BELL's target. This is especially noteworthy, given that students in the program participated voluntarily and that motivating middle school students is an especially difficult task. Thus, it *is* possible to operate an all-day, full-week summer academic program that middle school students will attend.

That said, there was also variation in the average attendance rate across the three study districts in summer 2012. In Districts A and B, the daily attendance rate was greater than 80 percent, both among all students in the BELL group and among the subset of students who attended for at least one day. In District C, however, daily attendance was only 73 percent even among students who attended at least one day of the program, which is below BELL's internal benchmark of 80 percent.

For context, it is important to note that truancy is an ongoing concern in District C — during the academic year as well as the summer — so BELL was operating in a challenging context in this respect. To reduce truancy during the school year, District C had hired social workers in each school to help interface with parents and to serve as a resource for students with particularly difficult home environments. The district provided these social workers to the BELL program as well, as a supplemental resource to help decrease truancy rates. Thus, although student attendance in BELL did not meet the target of 80 percent, given the context in which BELL was operating, an attendance rate of 73 percent is noteworthy. As a social worker shared, "There are several students who may have come with a history of attendance problems."

- **Given their attendance rate, students who were invited to participate in BELL (the BELL group) received about 23 hours of academic instruction per subject area.**

As explained in Chapter 1, BELL students are offered 30 hours of ELA instruction and 30 hours of math instruction across all five weeks of the program.[20] Given their average daily attendance rates (76 percent among all students and 82 percent among those who attended at least one day of the program), students in the BELL group in the average study district received about 23 hours of instruction in reading and 23 hours in math.[21] In Districts A and B, students in

[20]Each week for five weeks, students receive six hours of instruction per subject area per week, for a total of 30 hours per subject area.

[21]The number of hours of instruction received is equal to the average attendance rate among all BELL students (76 percent) times 30 hours of instruction provided in each subject area.

the BELL group received 24 hours of instruction in each subject area; in District C, students received about 20 hours of instruction in each subject area.[22]

To inform the field of summer learning, it is also worth examining why some students may not have attended the program as regularly, from the perspective of students. In the fall survey, students were asked to rate how often different factors prevented them from attending the BELL program (where a rating of 0 represents "Never" and a rating of 4 represents "Pretty often"). Table 2.5 presents the findings from this survey item. As reflected in the low ratings in this table, none of the listed factors was a significant barrier in terms of preventing students from attending the program, which is to be expected, given the generally high BELL program attendance rates. However, the most common self-reported reason for not attending BELL was a conflict with family vacations (which could include visiting relatives) and a conflict with more

The Evaluation of Building Educated Leaders for Life (BELL)

Table 2.5

Reasons That BELL Students Missed Days of Summer Program: Fall 2012 Analysis Sample

Attendance in Summer 2012	Average Across Districts[a]	By District		
		District A	District B	District C
Reasons for missing program days (0-4)[b]				
Had another activitity wanted to go to more	1.1	1.0	1.1	1.2
Could not get to or from the program	0.7	0.8	0.9	0.6
Did not like the program	1.0	0.6	1.1	1.2
Parent wanted student to do something else	0.6	0.5	0.7	0.6
Family went on vacation	1.1	1.0	1.1	1.3
Sample size[c]	324	142	38	144

SOURCE: MDRC calculations based on the fall 2012 student survey.

NOTES: The analyses reported in this table are based on the sample of students in the BELL group who took the GRADE and GMADE assessments and who responded to the student survey in fall 2012 (Fall 2012 Analysis Sample).

[a]Each of the three districts is given an equal weight when calculating the results in the "Average Across Districts" column.

[b]Students were asked to report how frequently they missed a day of their summer program for a given reason, using the following scale: 0 = Never; 1 = Hardly ever; 2 = Not very often; 3 = Sometimes; 4 = Pretty often.

[c]Includes only students who reported having missed days of the summer program in the survey.

[22]The number of hours of instruction received in each district is equal to the average attendance rate among all BELL students in each district (81 percent in District A, 80 percent in District B, and 66 percent in District C) times 30 hours of instruction provided in each subject area.

interesting activities that students wanted to attend instead. These findings are supported by interviews with program staff. In District C — which had the lowest program attendance rate — program leaders in all visited schools mentioned that they had to deal with parents pulling students out of the program for a week at a time for family vacations. At one school in District C, a program leader further mentioned that the parents of students involved in summer sports programs had been called, because their children were missing classes in the morning for sports practice and coming to the program only in the afternoon.

The Summer Activities of BELL and Non-BELL Students

An important driver of BELL's potential to improve student outcomes is the extent to which the program provides its students with summer services that are different from what they would have received otherwise. If students in the non-BELL group are engaging in unstructured or typical summer activities, then there is greater potential for a program like BELL to have an impact on the academic achievement of the students that they serve. In contrast, if non-BELL students also participated in structured academic enrichment activities during the summer, then this limits the extent to which students who participated in BELL can benefit from the program relative to what they would have experienced otherwise.

Based on data from the fall 2012 student survey, Table 2.6 examines the contrast in the summer activities between BELL and non-BELL students in summer 2012.[23] The survey asked students about the number of times that they participated in different types of academic and typical summer activities. Because the survey was administered in the fall, the mean frequencies in Table 2.6 should be interpreted as the *minimum* number of times that students participated in these activities, since students may have not have been able to remember all their summer activities. Fortunately, due to random assignment, the magnitude of this recall error should be the same for BELL and non-BELL students. This means that the *difference* in summer activities between the two groups of students should not be affected by recall bias and that, therefore, the results in Table 2.6 can be used to understand the ways in which BELL and non-BELL students' summer experiences differed in summer 2012. (Box 1.3 in Chapter 1 explains how to interpret the findings in this report's tables.)[24]

[23]These findings are based on students in the Fall 2012 Analysis Sample. For more information about the student survey, see Appendix B.

[24]As explained in Chapter 1, students' summer activities are secondary outcomes in this report because they are used to contextualize impacts on academic achievement (the primary outcomes). Thus, statistically significant differences between the summer activities of BELL and non-BELL students need not be adjusted for multiple hypothesis testing, based on standards provided by What Works Clearinghouse (2014).

The Evaluation of Building Educated Leaders for Life (BELL)

Table 2.6

Summer Activities of Students in the Fall 2012 Analysis Sample, by Treatment Group

Activity in Summer 2012	BELL Group	Non-BELL Group	Estimated Difference	Effect Size	P-Value for Estimated Difference
Summer program participation (%)					
Student went to a program that was academically focused	83.6	21.1	62.5 ***	1.50	<0.001
Math and reading activities	79.4	11.9			
Mostly reading or mostly math activities	4.2	9.2			
Student went to a nonacademic summer program	6.6	11.2	-4.6 **	-0.16	0.036
Student did not go to a summer program	9.8	67.7	-57.9 ***	-1.24	<0.001
Total number of times that student engaged in academic activities during the summer[a]	13.3	9.8	3.5 ***	0.29	0.004
Wrote a letter, poem, or story	4.3	3.8	0.5	0.07	0.434
Played math games or did math problems	9.0	6.0	3.0 ***	0.35	<0.001
Total number of times that student engaged in typical activities during the summer	42.7	46.1	-3.4	-0.15	0.118
Went to a library	4.0	4.9	-0.9	-0.13	0.145
Read a book during free time	7.1	7.2	-0.1	-0.01	0.922
Watched TV during the day on a weekday	15.4	16.4	-1.0	-0.10	0.302
Did activities at a club, community center, church, or day camp	7.7	8.0	-0.3	-0.03	0.729
Played in a sports program	8.4	9.5	-1.1	-0.10	0.261
Sample size (N = 919)	585	334			

(continued)

Table 2.6 (continued)

SOURCE: MDRC calculations based on the BELL student survey administered in fall 2012.

NOTES: The analyses reported in this table are based on the sample of students who took the GRADE and GMADE assessments and who responded to the student survey in fall 2012 (Fall 2012 Analysis Sample). The estimated differences between the BELL group and the non-BELL group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012, as well as random differences between the BELL and non-BELL groups with respect to the following variables: a student's' score on state reading and math tests taken in spring 2012, whether a student has an individualized education plan (IEP), whether the student has English as a Second Language (ESL), whether a student is eligible for free or reduced-price lunch, parent education, race/ethnicity, and gender. The values in the column labeled "BELL Group" are the observed means for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-BELL group, using the observed mean covariate values for the BELL group as the basis for the adjustment. Each of the three study districts is given an equal weight when estimating the results reported in this table. Rounding may cause slight discrepancies in calculating sums and differences.

Effect sizes are calculated by dividing the estimated difference by the standard deviation of the summer activity measure for students in the Fall 2012 Analysis Sample who are in the non-BELL group.

A two-tailed t-test was used to test differences between the BELL and non-BELL groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]Students reported the frequency of each summer activity on a 5-point scale, which was converted to number of times per summer as follows: "Never" = 0; "Hardly ever" (1 or 2 times) = 1.5; "Not very often (once a month)" = 3; "Sometimes (about once a week)" = 13; "Pretty often (a couple times or more a week)" = 26.

Academic Activities

The first two panels of Table 2.6 examine whether BELL students participated in a greater number of academically focused summer activities than non-BELL students.

- **In summer 2012, BELL students received about 18 more hours of academic instruction per subject area than non-BELL students.**

As shown in the first panel of Table 2.6, 84 percent of students in the BELL group reported going to a summer program that was academically focused.[25] However, a substantial proportion of non-BELL students (21 percent) also ended up receiving academic instruction from other programs during the summer.[26] Assuming that the programs attended by these non-BELL students provided the same amount of academic instruction as BELL — and that the attendance rate at these programs was the same as for BELL — then non-BELL students in the study's average district would have received about 5 hours of instruction in each subject area (compared with 23 hours for students admitted to BELL, or a difference of 18 hours between the two groups).[27]

The second panel of Table 2.6 looks at the extent to which BELL and non-BELL students engaged in two types of academic activity that are part of the BELL program: (1) writing a letter, poem, or story and (2) playing math games or doing math problems. All else being equal, BELL students would be expected to engage in these activities more often than non-BELL students.

---

[25]As shown in Table 2.4, based on attendance records, 92 percent of students in the BELL group attended the program. In the survey, however, 84 percent of students reported that they had attended an academically focused program in the summer. This discrepancy is due to recall error: students forgetting what they did over the summer. As noted above, the degree of recall bias should be the same for BELL and non-BELL students; therefore, the service contrast is unbiased.

[26]It is unclear what academic programs these students were attending. A very small number of them received summer services from BELL, whether inadvertently or mistakenly (2 percent of non-BELL students ended up enrolling in the program). However, the remaining 19 percent of non-BELL students who received academic instruction would have gotten it from a program other than BELL. Only a few programs were specifically mentioned as alternatives by BELL staff (three programs were noted in District A; two math or science camps were noted in District B; and one summer program was mentioned in District C).

[27]The total hours of instruction received by BELL students (23 hours) is equal to 92 percent (the proportion of BELL students who attended at least one day of the program) times 82 percent (the attendance rate among students who attended at least one day of the program) times 30 hours of instruction provided in each subject area. The number of hours of instruction received by non-BELL students (5 hours) is equal to 21 percent (the proportion of non-BELL students who attended an academic summer program) times 82 percent (the attendance rate among BELL students who attended at least one day of the program, which is assumed to be the same for non-BELL students enrolled in other summer programs) times 30 hours of instruction in each subject (the amount of instruction provided by BELL, which is assumed to be the same for the academic programs attended by non-BELL students).

- **In the average study district in summer 2012, students in the BELL group did not report writing a letter, poem, or story more often than students in the non-BELL group.**

BELL and non-BELL students reported writing a letter, poem, or story at about the same frequency over the summer. In the average study district, students in both groups participated in these types of activities about 1.3 times per month (about 4 times during the summer).[28]

- **In the average study district in summer 2012, students in the BELL group reported playing math games or doing math problems more often than non-BELL students.**

On average, students in the BELL group played math games or did math problems nine times during the summer (three times per month) while the non-BELL group did so six times (two times per month). This difference is statistically significant.

Typical Summer Activities

The bottom panel of Table 2.6 focuses on activities that are more typical of what students do during the summer: going to the library; reading a book during free time; watching TV on a weekday; going to a club, community center, church, or day camp; and playing in a sports program. Because students spend their day at BELL, attending the program could potentially "crowd out" the amount of time that students have available to engage in these typical activities. For example, BELL students might read less often during their free time than non-BELL students — and might play sports less frequently — because they have less spare time to engage in these activities.

Based on the survey findings, however, it does not appear as though attending BELL prevented students from participating in these summer activities. For instance, both BELL and non-BELL students reported reading a book during their free time about once every two weeks (seven times during the summer). Both groups of students were also equally likely to play sports or to attend other organized enrichment programs; both groups engaged in these activities about once every two weeks, or eight to nine times during the summer.

Summer Activities, by District

Tables 2.7 to 2.9 present the contrast in the summer activities of BELL and non-BELL students in each study district in summer 2012. Two findings are notable:

[28]These calculations assume that the summer has three months and 13 weeks.

- The contrast in students' writing activities is not statistically significant in any of the three districts, though the size of the contrast is largest in District B (Table 2.8).

- The contrast in students' math activities is statistically significant in Districts B and C. In District B (Table 2.8), BELL students played math games almost four times per month, on average, whereas non-BELL students played math games about two times per month. In District C (Table 2.9), the contrast in math-related activities is smaller than in District B, but it is also statistically significant. In District A (Table 2.7), the contrast in math activities is smallest in size and is not statistically significant.

Summary

The findings reported in this chapter indicate that, in summer 2012, the BELL program was well implemented in the study districts and that student attendance met BELL's internal target of 80 percent in all but one district. Overall, BELL students received about 18 hours more reading and math instruction that non-BELL students. BELL students reported playing math games and doing math problems more often during the summer, though they did not engage in writing activities (writing letters, poems, or stories) more often than non-BELL students. The findings also point to two challenges. First, delays in providing teachers with complete curricular materials and diagnostic tests arose (due in part to problems with the curriculum vendor and exacerbated by the study's random assignment process). Second, BELL's teacher training was not well aligned with the qualifications of its teachers. In particular, teachers expressed that they would have benefited more from the training if it had focused more on the BELL curricula, rather than on general instructional practices. Challenges related to start-up and the appropriate tailoring of staff training may be typical of summer programs, because they start fresh every summer and must hire and train a new set of instructors each year. As is discussed in Chapter 4, BELL has made several changes to its program model since 2012 — some of which aim to address these challenges.

The Evaluation of Building Educated Leaders for Life (BELL)

Table 2.7

Summer Activities of Students in the Fall 2012 Analysis Sample, by Treatment Group: District A

Activity in Summer 2012	BELL Group	Non-BELL Group	Estimated Difference	Effect Size	P-Value for Estimated Difference
Summer program participation (%)					
Student went to a program that was academically focused	91.0	22.8	68.2 ***	1.63	<0.001
Math and reading activities	85.7	4.0			
Mostly reading or mostly math activities	5.4	18.8			
Student went to a nonacademic summer program	2.9	3.0	-0.1	0.00	0.963
Student did not go to a summer program	6.1	74.2	-68.1 ***	-1.46	<0.001
Total number of times that student engaged in academic activities during the summer [a]	12.0	10.8	1.2	0.10	0.475
Wrote a letter, poem, or story	3.5	3.0	0.4	0.06	0.637
Played math games or did math problems	8.6	7.8	0.8	0.09	0.530
Total number of times that student engaged in typical activities during the summer	36.1	40.4	-4.3	-0.19	0.155
Went to a library	3.3	3.5	-0.2	-0.03	0.783
Read a book during free time	7.3	9.6	-2.3 **	-0.26	0.044
Watched TV during the day on a weekday	14.6	16.0	-1.4	-0.13	0.320
Did activities at a club, community center, church, or day camp	4.2	4.0	0.2	0.02	0.895
Played in a sports program	6.7	7.3	-0.6	-0.05	0.674
Sample size (N = 358)	279	79			

(continued)

Table 2.7 (continued)

SOURCE: MDRC calculations based on the BELL student survey administered in fall 2012.

NOTES: The analyses reported in this table are based on the sample of students who took the GRADE and GMADE assessments and who responded to the student survey in fall 2012 (Fall 2012 Analysis Sample). The estimated differences between the BELL group and the non-BELL group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012, as well as random differences between the BELL and non-BELL groups with respect to the following variables: a student's score on state reading and math tests taken in spring 2012, whether a student has an individualized education plan (IEP), whether the student has English as a Second Language (ESL), whether a student is eligible for free or reduced-price lunch, parent education, race/ethnicity, and gender. The values in the column labeled "BELL Group" are the observed means for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-BELL group, using the observed mean covariate values for the BELL group as the basis for the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

Effect sizes are calculated by dividing the estimated difference by the standard deviation of the summer activity measure for students in the Fall 2012 Analysis Sample who are in the non-BELL group.

A two-tailed t-test was used to test differences between the BELL and non-BELL groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]Students reported the frequency of each summer activity on a 5-point scale, which was converted to number of times per summer as follows: "Never" = 0; "Hardly ever (1 or 2 times)" = 1.5; "Not very often (once a month)" = 3; "Sometimes (about once a week)" = 13; "Pretty often (a couple times or more a week)" = 26.

The Evaluation of Building Educated Leaders for Life (BELL)

Table 2.8

Summer Activities of Students in the Fall 2012 Analysis Sample, by Treatment Group: District B

Activity in Summer 2012	BELL Group	Non-BELL Group	Estimated Difference	Effect Size	P-Value for Estimated Difference
Summer program participation (%)					
Student went to a program that was academically focused	76.2	21.7	54.5 ***	1.30	<0.001
Math and reading activities	71.4	19.9			
Mostly reading or mostly math activities	4.8	1.8			
Student went to a nonacademic summer program	12.7	20.3	-7.6	-0.26	0.155
Student did not go to a summer program	11.1	58.0	-46.8 ***	-1.00	<0.001
Total number of times that student engaged in academic activities during the summer [a]	16.0	8.7	7.4 **	0.61	0.012
Wrote a letter, poem, or story	5.3	4.0	1.3	0.19	0.397
Played math games or did math problems	10.7	4.6	6.1 ***	0.71	0.006
Total number of times that student engaged in typical activities during the summer	47.5	50.7	-3.2	-0.14	0.545
Went to a library	5.2	7.5	-2.4	-0.34	0.108
Read a book during free time	7.5	6.0	1.6	0.18	0.430
Watched TV during the day on a weekday	15.7	16.3	-0.6	-0.06	0.799
Did activities at a club, community center, church, or day camp	10.2	10.5	-0.3	-0.03	0.876
Played in a sports program	9.0	10.4	-1.4	-0.13	0.546
Sample size (N = 117)	63	54			

(continued)

Table 2.8 (continued)

SOURCE: MDRC calculations based on the BELL student survey administered in fall 2012.

NOTES: The analyses reported in this table are based on the sample of students who took the GRADE and GMADE assessments and who responded to the student survey in fall 2012 (Fall 2012 Analysis Sample). The estimated differences between the BELL group and the non-BELL group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012, as well as random differences between the BELL and non-BELL groups with respect to the following variables: a student's score on state reading and math tests taken in spring 2012, whether a student has an individualized education plan (IEP), whether the student has English as a Second Language (ESL), whether a student is eligible for free or reduced price lunch, parent education, race/ethnicity, and gender. The values in the column labeled "BELL Group" are the observed means for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-BELL group, using the observed mean covariate values for the BELL group as the basis for the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

Effect sizes are calculated by dividing the estimated difference by the standard deviation of the summer activity measure for students in the Fall 2012 Analysis Sample who are in the non-BELL group.

A two-tailed t-test was used to test differences between the BELL and non-BELL groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]Students reported the frequency of each summer activity on a 5-point scale, which was converted to number of times per summer as follows: "Never" = 0; "Hardly ever (1 or 2 times)" = 1.5; "Not very often (once a month)" = 3; "Sometimes (about once a week)" = 13; "Pretty often (a couple times or more a week)" = 26.

The Evaluation of Building Educated Leaders for Life (BELL)

Table 2.9

Summer Activities of Students in the Fall 2012 Analysis Sample, by Treatment Group: District C

Activity in Summer 2012	BELL Group	Non-BELL Group	Estimated Difference	Effect Size	P-Value for Estimated Difference
Summer program participation (%)					
Student went to a program that was academically focused	83.5	18.6	64.9 ***	1.55	<0.001
Math and reading activities	81.1	11.6			
Mostly reading or mostly math activities	2.5	7.0			
Student went to a nonacademic summer program	4.1	10.3	-6.2 ***	-0.21	0.010
Student did not go to a summer program	12.3	71.1	-58.7 ***	-1.26	<0.001
Total number of times that student engaged in academic activities during the summer [a]	11.9	10.0	1.9	0.16	0.144
Wrote a letter, poem, or story	4.2	4.4	-0.2	-0.03	0.735
Played math games or did math problems	7.7	5.6	2.1 **	0.25	0.028
Total number of times that student engaged in typical activities during the summer	44.5	47.1	-2.7	-0.11	0.251
Went to a library	3.6	3.6	-0.1	-0.01	0.937
Read a book during free time	6.5	6.0	0.5	0.06	0.573
Watched TV during the day on a weekday	15.9	17.0	-1.1	-0.10	0.310
Did activities at a club, community center, church, or day camp	8.9	9.6	-0.8	-0.07	0.433
Played in a sports program	9.6	10.9	-1.3	-0.11	0.222
Sample size (N = 444)	243	201			

(continued)

Table 2.9 (continued)

SOURCE: MDRC calculations based on the BELL student survey administered in fall 2012.

NOTES: The analyses reported in this table are based on the sample of students who took the GRADE and GMADE assessments and who responded to the student survey in fall 2012 (Fall 2012 Analysis Sample). The estimated differences between the BELL group and the non-BELL group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012, as well as random differences between the BELL and non-BELL groups with respect to the following variables: a student's score on state reading and math tests taken in spring 2012, whether a student has an individualized education plan (IEP), whether the student has English as a Second Language (ESL), whether a student is eligible for free or reduced price lunch, parent education, race/ethnicity, and gender. The values in the column labeled "BELL Group" are the observed means for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-BELL group, using the observed mean covariate values for the BELL group as the basis for the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

Effect sizes are calculated by dividing the estimated difference by the standard deviation of the summer activity measure for students in the Fall 2012 Analysis Sample who are in the non-BELL group.

A two-tailed t-test was used to test differences between the BELL and non-BELL groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]Students reported the frequency of each summer activity on a 5-point scale, which was converted to number of times per summer as follows: "Never" = 0; "Hardly ever (1 or 2 times)" = 1.5; "Not very often (once a month)" = 3; "Sometimes (about once a week)" = 13; "Pretty often (a couple times or more a week)" = 26.

Chapter 3

Impacts on Academic Achievement and Engagement

Chapter 3 examines whether the Building Educated Leaders for Life (BELL) middle school summer program had a positive impact on students' academic outcomes in fall 2012, after they had participated in the program. As explained in Chapter 1 (Box 1.2), random assignment was used to decide which eligible students would be invited to participate in BELL. Thus, the impact of the program can be estimated by comparing the outcomes of students who were invited to participate in BELL (the BELL group) and the outcomes of students who were assigned to remain in "business as usual" summer activities (the non-BELL group). Because this evaluation is underpowered, it lacks the ability to statistically detect effects of the size seen in prior evaluations of summer programs. This means that effects on academic achievement would have to be very large (equivalent to about 14 to 17 weeks of regular schooling) to conclude that they are not due simply to chance.[1] However, the impact estimates themselves are still rigorous and unbiased; thus, the results can be used to identify promising or preliminary patterns of effects to inform the field of summer learning. The key findings from the impact evaluation follow.

- **What was BELL's impact on middle school students' reading achievement when they returned to school in the fall?** In the average study district in fall 2012, BELL students did not have higher reading test scores than non-BELL students (effect size = 0.01; p-value = 0.929). These results are consistent across reading subtests. In one of the three study districts, the effect on reading scores is negative and statistically significant. Thus, this study provides no evidence that BELL had a positive impact on students' reading achievement in the fall after program participation.

- **What was BELL's impact on middle school students' math achievement when they returned to school in the fall?** In the average study district in fall 2012, BELL students outperformed non-BELL students in math by an effect size of 0.07, which is equivalent to a little over one month of additional learning and is the amount by which students are expected to grow during a five-week period during the regular school year. The magnitude of this effect is also similar in size to what has been found in prior evaluations of voluntary summer programs at the elementary school level. On the one hand,

[1]The minimum detectable effect sizes for the study are 0.15 for reading scores and 0.17 for math scores, which is equivalent to the effect of 17 weeks of regular schooling in reading and 14 weeks of regular schooling in math (based on middle school benchmarks in Hill, Bloom, Black, and Lipsey, 2007). See Appendix A for further discussion of the MDES.

this difference is not statistically significant, which means that this result could simply be due to chance rather than to the effect of BELL. On the other hand, some of the study's ancillary findings support the hypothesis that BELL had a small but positive effect on math achievement. For instance, in one of the study districts, BELL had a statistically positive impact on students' math scores in one subdomain. BELL also had a statistically significant effect on students' participation in math-related activities during the summer, which is an important precursor to impacts on math achievement.

- **What was BELL's impact on middle school students' emotional and behavioral engagement when they returned to school in the fall?** In the average study district in fall 2012, BELL students appear to have been no more (or no less) engaged than non-BELL students when they returned to school (effect size = –0.01; p-value = 0.927). Thus, despite having attended an academically focused program for five weeks during the summer, the BELL group did not "burn out" and return to school with less motivation to learn.

In general, these findings provide preliminary evidence that the BELL middle school program did not have an impact on students' reading skills but that it may have had a positive effect on students' math skills.

The remainder of this chapter provides more detailed findings. The next section examines BELL's impact on students' reading and math achievement in the fall, after the summer program. Then the chapter looks at BELL's effect on students' engagement when they returned to school in the fall. The final section examines BELL's effect on academic achievement in each of the three study districts.

As already noted, the purpose of this study is to examine the effect of the average BELL program. For this reason, the three study districts (Districts A, B, and C) are weighted equally in the pooled analyses reported in this chapter. Thus, the pooled impact findings represent the effect of BELL for the average study district or for the average BELL program.[2]

Impacts on Academic Achievement in the Fall

Student achievement in math and reading in fall 2012 was measured using the Group Reading Assessment and Diagnostic Examination (GRADE) and the Group Mathematics Assessment and Diagnostic Examination (GMADE) assessments. Table 3.1 presents the estimated impact

[2]To obtain the pooled impact estimates, the impact of BELL is estimated for each district separately, and these estimates are then averaged together. Appendix A provides details about the statistical model.

The Evaluation of Building Educated Leaders for Life (BELL)

Table 3.1

Impacts on Academic Achievement in the Fall:
Fall 2012 Analysis Sample

Outcome	BELL Group	Non-BELL Group	Estimated Impact	Effect Size	P-Value for Estimated Impact
Reading achievement (standard score)[a]	91.6	91.5	0.1	0.01	0.929
Corresponding grade equivalent	5.2	5.2			
Corresponding percentile	32	32			
Corresponding normal curve equivalent (NCE)	38	38			
Reading comprehension (standard score)	90.3	90.1	0.2	0.01	0.822
Reading vocabulary (standard score)	94.7	94.9	-0.2	-0.01	0.843
Math achievement (standard score)[a]	87.6	86.6	0.9	0.07	0.286
Corresponding grade equivalent	5.1	4.9			
Corresponding percentile	27	25			
Corresponding normal curve equivalent (NCE)	33	32			
Math concepts (standard score)	88.3	88.6	-0.2	-0.02	0.814
Math operations (standard score)	89.3	87.6	1.7 *	0.12	0.094
Math processes (standard score)	85.3	84.1	1.3	0.10	0.233
Sample size (N = 919)	585	334			

SOURCES: MDRC calculations based on the GRADE and GMADE assessments administered in fall 2012.

NOTES: The analyses reported in this table are based on the sample of students who took the GRADE and GMADE assessments and who responded to the student survey in fall 2012 (Fall 2012 Analysis Sample). Estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012, as well as random differences between the BELL and non-BELL groups with respect to the following variables: a student's score on state reading and math tests taken in spring 2012, whether a student has an individualized education plan (IEP), whether the student has English as a Second Language (ESL), whether a student is eligible for free or reduced-price lunch, parent education, race/ethnicity, and gender. The values in the column labeled "BELL Group" are the observed means for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-BELL group, using the observed mean covariate values for the BELL group as the basis for the adjustment. Each of the three study districts is given an equal weight when estimating the results reported in this table. Rounding may cause slight discrepancies in calculating sums and differences.

Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome measure for students in the Fall 2012 Analysis Sample who are in the non-BELL group.

A two-tailed t-test was applied to differences between BELL and non-BELL groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]Students enrolled in fifth grade in spring 2012 were given Level 5 of the GRADE and GMADE; students in sixth grade were given Level 6; and students in seventh grade were given Level M. The national average for GRADE and GMADE standard scores is 100, and the standard deviation is 15. No statistical tests or arithmetic operations were performed on grade equivalents and percentiles because these are not equal-interval scales of measurement.

of BELL on these two assessments. Recall that BELL's effects on *total* scores are the primary indicators of program effectiveness in this study. Impacts on subtest scores are secondary and are presented only for the purpose of contextualizing the primary findings.

Because impacts on scaled scores are difficult to interpret, Table 3.1 also presents the impact estimates as effect sizes. The "effect size" is a metric that is widely used for gauging whether the magnitude of a program's impact is large or small. Conventional guidelines suggest that an effect size of about 0.20 is "small"; 0.50 is "medium"; and 0.80 is "large."[3] These guidelines are crude, however, because they cover many types of social interventions and programs. More recently in education research, it has become standard practice to use effect-size benchmarks that take into account the length of a program and the population of students that it serves. In this study, for example, the most relevant benchmark is the amount by which middle school students are expected to grow during a five-week period during the school year, which is an effect size of 0.04 in reading and 0.06 in math.[4] Thus, these are the benchmarks that should be used in interpreting the magnitude of the findings in this chapter. (For an explanation of how effect sizes are calculated, see Box 1.3 in Chapter 1.)[5]

- **In the average study district, BELL students and non-BELL students performed at a very similar level in reading at the beginning of fall 2012.**

As shown in Table 3.1, both BELL and non-BELL students were still performing below grade level in reading in the fall — at a fifth-grade level, on average. This is lower than would be expected for these students, because they were enrolled in the sixth grade or higher. Relative to a national sample of students, students in the study scored at the 32nd percentile in reading in the fall.

Furthermore, students in both groups had very similar reading scores. In the average study district, students in the BELL group had an average scaled score of 91.6 on the GRADE, while students in the non-BELL group had an average score that was only slightly lower: 91.5. The difference in reading scores between BELL and non-BELL students is not statistically significant; therefore, it cannot be concluded that BELL students outperformed non-BELL students in reading. Moreover, the magnitude of the impact is small, representing an effect size of 0.01, or about one week of additional learning, which is less than one would expect from a five-week program.[6] Estimated impacts on the two reading subtests — reading comprehension and

[3]Cohen (1988).

[4]This is based on annual growth estimates from Hill, Bloom, Black, and Lipsey (2007) for students enrolled in grades 5 through 7, which are the grade levels from which students in the study sample were rising. During the regular 36-week school year, the achievement of students in these grade levels is expected to grow by an effect size of about 0.32 in reading and 0.42 in math. Thus, students grow by 0.009 per week in reading and 0.012 per week in math; for a five-week period, they should grow by 0.04 in reading and 0.06 in math.

[5]The standard deviations used to calculate effect sizes can be found in Appendix A. Standard errors for the main impact estimates can be found in Appendix E.

[6]In this chapter, effect sizes for reading and math scores are converted to weeks of learning based on benchmarks for middle school students from Hill, Bloom, Black, and Lipsey (2007).

reading vocabulary — are similar in magnitude; the effect sizes are 0.01 and –0.01, respectively, and neither estimate is statistically significant.

- **In the average study district in fall 2012, BELL students outperformed non-BELL students in math by the equivalent of one month of learning (effect size = 0.07), which is what one would expect from a five-week program. However, this effect is not statistically significant.**

In math, both BELL and non-BELL students were still performing below grade level in the fall — at a fifth-grade level, on average, or the 26th percentile nationally. However, students in the BELL group had slightly higher math scores. They had an average scaled score of 87.6 on the GMADE at the beginning of the school year, while students in the non-BELL group had a score of 86.6. Thus, in the average study district, BELL students outperformed non-BELL students by 1.0 scaled score point, which corresponds to an effect size of 0.07. This is approximately equivalent to a little over one month's worth of academic learning (six weeks).[7] On the math subtests, BELL students outperformed non-BELL students in operations and processes but not on math concepts; the effect sizes are 0.12, 0.10, and –0.02, respectively.

Although none of the differences in math between BELL and non-BELL students is statistically significant, these preliminary findings are promising from a programmatic perspective. As noted above, if the BELL program is as effective as regular schooling, then a five-week academically oriented program like BELL should help middle school students grow by an effect size of 0.06 in math, which is about the effect size that was found in this study (0.07). Also of note, the magnitude of BELL's effect on middle school students' math achievement — a little over one month's worth of learning — is similar in size to the impact of previously evaluated voluntary summer programs for elementary school students, including BELL's elementary school program. (See Chapter 1.)

Impacts on Student Engagement in the Fall

Students' engagement in fall 2012 was measured using a set of items in the student survey, which are based on a scale developed by Ellen Skinner and colleagues.[8] The items in the scale

[7]This conversion to weeks of learning is based on Hill, Bloom, Black, and Lipsey (2007), rather than on the grade equivalent conversions for GMADE scores provided in Table 3.1. However, the two methods provide similar results. Based on the GMADE grade equivalents in Table 3.1, non-BELL students scored at a grade level of 4.9 in math (the performance of a student at the end of fourth grade) while BELL students scored at a grade level of 5.1 (the performance of a student after one month of fifth grade). Based on these numbers, the difference between the two groups is also about one month of instruction during the regular school year.

[8]Skinner, Furrer, Marchand, and Kindermann (2008).

asked students to rate their level of engagement in various types of classroom activities.[9] Responses to these items were then used to create a measure of students' overall engagement, as well as a measure of two specific facets of engagement: behavioral engagement and emotional engagement.[10]

- **In the average study district, BELL students appear to have been no more (or no less) engaged than non-BELL students when they returned to school in fall 2012.**

Table 3.2 presents the estimated effect of BELL on student engagement. BELL students reported being somewhat less engaged in their schoolwork than non-BELL students when they returned to school in the fall (effect size = –0.01), but the difference between the two groups is not statistically significant. Impacts on the two subscales of engagement — behavioral engagement and emotional engagement — are similar in magnitude; the effect sizes are –0.02 and –0.01, respectively, and neither effect is statistically significant.

On the one hand, these findings indicate that BELL did not meet its objective of increasing students' engagement in school. On the other hand, the results can also be viewed in a positive light: Despite having attended an academically focused program for several weeks during the summer, students in the BELL group did not "burn out" and return to school with less motivation to learn.

Impacts Analyzed by Study District

Tables 3.3 and 3.4 present the estimated impact of BELL on reading and math achievement in fall 2012 for each of the three districts in the study. The purpose of these analyses is to explore whether the effect of BELL on academic achievement is similar in magnitude across study districts or, conversely, whether there is observable variation in the size of program effects across districts. Because the study is underpowered — and because the sample size for some of these districts is small — district-specific effects must be very large to be statistically significant, especially in Districts A and B, which have smaller sample sizes. Similarly, *differences* in effects across districts must be even larger to be able to conclude that the impact in one district is statistically larger or smaller than for another district. (Indeed, statistical tests show that BELL's effect on reading and math scores do not vary by a statistically significant amount across the

[9]All items are measured on a 4-point scale: 1 = "Not at all true"; 2 = "Not very true"; 3 = "Sort of true"; and 4 = "Very true."

[10]Appendix B provides further information about the survey items included in the student engagement scales. The internal consistency reliability (Cronbach's alpha) of these measures ranges from 0.75 to 0.84.

The Evaluation of Building Educated Leaders for Life (BELL)

Table 3.2

Impacts on Student Engagement in the Fall: Fall 2012 Analysis Sample

Outcome	BELL Group	Non-BELL Group	Estimated Impact	Effect Size	P-Value for Estimated Impact
Student engagement (1-4)[a]	3.10	3.10	0.00	-0.01	0.927
Behavioral engagement	3.45	3.46	-0.01	-0.02	0.809
Emotional engagement	3.05	3.06	-0.01	-0.01	0.874
Sample size (N = 919)	585	334			

SOURCES: MDRC calculations based on the GRADE and GMADE assessments and the student survey administered in fall 2012.

NOTES: The analyses reported in this table are based on the sample of students who took the GRADE and GMADE assessments and who responded to the student survey in fall 2012 (Fall 2012 Analysis Sample). Estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012, as well as random differences between the BELL and non-BELL groups with respect to the following variables: a student's score on state reading and math tests taken in spring 2012, whether a student has an individualized education plan (IEP), whether the student has English as a Second Language (ESL), whether a student is eligible for free or reduced-price lunch, parent education, race/ethnicity, and gender. The values in the column labeled "BELL Group" are the observed means for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-BELL group, using the observed mean covariate values for the BELL group as the basis for the adjustment. Each of the three districts is given an equal weight when estimating the results reported in this table; therefore, means and estimated impacts are for the average district in the study sample. Rounding may cause slight discrepancies in calculating sums and differences.

Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome measure for students in the Fall 2012 Analysis Sample who are in the non-BELL group.

A two-tailed t-test was applied to the impact estimate. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]The student engagement scale and subscales are based on Skinner, Furrer, Marchand, and Kindermann (2008). A student's overall engagement score is based on his or her average response to 16 items with a 4-point truth scale (1 = "not at all true," 2 = "not very true," 3 = "sort of true," and 4 = "very true"). The behavioral and emotional engagement subscales are each based on a subset of five items.The internal consistency reliability (Cronbach's alpha) of these scales for the Fall 2012 Analysis Sample follow: student engagement = 0.84; behavioral engagement = 0.75; emotional engagement = 0.78.

study districts; for the results of these tests, see Tables 3.3 and 3.4.) Nonetheless, the district-specific effects of BELL reveal interesting patterns that can help one to better understand the pooled results.[11]

[11]As explained in Chapter 1, district-specific findings are considered secondary analyses in this report because they are conducted for hypothesis-generating purposes. Thus, statistically significant district-specific effects need not be adjusted for multiple hypothesis testing, based on standards provided by the What Works Clearinghouse (2014).

Table 3.3 presents district-specific impacts on reading achievement. As shown, the pattern of BELL's effects on reading scores varies across the three study districts, with the magnitude of effects being largest in District B and smallest in District C.[12] Though not statistically significant, BELL's effect on total reading scores in District B is numerically large and positive (effect size = 0.18), with similar effect sizes for the reading subtests. Conversely, in District C, the estimated effect of BELL on total reading scores is negative and statistically significant (effect size = –0.12; p-value = 0.070), which means that this result is not due to chance. In District C, effects on the two reading subtests are also negative in direction, and the effect for reading comprehension is statistically significant.

It is important to point out that the negative effect on reading scores in District C does not necessarily mean that BELL students in this district experienced a *decrease* (or summer loss) in their reading skills relative to non-BELL students. Rather, both groups of students may have made gains, but the reading gains of BELL students may have been smaller than the gains made by non-BELL students. It is not possible to determine whether this supposition of smaller gains is the correct one, because, for reasons explained in Chapter 1, the pretest-to-posttest change in students' achievement over the summer cannot be measured.

Table 3.4 presents district-specific impacts on math achievement. As shown, the magnitude and pattern of BELL's effect on math scores also appear to vary across districts. In District A, the effect on total math scores is slightly negative and is not statistically significant (effect size = –0.08). In contrast, in District B, BELL's effect on total math scores is numerically large and positive (effect size = 0.24). Although the estimated effect on total math scores in District B is not statistically significant, BELL did have a statistically significant impact on one particular type of math skill there: math operations (effect size = 0.37; p-value = 0.036). The effect size on this subtest corresponds to approximately seven months of learning, which is greater than the amount by which students are expected to grow during a five-week period during the regular school year. In District C — where impacts on *reading* scores are statistically negative — BELL's effect on math scores is positive in magnitude. The effect on total math scores in this district is 0.06, which is equivalent to about one month of additional learning and is the effect size that one would expect from a five-week program.

Given the apparent variation in the pattern of BELL's effect across districts, an important question is whether these district-specific impact findings are associated with the way in which each district implemented the BELL program (as reported in Chapter 2), and, by extension, whether any lessons can be gleaned as to which program features may matter most for the success of summer programs for middle school students.

[12]Because the study is underpowered, it cannot be concluded that the variation in effects across districts is statistically significant.

The Evaluation of Building Educated Leaders for Life (BELL)

Table 3.3

Impacts on Reading Achievement in the Fall, by District: Fall 2012 Analysis Sample

Outcome	Sample Size	BELL Group	Non-BELL Group	Estimated Impact	Effect Size	P-Value
District A						
Total reading (standard score)	358	93.1	93.6	-0.6	-0.04	0.604
Reading comprehension (standard score)	358	91.8	92.0	-0.2	-0.02	0.829
Reading vocabulary (standard score)	358	96.5	97.4	-0.9	-0.07	0.494
District B						
Total reading (standard score)	117	93.0	90.7	2.2	0.18	0.225
Reading comprehension (standard score)	117	91.6	89.2	2.4	0.19	0.231
Reading vocabulary (standard score)	117	95.7	94.1	1.6	0.12	0.471
District C						
Total reading (standard score)	444	88.7	90.2	-1.5 *	-0.12	0.070
Reading comprehension (standard score)	444	87.7	89.3	-1.6 *	-0.13	0.073
Reading vocabulary (standard score)	444	92.0	93.2	-1.2	-0.10	0.198
Test of variation in impacts across districts						
Total reading (standard score)						0.177
Reading comprehension (standard score)						0.169
Reading vocabulary (standard score)						0.497

SOURCES: MDRC calculations based on the GRADE and GMADE assessments administered in fall 2012.

NOTES: The analyses reported in this table are based on the sample of students who took the GRADE and GMADE assessments and who responded to the student survey in fall 2012 (Fall 2012 Analysis Sample). Estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012, as well as random differences between the BELL and non-BELL groups with respect to the following variables: a student's score on state reading and math tests taken in spring 2012, whether a student has an individualized education plan (IEP), whether the student has English as a Second Language (ESL), whether a student is eligible for free or reduced-price lunch, parent education, race/ethnicity, and gender. The values in the column labeled "BELL Group" are the observed means for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-BELL group, using the observed mean covariate values for the BELL group as the basis for the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome measure for students in the Fall 2012 Analysis Sample who are in the non-BELL group.

A two-tailed t-test was applied to differences between BELL and non-BELL groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]Students enrolled in fifth grade in spring 2012 were given Level 5 of the GRADE and GMADE; students in sixth grade were given Level 6; and students in seventh grade were given Level M. The national average for GRADE and GMADE standard scores is 100, and the standard deviation is 15. No statistical tests or arithmetic operations were performed on grade equivalents and percentiles because these are not equal-interval scales of measurement.

The Evaluation of Building Educated Leaders for Life (BELL)

Table 3.4

Impacts on Math Achievement in the Fall, by District: Fall 2012 Analysis Sample

Outcome	Sample Size	BELL Group	Non-BELL Group	Estimated Impact	Effect Size	P-Value
District A						
Total math (standard score)	358	90.7	91.7	-1.0	-0.08	0.409
Math concepts (standard score)	358	88.9	91.1	-2.2	-0.16	0.111
Math operations (standard score)	358	94.0	94.6	-0.6	-0.04	0.688
Math processes (standard score)	358	87.7	88.1	-0.4	-0.03	0.802
District B						
Total math (standard score)	117	85.5	82.4	3.1	0.24	0.151
Math concepts (standard score)	117	88.7	88.3	0.4	0.03	0.857
Math operations (standard score)	117	87.7	82.5	5.2 **	0.37	0.036
Math processes (standard score)	117	84.1	80.9	3.2	0.25	0.208
District C						
Total math (standard score)	444	86.6	85.8	0.8	0.06	0.417
Math concepts (standard score)	444	87.4	86.4	1.1	0.08	0.314
Math operations (standard score)	444	86.3	85.8	0.5	0.04	0.653
Math processes (standard score)	444	84.1	83.2	0.9	0.07	0.416
Test of variation in impacts across districts						
Total math (standard score)						0.220
Math concepts (standard score)						0.168
Math operations (standard score)						0.126
Math processes (standard score)						0.464

SOURCES: MDRC calculations based on the GRADE and GMADE assessments administered in fall 2012.

NOTES: The analyses reported in this table are based on the sample of students who took the GRADE and GMADE assessments and who responded to the student survey in fall 2012 (Fall 2012 Analysis Sample). Estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012, as well as random differences between the BELL and non-BELL groups with respect to the following variables: a student's score on state reading and math tests taken in spring 2012, whether a student has an individualized education plan (IEP), whether the student has English as a Second Language (ESL), whether a student is eligible for free or reduced-price lunch, parent education, race/ethnicity, and gender. The values in the column labeled "BELL Group" are the observed means for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-BELL group, using the observed mean covariate values for the BELL group as the basis for the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome measure for students in the Fall 2012 Analysis Sample who are in the non-BELL group.

A two-tailed t-test was applied to differences between BELL and non-BELL groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]Students enrolled in fifth grade in spring 2012 were given Level 5 of the GRADE and GMADE; students in sixth grade were given Level 6; and students in seventh grade were given Level M. The national average for GRADE and GMADE standard scores is 100, and the standard deviation is 15. No statistical tests or arithmetic operations were performed on grade equivalents and percentiles because these are not equal-interval scales of measurement.

In general, it appears that program effects are most strongly associated with the size of the service contrast between BELL and non-BELL students' summer academic activities. Effects on reading scores are largest in magnitude in District B, followed by District A and then District C (Table 3.3). The magnitude of the service contrast in students' writing activities (whether they wrote a letter, poem, or story) follows the same rank order (Chapter 2, Tables 2.7 to 2.9). The same is true for the math findings. The largest effects in math are in District B, followed by District C and then District A (Table 3.4). The size of the service contrast in students' math-related activities follows the same rank order (Tables 2.7 to 2.9). Moreover, in Districts B and C — where effects on math scores are positive in direction — the service contrast in math activities is statistically significant.

Aside from the service contrast, no other implementation features appear to be associated with the district-level impact findings, and, in fact, the pattern of results is at times the opposite of what one would expect. In District B, where effects on math scores are the largest in magnitude, the program actually experienced the greatest delays in receiving curricular materials and diagnostic test scores, and attendance in the program was satisfactory (85 percent among students who attended at least one day of the program) but lower than in District A. Nor is there any clear explanation of why effects on reading scores are statistically negative in District C while effects on math scores are positive in direction. Of the three study districts, program attendance was the lowest in District C; in addition, only a small percentage of teachers in the district worked at the program school during the school year, which may have made it more difficult for teachers in District C to ensure continuity of instruction between the regular school year and the BELL program. Such explanations are not convincing, however, because these factors would explain small or null effects — not negative effects. Nor do these factors explain why effects for math and reading in District C are in different directions.

Summary

On the one hand, this study's findings offer preliminary evidence that BELL's middle school model may not have improved students' reading achievement in summer 2012. The effect of BELL on reading scores in the average study district is numerically zero, and the effect in one of the study districts is negative. Thus, it cannot be concluded that BELL's middle school program — as it was implemented in three school districts in summer 2012 — had an impact on the reading achievement of students in the BELL group compared with students in the non-BELL group.

On the other hand, there is suggestive preliminary evidence that BELL's middle school model may have had a positive effect on students' math achievement. BELL students outperformed non-BELL students by the equivalent of about one month of learning, which is the size of effect that one would expect from a five-week program during the regular school year. The

magnitude of this effect is also similar in size to findings from prior studies of voluntary summer programs for elementary school students. It is important to note that BELL's impact on math scores is not statistically significant, so the difference between BELL and non-BELL students in math could be due to chance rather than to the effect of BELL. Yet some of the ancillary findings from the study support the hypothesis that BELL had a small but positive effect on math achievement. First, BELL also had a statistically significant effect on students' participation in math-related activities during the summer, which is an important precursor to impacts on math achievement. (See Chapter 2.) Second, BELL appears to have had a positive effect on students' math scores in one subdomain in District B, which also happens to be the district where BELL had the largest effect on students' math-related activities.

Chapter 4

Conclusion

Middle school is a critical transition point for students, both academically and in terms of setting the foundations for the adults that they will become. Students who flounder in middle school rarely graduate from high school on time, and so finding ways to help them succeed is crucial.[1] Summer school — which provides students with more time and help with mastering difficult material — seems a logical aid. But getting middle school students to engage voluntarily in academic material in which they are weak — be it during the regular school day or during the summer — is a more difficult task than when students are younger.

The purpose of this study of the Building Educated Leaders for Life (BELL) middle school program is to help expand knowledge about the potential effects of voluntary full-day summer academic programs. As explained throughout this report, several limitations of the study make it inadvisable to use the findings to make definitive conclusions about the effect of academic summer programs for middle school students. Instead, the study's findings should be viewed as *preliminary.* Its results should be used to generate hypotheses about the potential for such programs to improve the academic achievement of this tough-to-reach age group and to better understand the context in which summer programs are implemented.

Chapter 4 summarizes the key findings about the impact and implementation of the BELL program and discusses their implications for similar academically focused middle school summer programs. The chapter also describes how the BELL middle school model has evolved and changed since summer 2012 and the next steps for the program. The chapter concludes by offering a set of lessons and recommendations for future studies of academic summer programs for middle school students.

Discussion of the Key Findings

The findings from this study provide suggestive preliminary evidence that BELL's middle school model may have had a positive effect on students' math achievement. Although not statistically significant, BELL's effect on middle school students' math scores is encouraging. At the beginning of fall 2012, BELL students outperformed non-BELL students in math by an amount equivalent to about one month of learning. The magnitude of this effect is similar in size to what has been found in prior evaluations of voluntary summer programs at the elementary school level. On the one hand, an important caveat is that the observed difference in math

[1]Balfanz (2009); Balfanz, Herzog, and Mac Iver (2007).

scores between BELL and non-BELL students is not statistically significant and may simply be due to chance. On the other hand, the fact that BELL had a statistically significant impact on students' math-related activities during the summer — and that impacts on math scores in one subdomain were positive and are statistically significant in one of the three study districts — lends support to the hypothesis that BELL had a small but positive effect on math achievement.

This study also provides preliminary evidence that BELL's middle school model may not have improved students' reading achievement. Both BELL and non-BELL students had very similar average reading scores when they returned to school in fall 2012, and the impact on reading scores is negative in one study district. In addition, the program does not appear to have had an impact on BELL students' writing activities during the summer, compared with those of non-BELL students.

Thus, the big question is: Why did BELL not affect students' reading achievement and writing behaviors when it does seem to have had an impact on their math achievement and math-related activities? One hypothesis for this pattern of results is that students were more engaged in the math curriculum or, conversely, that the BELL reading curriculum may not have been effective with the average student in the study, who is below grade level in reading. Another hypothesis is that non-BELL students were able to maintain their reading and writing skills through a more informal route. Perhaps in era of smartphones and the Internet, students are consuming a larger amount of written material outside school, which would reduce the potential benefits of a short academic summer program on reading achievement.

Another noteworthy finding is that, despite attending an academic summer program, BELL students were no less engaged in school than non-BELL students when they returned to school in fall 2012. Some fear that sending their children to a summer program will fatigue them and make them less engaged academically when they return to school, but this was not the case with the BELL program.

Lessons for Implementing Academic Summer School Programs

In 2009, the Institute of Education Sciences (IES) at the U.S. Department of Education commissioned an exhaustive review of the evidence on the effectiveness of academic summer and afterschool programs. Although there is little rigorous research on this topic for middle school students as an age group, this IES review did uncover several promising practices for middle school programs that aim to improve students' academic outcomes.[2] But some of these practices can be difficult to implement. Below are reflections on the challenges involved, based on the implementation findings from this study (Chapter 2).

[2]Beckett et al. (2009).

- **Teachers and mentors (teaching assistants) should be trained in and familiar with the summer program's curriculum, and materials should be in place on Day One.**

A particular strength of the BELL program is that it hires experienced certified teachers. During focus groups, however, teachers provided feedback on possible areas of improvement with respect to the staff training. First, teachers noted that they would have preferred to spend more time familiarizing themselves with the BELL curriculum for their subject area, rather than spending time on pedagogy or instructional practices. Second, some teachers also noted that BELL's training could focus more on helping academic teachers work effectively with the mentors. Some teachers successfully used their mentors to assist with lessons and support student learning, while other teachers struggled with the collaboration.

In terms of program start-up, a five-week program is very short and leaves little time for getting organized. While this sounds easy to fix, the exact number of students is often uncertain until shortly before they program starts, so teachers are sometimes hired and materials are ordered within days of beginning the program. As happened in this study, there can be unexpected delays in getting materials to program sites on time. Thus, summer programs should make a concerted effort to be ready to start on Day One of the program and to have contingency plans in place in the event that distribution problems or other delays arise.

- **The curriculum has to be relevant, interactive, and hands-on so that middle school students will stay engaged.**

Academic motivation decreases steadily from early elementary school into high school, and adolescents' desire for autonomy makes it is harder to get older students to engage in learning.[3] Middle school students tend to engage more actively in material that they consider "relevant." For instance, the IES review of summer and after-school programs recommends that successful academic programs should "be interactive, hands on, learner directed and related to the real world while remaining grounded in academic learning goals," suggesting that activities capitalize on students' interests.[4]

This report's findings from the BELL evaluation suggest that it may be more difficult for academic summer programs to engage middle school students in reading than in math (Chapter 3). Engaging them in reading and writing instruction may require even more concerted and hands-on approaches.

- **Instruction should be adapted to individual and small-group needs.**

[3]National Research Council and Institute of Medicine (2004), p. 2.
[4]Beckett et al. (2009), p. 29.

This IES recommendation is the one that is best supported by research. Given that summer programs only have a short period of time to improve academic achievement, teachers need to target instruction to the skills that a student has not yet mastered. Individualized instruction is even more important in shorter programs than during the regular school day.

BELL knows that providing teachers with student diagnostic data is critical for individualizing instruction, which is why students are tested at the start of the program. As noted, however, it is sometimes operationally difficult to provide timely diagnostic information. Some teachers who were interviewed also felt that testing students during the first few days of the program took away from precious teaching and learning time in an already-short program. One possible solution is to test students during the application process or in the week before the program starts. Alternatively, students' teachers in the regular school year might be able to provide the summer instructors with critical information about those students' specific academic needs.

- **The summer curriculum should be aligned with the approaches and level of instruction that are used with students during the regular school year.**

The IES review suggests that the effectiveness of a summer program might be improved if the curriculum is aligned with the curriculum that is used during the regular school year. In particular, the review recommends that program coordinators talk with staff about how summer activities can best connect to (and align with) school-based learning objectives. Such conversations can also be useful for understanding the needs of particular students. Hiring summer instructors who teach at the school during the regular school year might also improve instructional alignment.

The Evolution of BELL's Middle School Model and Next Steps

As a continuous learning organization, BELL's efforts to refine and strengthen its middle school model have been ongoing since summer 2012. Even before the findings from this study were known, BELL had started implementing the following modifications to its model, with the goal of improving instructional quality:

- **Online teacher training and staff resources.** BELL has continued to improve and add new content to BELL University, its internal online e-learning platform, and has developed a digital library in BELL University that stores important resources for program and regional leaders.
- **Decentralization of staff training.** In summer 2012, BELL's training was highly centralized: a training team from BELL's national headquarters would visit the regions and offer a standardized training to staff. More recently,

however, BELL has implemented a "train-the-trainer" model that is less centralized and more regionally driven by individual BELL partners. Each of BELL's regional offices sends a designated representative to national headquarters to undergo a train-the-trainer boot camp. The regional representatives then take the information learned and related materials back to their regional offices, which develop and coordinate their own training for program leaders (program managers and lead teachers). In turn, program leaders then coordinate their own school-specific training for teachers and mentors at the individual sites. This approach maintains fidelity to the BELL model while also allowing regions to accommodate local training needs.

- **Curricula aligned with Common Core State Standards (CCSS).** During core academic instruction time, BELL is now using English Language Arts (ELA) and math curricula published by Pearson: *Reader's Journey* and *Math Navigator*. Both curricula are fully aligned with the Common Core State Standards. The ELA curriculum has an increased focus on nonfiction (informational) texts, and the math curriculum has a strong focus on algebraic reasoning. These new curricula were chosen because they can be customized to a five-week program; they are structured in a way that provides teachers with opportunities to individualize instruction (through one-on-one and small-group activities); they include hands-on project-based activities that are engaging to middle school students; and they include an optional blended learning (online) module that teachers can use if they have access to computers.

- **Distribution of curricular materials.** The instructional materials for *Reader's Journey* and *Math Navigator* are fully consumable, meaning that students can keep their workbooks and other materials after completing the program. This has eliminated the need to store textbooks and other materials in a central BELL warehouse during the school year and having to return them to sites at the start of the following summer. Instead, curricular materials are shipped directly to sites from Pearson at the start of each summer, resulting in more timely arrival of key material resources.

- **Student assessment.** BELL is using a new, computer-adaptive assessment tool to measure student achievement at both the beginning and the end of the program. This assessment is more rigorous than the previously used instrument, is aligned with the CCSS, and provides immediate data on student performance. Of note, the new assessment tool can tell teachers in which specific subdomains of reading and math their students are most deficient, and it can also group students based on their level of need. Because the assessment

tool is computer based (rather than paper based), teachers have access to their students' test scores sooner.

- **The role of the lead teacher.** The responsibilities of the lead teacher are now those of a full-time "instructional coach." Instructional coaches are expected to observe classrooms each week, to provide advice to teachers on how to improve instruction and engage students, to give teachers feedback on their weekly lesson plans, and more generally to support teachers in implementing the new curricula. Instructional coaches are also tasked with helping teachers make more efficient use of mentors (teaching assistants) for the purposes of delivering and individualizing ELA and math instruction.

- **Quality assurance.** BELL has implemented a new quality-assurance tool that is more focused on measuring classroom instructional quality and rigor and is less focused on measuring compliance to the structural aspects of the program model.

These programmatic enhancements are in line with the best practices recommended by IES and are a positive step toward strengthening BELL's middle school model. For instance, the decentralization of BELL's training may lead to better alignment between teacher training and the qualifications and needs of the teaching staff in each district. BELL's new ELA and math curricula — because they are aligned with the CCSS — will likely be better aligned with what students are taught during the school year. In addition, the new curricula may also be more engaging to students, because the content is more rigorous and includes hands-on activities. Finally, BELL's new distribution process will help to ensure that there are fewer delays in getting program materials to the sites, thereby facilitating strong program start-up.

In the coming summers, BELL will continue to strengthen and refine its middle school model so that it can improve the academic trajectory of children in communities with resource challenges. With the help of its steadfast funders, the organization has embarked on a multiyear process to look for ways to better engage and teach struggling middle school students. As part of this process, BELL has created a Middle School Advisory Board whose membership includes researchers and practitioners with expertise in middle school interventions and summer programs, who will advise BELL on best practices for teaching middle school students. Being a data-driven organization, BELL plans to implement further modifications to its program, based on the board's recommendations, and to assess whether these modifications have the potential to improve student outcomes. Given that there are so few examples of effective models for middle school summer programs, these improvements to the BELL model — and the evaluation of their implementation and effects — will be of interest to the whole field.

Lessons for Future Studies of Academic Summer Programs

Reflecting on the challenges encountered in this research study, lessons like the following can be drawn that might inform and improve the design of future evaluations of academic summer programs for middle school students.

- **Experienced sites.** Future evaluations should be conducted, if possible, in schools that have prior experience with the summer program. Among other benefits, this will ensure that the programs being evaluated have an existing pool of summer teaching staff to draw from and that they already have a distribution process in place for getting program materials to schools. This, in turn, will ensure that the programs can get going more quickly at start-up, thus providing a stronger test of their effects.

- **Number of study participants.** Future evaluations should be powered to detect effects on academic achievement of about 0.04 to 0.06 standard deviation, which are the gains in reading and math test scores that one would expect from a five-week summer program. Recruiting a sample of students that is large enough to detect effects of this magnitude is challenging, because random assignment can be used only in programs and at grade levels when participation is voluntary and oversubscribed. Thus, to build a large enough sample, it might be necessary to recruit students across two or more summers (multiple cohorts).

- **Administering pretests.** For two reasons, future studies of summer programs should consider administering parallel academic assessments at the beginning of the program (pretest) and at the end of the program or in the fall (posttest). First, this would make it possible to measure summer loss, which is an important piece of contextual information for interpreting program effects. In the present study, for example, BELL and non-BELL students performed at similar levels in reading in the fall, but it is not possible to determine whether both groups gained skills over the summer or whether both groups lost skills. Knowing whether students gained or lost skills would be useful for identifying specific areas for program improvement. The second reason for administering a pretest is that using pretest scores as a baseline covariate in the analysis could potentially improve the precision of estimated program impacts.[5]

[5]This is because, presumably, pretests would be more highly correlated with the posttests than spring state test scores (the baseline achievement measure used in the present study).

- **"Business as usual" academic summer activities.** Future studies should collect detailed information from control group students about their formal and informal summer academic activities (such as time spent reading and which types of materials, time spent writing emails and texting, and so on). In the present study, for example, students' summer reading activities were not measured at the level of detail needed to fully understand why non-BELL students' reading scores were at the same level in the fall as the scores of non-BELL students.

- **Effect of multiple summers.** Future studies of summer programs should look at the effect of participating in the program for consecutive summers and should follow students for a longer time frame. For example, Higher Achievement — an intensive year-round academic intervention that includes after-school programming during the school year as well as academic instruction during the summer — had no impact on students' test scores after their first year in the program (effect size = 0.02 to 0.03), but it had effects after two years of participation (effect size = 0.08 to 0.10).[6]

Final Thoughts

BELL is a strong, data-driven organization that is determined to help less advantaged youth succeed. It has shown that it can operate an effective summer program for elementary school students and that students who participate in the program outperform similar elementary school students by a little more than an extra month of schooling.[7] By participating in this randomized controlled trial, BELL, as a learning organization, undertook to rigorously investigate whether it had effectively translated its elementary school model into a middle school model. It learned that it was able to attract and retain struggling middle school students to a fairly well-run academic summer program and that the program may have positively affected middle school students' math scores but not their reading scores. BELL is now using what it has learned as input in its continuous efforts to improve its middle school model.

[6]Herrera, Grossman, and Linden (2013).
[7]Chaplin and Capizanno (2006).

Appendix A

The Statistical Model and Statistical Power of the Evaluation

Appendix A discusses various technical issues related to the estimation of program impacts. The first section discusses the statistical model used to estimate the impact of the Building Educated Leaders for Life (BELL) summer academic program for middle school students. The second section discusses the minimum detectable effect size (MDES) for the main impact findings in the study. The final section includes tables of the standard deviations used to calculate effect sizes in this report. (For an explanation of how effect sizes are calculated, see Box 1.3 in Chapter 1.)

The Statistical Model for Estimating Impacts

The impact of the BELL middle school program on student outcomes is estimated by fitting regression model (1) to the Fall 2012 Analysis Sample:

$$Y_i = \beta_{DA} T_i * DA_i + \beta_{DB} T_i * DB_i + \beta_{DC} T_i * DC_i + \sum_K \lambda_k B_{ki} + \sum_S \delta_s X_{si} * DA_i + \sum_S \gamma_s X_{si} * DB_i + \sum_S \theta_s X_{si} * DC_i + \sum_S \omega_s M_{si} * DA_i + \sum_S \phi_s M_{si} * DB_i + \sum_S \psi_s M_{si} * DC_i + \varepsilon_i \quad (1)$$

Where:

Y_i = Outcome of interest for student i (test scores or student engagement).

T_i = Indicator for BELL group membership (treatment status). This indicator is equal to 1 if student i was assigned to the BELL program and zero if student i was assigned to the non-BELL group.

DA_i = Indicator for District A, equal to 1 for students in District A and 0 for other students.

DB_i = Indicator for District B, equal to 1 for students in District B and 0 for other students.

DC_i = Indicator for District C, equal to 1 for students in District C and 0 for other students.

Bki = Set of K random assignment block indicators, equal to 1 if student i is in random assignment block k and zero otherwise. These blocks are included in model (1) to capture a central feature of the research design in which random assignment was conducted separately for each grade level and by the school

that students attended in spring 2012.[1] Controlling for random assignment blocks in the model also accounts for the clustering of student outcomes by school and grade level, because it explains all the between-school and between-grade variation in student outcomes.[2]

Xsi = Set of S baseline characteristics for student i.[3] To obtain unbiased impact estimates, it is not necessary to control for students' baseline characteristics, because random assignment should ensure that the program and control groups have similar observed and unobserved characteristics at baseline. However, controlling for student characteristics can increase the precision of the impact estimates, because these characteristics explain part of the within-block variation in the outcome measure. Controlling for student characteristics can also be used as a "safeguard" to ensure that the program and control groups are comparable on all characteristics.[4] In model (1), note that the student characteristics are interacted with indicators of school district (DA, DB, DC) to allow their effect to vary across districts.

[1]There are 44 random assignment blocks in the full study sample and 43 in the Fall 2012 Analysis Sample used to estimate program impacts. (Appendix C discusses this sample.) One block is excluded from the analysis because it included either only BELL or only non-BELL students. These blocks represent different combinations of students' grade level and their school in spring 2012. It is important to note that the blocks are defined based on students' school *in the 2011-2012 school year,* not the school where the summer program was held (each of which serves students from many feeder schools). This was done to ensure that the BELL and non-BELL groups were similar before entering the program in terms of the distribution of schools that they attended during the school year.

[2]The random assignment ratio in the Fall 2012 Analysis Sample differs across blocks (minimum = 0.15; maximum = 0.88; median = 0.58). These differences in the random assignment ratio must be accounted for to obtain an unbiased estimate of impacts. There are several ways to account for variation in the random assignment ratio. The two most common methods are (1) to "block-mean" center the covariates on the right-hand side of the model and (2) to include block fixed-effects in the model. Raudenbush (2009) shows that these two methods produce the same impact estimate. Model 1 is based on the latter approach.

[3]The following covariates are included in Model 1: students' scores on state reading and math tests taken in spring 2012 (interacted with indicators of district and grade level, to account for differences in the scales of the tests across districts and grades), whether a student has an individualized education plan (IEP), whether the student has English as a Second Language (ESL), whether a student is eligible for free or reduced-price lunch, parents' education, race/ethnicity, and gender. These covariates were chosen because they are strong predictors of academic achievement; the decision about which covariates to include in the model was made before starting the impact analysis.

[4]Specifically, when differences between the BELL group and the non-BELL group are between 0.05 and 0.25 standard deviation (as in this study; see Appendix C), the What Works Clearinghouse recommends that these characteristics be included as covariates in the impact model (What Works Clearinghouse, 2014).

Mi = A set of S missing indicators for each of the student characteristics, coded 1 if missing and 0 otherwise.[5] The effect of these indicators is also allowed to vary across district.

ε_i = A within-student error term.

Therefore:

β_{DA} = The estimated impact of BELL on outcome Y in District A.

β_{DB} = The estimated impact of BELL on outcome Y in District B.

β_{DC} = The estimated impact of BELL on outcome Y in District C.

The average impact of BELL on outcome Y (β_{ALL}) is then estimated by taking a linear combination of the three district-specific impact estimates, as follows:

$$\beta_{ALL} = \frac{(\beta_{DA} + \beta_{DB} + \beta_{DC})}{3}$$

The standard error (s.e.) for the average impact is:

$$s.e.(\beta_{ALL}) = \frac{\sqrt{var(\beta_{DA}) + var(\beta_{DB}) + var(\beta_{DC})}}{3}$$

In this way, each of the districts is weighted equally in the overall impact findings, which means that the overall impact estimates in the report represent the effect of BELL for the average district in the study. As explained in the report, districts are weighted equally because the effect of BELL differs numerically across the study districts and because the research sample in District C is much larger than in Districts A and B.

The statistical significance of impact estimates (and other estimates) in this report is assessed using a two-tailed t-test. *Statistical significance* is a measure of the degree of certainty that one may have that a program's impact is actually nonzero. If an impact estimate is statistically significant, then one may conclude with some confidence that the program really had an effect on the outcome being assessed. If an impact estimate is not statistically significant, then the nonzero estimate is more likely to be a product of chance. In this report, statistical significance is indicated by asterisks (*) when the p-value is less than or equal to 10 percent.

[5]Missing information on each student characteristic X was imputed using a *dummy variable approach,* which consists of (1) imputing a value of zero for missing values in each covariate, (2) creating a dichotomous indicator of missingness for each covariate, and (3) including these indicators alongside the imputed covariates in the statistical model (Puma, Olsen, Bell, and Price, 2009). In the Fall 2012 Analysis Sample, the percentage of missing data ranges from 1 percent (for free or reduced-price lunch status) to 10 percent (for ESL status).

Finally, it is important to note that the impact estimates presented in this report are "intent-to-treat" estimates of the effect of the BELL program. Some students assigned to BELL chose not to participate in the program. (Among the Fall 2012 Analysis Sample, 8 percent of students in the BELL group did not attend the program at all; see Chapter 2.) Thus, the findings in this report represent the estimated impact of offering students the opportunity to enroll in BELL (intent to treat), rather than the impact of BELL on students who actually enrolled (treatment on the treated). However, because students' participation in educational interventions is typically voluntary, intent-to-treat estimates of the impact of offering a program or service are also policy relevant.[6]

Minimum Detectable Effect Sizes

This section examines how large an impact BELL would have had to produce in order for the evaluation to be able to detect it. A common way to convey a study's statistical power is through the minimum detectable effect or the minimum detectable effect size. Formally, the *minimum detectable effect (MDE)* is the smallest true program impact that can be detected with a reasonable degree of power (in this case, 80 percent) for a given level of statistical significance (in this case, 10 percent for a two-tailed test). The *minimum detectable effect size (MDES)* is the minimum detectable effect scaled as an effect size; in other words, it is the MDE divided by the standard deviation of the outcome of interest. Effect sizes are used widely for measuring the impacts of educational programs and are defined in terms of the underlying population's standard deviation of student achievement. For example, an MDES of 0.20 indicates that an impact estimator can reliably detect a program-induced increase in student achievement that is equal to or greater than the 0.20 standard deviation of the existing student distribution.

The minimum detectable effect (MDE) and effect size (MDES) for a study are a function of the standard error (s.e.) of the estimated program impact:[7]

$$MDE = M_{N-B-X} * s.e.(\hat{\beta}) \tag{2a}$$

$$MDES = M_{N-B-X} * \frac{s.e.(\hat{\beta})}{\sigma} \tag{2b}$$

[6]The estimated effect of the treatment on the treated can be obtained by dividing the intent-to-treat impact estimates in this report by 90 percent, which is the difference between the percentage of students in the BELL group in the Fall 2012 Analysis Sample who actually attended the program (92 percent) and the percentage of non-BELL students in the sample who attended the program (2 percent). For a discussion, see Bloom (2006).

[7]This is because the standard error of the impact estimate is what determines whether the impact estimate is statistically significant.

Where:

$s.e.(\hat{\beta})$ = Standard error of the impact estimate.

σ = The standard deviation that is used to calculate effect sizes. (In this study, for example, it is the standard deviation of the non-BELL group.)

N = Number of students in the sample.

B = Number of random assignment blocks in the impact analysis.

X = Number of student baseline characteristics and missing data indicator variables included as covariates in Model 1. (See the preceding section.)

M_{N-B-X} = The "degrees of freedom" multiplier, which is calculated to be 2.5 in this study, assuming a two-tailed test with a statistical power level of 0.80 standard deviation and a statistical significance level of 10 percent.

However, during the study's design phase (during student recruitment), the standard error is not known and therefore must be estimated. The following equations can be used to approximate the minimum detectable effect:

$$MDE = M_{N-B-X}\sqrt{\frac{(1-R^2)\sigma^2}{P(1-P)N}} \quad (3a)$$

$$MDES = M_{N-B-X}\sqrt{\frac{(1-R^2)}{P(1-P)N}} \quad (3b)$$

Where:

P = Proportion of sample members assigned to the program group.

R^2 = Proportion of the variation in the outcome measure that is explained by the covariates and missing data indicator variables in Model 1.

As seen here, the sample size is a key determinant of the MDES. The *greater* the number of students in the study, the *smaller* the impact that can be detected. In practice, this means that there is a trade-off between the ability to detect effects and the cost of the study. On the one hand, a smaller MDES increases the likelihood that the study will be able to conclude that the program's effect is statistically significant. On the other hand, the smaller the MDES, the greater the number of students who need to be recruited into the study, which, in turn, increases the

cost and complexity of the evaluation. Therefore, evaluators must decide on a sample size that strikes the right balance between cost and the ability to detect effects.

An important step in this process is to figure out what size of effect the intervention is *expected* to have and to choose a sample size that will make it possible to detect that effect. For example, suppose that an intervention is expected to improve students' reading achievement by an effect size of 0.10; in that case, the most cost-efficient approach would be to choose a sample size that is just large enough to be able to detect an impact of that magnitude.

Researchers' "best guess" about the expected effect of an intervention typically comes from (1) prior evaluations of the intervention itself and/or (2) evaluations of programs that are similar to the intervention being studied. For the BELL evaluation, there are two sources of such information from which to draw. The first is BELL's earlier study of its elementary school program, which found an improvement of 0.08 in students' reading achievement.[8] The second is a meta-analysis of summer programs by Cooper and colleagues.[9] As discussed in Chapter 1, these evaluations of summer programs had found impacts similar in size to those for BELL's elementary school program (equivalent to about one month's worth of regular schooling).

After the spring 2012 recruiting period ended, BELL had recruited 1,226 students in the three study districts. It was determined that, with this sample size, the study would be able to detect an effect of 0.10 or larger.[10] This was considered to be an acceptable MDES, because it means that the study would be able to detect an effect of about the same size as the effect of other summer school programs. Therefore, it was decided to move forward with the impact evaluation.

Now that data have been collected, it is possible to look at the *actual* MDE and MDES for the study, based on equations (2a) and (2b). Appendix Table A.1 shows that the MDES is 0.15 for impacts on the reading total scores (based on the Group Reading Assessment and Diagnostic Examination, or GRADE) and 0.17 for impacts on math total scores (based on the Group Mathematics Assessment and Diagnostic Examination, or GMADE). An MDES like these is equivalent to about 40 percent to 47 percent of the growth in test scores expected over the course of a full year of middle school, which is more than one can expect from a program that is only five weeks long.[11] As noted in Chapter 1, BELL would have to be three times more

[8]Chaplin and Capizzano (2006).

[9]Cooper, Charlton, Valentine, and Muhlenbruck (2000).

[10]The MDES of 0.10 is based on equation (3b), with N = 1,032 — a response rate of 90 percent; P = 0.60; $R^2 = 0.55$; and a statistical significance level of 10 percent (two-tailed test).

[11]Hill, Bloom, Black, and Lipsey (2007) found that the expected growth in test scores for middle school students is 0.32 in reading for a full year of school and 0.42 in math. (These numbers are based on students in grades 5 to 7, which are the grades from which students in this study were rising.)

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table A.1

Minimum Detectable Effect (MDE) and Effect Size (MDES) for Impacts on Student Outcomes in the Fall: Fall 2012 Analysis Sample

Outcome	Number of Students	MDE	MDES
Reading achievement (standard score)[a]	919	1.90	0.15
Reading comprehension (standard score)	919	2.04	0.16
Reading vocabulary (standard score)	919	2.21	0.17
Math achievement (standard score)[a]	919	2.21	0.17
Math concepts (standard score)	919	2.43	0.18
Math operations (standard score)	919	2.54	0.18
Math processes (standard score)	919	2.64	0.21
Student engagement (1-4)[b]	919	0.10	0.22
Behavioral engagement (1-4)	919	0.12	0.23
Emotional engagement (1-4)	919	0.14	0.23

SOURCES: MDRC calculations based on the GRADE and GMADE assessments and the student survey administered in fall 2012.

NOTES: The MDE and MDES in this table are calculated based on the standard error of the impact estimate (adjusted for random assignment blocks and student baseline characteristics) and the number of students in the Fall 2012 Analysis Sample. A statistical significance level of 10 percent is assumed. The minimum detectable effect size (MDES) is calculated by dividing the MDE by the standard deviation of the outcome measure for students in the Fall 2012 Analysis Sample who are in the non-BELL group.

[a]Students enrolled in fifth grade in spring 2012 were given Level 5 of the GRADE and GMADE; students in sixth grade were given Level 6; and students in seventh grade were given Level M. The national average for GRADE and GMADE standard scores is 100, and the standard deviation is 15.

[b]The student engagement scale and subscales are based on Skinner, Furrer, Marchand, and Kindermann (2008). A student's overall engagement score is based on his or her average response to 16 items with a 4-point truth scale (1 = "not at all true," 2 = "not very true," 3 = "sort of true," and 4 = "very true"). The behavioral and emotional engagement subscales are each based on a subset of five items. The internal consistency reliability (Cronbach's alphas) of these scales for the Fall 2012 Analysis Sample are as follows: student engagement = 0.84; behavioral engagement = 0.75; emotional engagement = 0.78.

effective than regular schooling to produce effects this large in a five-week time frame. The MDES for reading (0.15) is smaller than for math (0.17) because the explanatory power (R^2) of the baseline covariates is higher for reading (0.61) than for math (0.53).

The actual minimal detectable effect sizes (0.15 and 0.17) are higher than what had been projected in the study's design phase (0.10), for three reasons:

- **The samples in each district had to be reweighted.** At the start of the study, before student recruitment began, it was expected that each district would contribute a similar number of students — and would have a similar impact — and that it would not be necessary to reweight the study districts.

But because the sample ended up being heavily skewed toward one district — and because the impacts are different across districts — it became necessary to reweight the impacts in order to estimate the impact of the average BELL program. Although weighting the districts equally is the right approach for obtaining the parameter of greatest interest (that is, the impact of BELL for the average study district), reweighting the sample increases the study's MDES by 0.03. Had the sites not been reweighted, the MDES for the effect of BELL would have been about 0.12 for reading and 0.14 for math (compared with 0.15 and 0.17, respectively).

- **Some students were excluded because they were randomly assigned (or not) to BELL's elementary school program.** In District C, students rising into sixth grade actually received BELL's elementary school model rather than the middle school model. Thus, of the 1,226 students recruited into this study, 194 were rising sixth-grade students in District C who ended up being randomly assigned to BELL's elementary school program rather than to its middle school program. These BELL students (and their non-BELL counterparts) were dropped from the study sample to avoid confounding the effect of BELL's two models; thus, the sample includes 1,032 students rather than 1,226.

- **The standard deviation for effect sizes is smaller than had been assumed.** In the study design phase, it was assumed that effect sizes would be calculated using the standard deviation of the entire analysis sample, whereas, prior to the analysis, it was decided to use the standard deviation of the non-BELL group instead. As shown in Appendix Table A.2, this is smaller than the overall standard deviation, especially in math. It was decided to use the standard deviation of the non-BELL group because this represents the amount of variation in outcomes for students who were unaffected by the program, which is a more stable reference point for calculating effect sizes.

As reported in Chapter 3, the estimated effect of BELL's middle school program on math is 0.07, which is below the MDES of 0.17; therefore, this effect is not statistically significant. To detect an effect as small as 0.07, the study would have needed to recruit a much larger sample of students — about 5,500.[12]

[12]This assumes that impacts for the three study districts would have to be reweighted, as happened in this evaluation.

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table A.2

Standard Deviations for Fall 2012 Student Outcomes: Fall 2012 Analysis Sample

Outcome	BELL Group	Non-BELL Group	All Students
Reading achievement[a]			
Total reading (standard score)	13.4	12.4	13.0
Reading comprehension (standard score)	13.4	12.6	13.1
Reading vocabulary (standard score)	13.5	12.8	13.2
Math achievement[a]			
Total math (standard score)	14.6	12.8	14.0
Math concepts (standard score)	15.2	13.8	14.7
Math operations (standard score)	15.5	14.0	15.0
Math processes (standard score)	13.4	12.8	13.2
Student engagement (1-4)[b]			
Overall engagement	0.43	0.46	0.44
Behavioral engagement	0.49	0.50	0.49
Emotional engagement	0.59	0.64	0.61
Sample size	585	334	919

SOURCES: MDRC calculations based on the GRADE and GMADE assessments and the student survey administered in fall 2012.

NOTES: The analyses reported in this table are based on the sample of students who took the GRADE and GMADE assessments and who responded to the student survey in fall 2012 (Fall 2012 Analysis Sample). The values in the column labeled "BELL Group" are the standard deviation for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the standard deviation for students randomly assigned to the non-BELL group.

[a]Students enrolled in fifth grade in spring 2012 were given Level 5 of the GRADE and GMADE; students in sixth grade were given Level 6; and students in seventh grade were given Level M. The national average for GRADE and GMADE standard score is 100, and the standard deviation is 15.

[b]The student engagement scale and subscales are based on Skinner, Furrer, Marchand, and Kindermann (2008). A student's overall engagement score is based on his or her average response to 16 items with a 4-point truth scale (1 = "not at all true," 2 = "not very true," 3 = "sort of true," and 4 = "very true"). The behavioral and emotional engagement subscales are each based on a subset of five items.

Standard Deviations Used to Calculate Effect Sizes

As explained in Chapter 3, the impact estimates in this report are presented both in their original metrics and as effect sizes. Effect sizes are based on the standard deviation of the student outcomes of interest for the non-BELL group in the Fall 2012 Analysis Sample. (Appendix C discusses this sample.)[13]

[13]The standard deviation used is the *total* standard deviation, which includes variation *between* grade levels and *within* grade levels. However, for the two primary outcomes (GRADE and GMADE scores), the *be-*

Appendix Table A.2 presents the standard deviations used to calculate effect sizes for impacts in this report (the "Non-BELL Group" column).[14] For reasons noted above, the standard deviation for the non-BELL group is used to calculate effect sizes. However, the table also presents standard deviations for the BELL group — as well as for the BELL and the non-BELL groups together — for use in future meta-analyses and research.

Appendix Table A.3 presents the standard deviations for the pre-program student characteristics that are used to describe the sample and to establish baseline equivalence in Appendix C. Similar to effect sizes for impacts, baseline differences are converted to effect sizes using the standard deviation based on non-BELL students only. However, the table also includes standard deviations for BELL students.

---

tween-grade variation in test scores is very small because each grade-level assessment is scaled to have a mean score of 100. Thus, the total standard deviation and the average within-grade standard deviation in test scores are very similar; they are the same at the first decimal point. By extension, this means that the decision to use the total standard deviation — instead of the average within-grade standard deviation — does not appreciably affect the magnitude of the effect size for the achievement outcomes.

[14]These standard deviations are used to calculate effect sizes for the overall impact of BELL as well as for impacts by district.

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table A.3

Standard Deviations for Baseline Characteristics of Students

Characteristic in Spring 2012 (%)	Study Sample		Fall 2012 Analysis Sample	
	BELL group	Non-BELL group	BELL group	Non-BELL group
Race/ethnicity				
Hispanic	49.3	45.6	49.5	46.7
Black, non-Hispanic	46.6	49.9	47.0	50.0
White, non-Hispanic	26.0	31.7	22.6	26.5
Asian	32.4	25.2	32.6	25.5
Other	26.3	25.7	25.5	25.5
Female	49.6	49.9	49.5	50.0
Eligible for free/reduced-price lunch	33.6	33.0	32.9	29.1
English as a Second Language	28.8	29.8	29.3	30.8
Parent education level[a]				
Did not finish high school	38.5	36.1	38.9	36.6
Has high school diploma or GED certificate	46.9	45.6	47.2	45.6
Completed some postsecondary education	44.8	46.7	44.3	46.9
Has bachelor's degree or higher	32.9	35.2	32.4	34.7
Other	29.1	28.5	29.3	27.8
Has an individualized education plan (IEP)	39.4	40.8	38.8	41.0
Proficient on state test in spring 2012[b]				
Reading	47.2	45.4	47.1	44.9
Math	49.4	49.5	49.3	49.4
Sample size[c]	643	389	585	334

SOURCES: MDRC calculations based on the BELL baseline intake form administered in spring 2012 and student records obtained from school districts.

NOTES: [a]For students with two guardians, this is the maximum education level of the two guardians.
[b]A student's proficiency is based on the standards in the state where he or she is attending school.
[c]Due to missing values, the number of students included varies by characteristic. The sample sizes reported in this table are for the full study sample and the Fall 2012 Analysis Sample.

Appendix B

Surveys and Testing: Timeline, Survey Scales, and Data Collection Instruments

Appendix B provides information about the timeline for administering student surveys and achievement tests that were used in fall 2012 to measure student outcomes in the Building Educated Leaders for Life (BELL) evaluation. The appendix also describes the survey scales and composite measures that were constructed from the fall 2012 student survey and from the BELL teacher survey.

Timeline for Collecting Student Data

Three instruments were used in the BELL evaluation to collect data from students: the Group Reading Assessment and Diagnostic Examination (GRADE), the Group Mathematics Assessment and Diagnostic Examination (GMADE), and the student survey. All three were administered in fall 2012, after the summer program ended, in order to make it possible to assess the outcomes of BELL and non-BELL students at the same time. These instruments were administered to students during weekends.

Appendix Table B.1 shows the survey and testing dates for the three study districts (Districts A, B, and C) and the number and percentage of students in the Fall 2012 Analysis Sample who were tested on those days. On average, across districts, about 91 percent of students took the tests and survey in the first session. In District A, 90 percent took the tests and survey in the first session; in Districts B, 95 percent did so; and, in District C, 90 percent of students took the tests and survey in the first session.[1]

Appendix Table B.2 presents the percentages of BELL students and non-BELL students who were tested in the first session. On average, across all three districts, the BELL and non-BELL groups were tested at similar times. The percentage of students in the BELL group who were tested in the first session (92 percent) does not differ statistically from the percentage of students in the non-BELL group who were tested in the first session (90 percent). In District A, however, a statistically larger proportion of BELL students than of non-BELL students were tested in the first session: 92 percent, compared with 84 percent.

Appendix Table B.3 shows the average number of days elapsed between testing and the last day of the BELL program or the first day of the school year. Students in both groups were tested about 40 days after the end of the summer program (range across districts: 33 to 46 calendar days) and 9 calendar days after the start of the school year (range across districts: 7 to 13 calendar days).

[1]The reference point (denominator) for these percentages is the number of students who took the tests and the survey (the Fall 2012 Analysis Sample).

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table B.1

Number of Students Tested, by Date, in Each Study District: Fall 2012 Analysis Sample

Date (2012)	District A	District B	District C
August 25	322 (89.9%)	111 (94.9%)	
September 8	28 (7.8%)		401 (90.3%)
September 15		2 (1.7%)	31 (7%)
September 22		2 (1.7%)	12 (2.7%)
September 26	8 (2.2%)		
September 29			
October 7			
Sample size	358	117	444

SOURCE: MDRC calculations based on testing dates for the GRADE and GMADE assessments.

NOTE: The analyses reported in this table are based on the sample of students who took the GRADE and GMADE assessments and who responded to the student survey in fall 2012 (Fall 2012 Analysis Sample).

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table B.2

Percentage of Students in the Fall 2012 Analysis Sample Who Were Tested in the First Session

Outcome	Sample Size	BELL Group	Non-BELL Group	Estimated Difference	P-Value
All districts	919	92%	90%	2%	0.482
By district					
District A	358	92%	84%	8% **	0.032
District B	117	94%	98%	-5%	0.438
District C	444	91%	89%	2%	0.454

SOURCE: MDRC calculations based on testing dates for the GRADE and GMADE assessments.

NOTES: The analyses reported in this table are based on the sample of students who took the GRADE and GMADE assessments and who responded to the student survey in fall 2012 (Fall 2012 Analysis Sample). The estimated differences between the BELL group and the non-BELL group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012. The values in the column labeled "BELL Group" are the observed means for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-BELL group, using the observed distribution of the BELL group across random assignment blocks as the basis for the adjustment. Each of the three study districts is given an equal weight when estimating the pooled results reported in this table. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to differences between BELL and non-BELL groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Appendix Table B.3

Average Number of Days Elapsed Between Fall Testing and the End of the BELL Program or the Start of the School Year

Reference Point (Date)	Sample Size	BELL Group	Non-BELL Group	Estimated Difference	P-Value
Since last day of the BELL program					
All districts	919	39.2	39.9	-0.7 *	0.082
District A (July 25, 2012)	358	32.2	35.0	-2.9 ***	0.000
District B (July 12, 2012)	117	45.6	44.8	0.8	0.442
District C (July 31, 2012)	444	39.8	40.0	-0.2	0.713
Since start of the school year					
All districts	919	8.5	9.2	-0.7 *	0.082
District A (Aug. 20, 2012)	358	6.2	9.0	-2.9 ***	0.000
District B (Aug. 20, 2012)	117	6.6	5.8	0.8	0.442
District C (Aug. 27, 2012)	444	12.8	13.0	-0.2	0.713
Since start of the school year (school days)					
All districts	919	6.5	7.0	-0.5 *	0.082
District A (Aug. 20, 2012)	358	4.8	6.9	-2.1 ***	0.000
District B (Aug. 20, 2012)	117	5.1	4.5	0.6	0.435
District C (Aug. 27, 2012)	444	9.6	9.7	-0.1	0.713

SOURCE: MDRC calculations based on testing dates for the GRADE and GMADE assessments.

NOTES: The analyses reported in this table are based on the sample of students who took the GRADE and GMADE assessments and who responded to the student survey in fall 2012 (Fall 2012 Analysis Sample). The estimated differences between the BELL group and the non-BELL group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012. The values in the column labeled "BELL Group" are the observed means for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-BELL group, using the observed distribution of the BELL group across random assignment blocks as the basis for the adjustment. Each of the three study districts is given an equal weight when estimating the pooled results reported in this table. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to differences between BELL and non-BELL groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Appendix Table B.3 also shows that, on average, testing for BELL and non-BELL students across the three districts happened at similar times. In the average study district, BELL students were tested about three-quarters of a day sooner than non-BELL students; although this difference is statistically significant, it is small in magnitude. In District A, BELL students were tested about two or three days sooner than non-BELL students.

Student Survey Scales

Student Engagement Scales

The student engagement scales used in this study are based on a scale developed by Ellen Skinner and colleagues.[2] Skinner's student engagement scale has been shown to be associated with students' academic outcomes: Students who have higher engagement scores earn higher grades and score higher on standardized achievement tests.[3]

Skinner's student engagement scale includes four subscales representing different facets of engagement — behavioral engagement, emotional engagement, behavioral disaffection, and emotional disaffection:

- The *behavioral engagement* subscale measures students' effort, attention, and persistence during learning activities. (For example, "When I'm in class, I usually think about other things.")
- The *emotional engagement* subscale measures the emotional reactions that a student experiences in the classroom, especially interest and happiness. (For example, "When I'm in class, I feel happy.")
- The *behavioral disaffection* subscale measures the opposite of behavioral engagement; it measures the extent to which students are passive, do not try, or give up. (For example, "In class, I do just enough to get by.")
- The *emotional disaffection* subscale — again measuring the opposite of emotional engagement — captures the extent to which students are bored, sad, anxious, or angry. (For example, "When we work on something in class, I feel discouraged.")

All items in the Skinner instrument are on a 4-point response scale: 1 = "Not at all true"; 2 = "Not very true"; 3 = "Sort of true"; 4 = "Very true." Students' scores on the student engagement scale and subscales are calculated by averaging their responses across all relevant survey items. If a student did not respond to an item, the value for that item is imputed using the mean of the values for the other items. By definition, average scores range from a minimum of 1 (none of the items is at all true for the student) to a maximum of 4 (all of the items are "very true" for the student).

[2]See Skinner, Furrer, Marchand, and Kindermann (2008). The Skinner scale is adapted from work by Wellborn (1991).

[3]Skinner, Wellborn, and Connell (1990).

To limit its length, the BELL student survey included only items for three of these four subscales: behavioral engagement, emotional engagement, and behavioral disaffection; emotional disaffection was not included. Appendix Table B.4 shows which specific survey items are included in the student engagement scale and subscales, as well as the reliability of each scale, based on students in the Fall 2012 Analysis Sample. The internal consistency reliability (Cronbach's alpha) of the overall student engagement scale is 0.84, while the reliability of the behavioral engagement and emotional engagement subscales is 0.75 and 0.77, respectively. Because the reliability of the behavioral disaffection subscale is only 0.61, this subscale was not included as an outcome measure in the analysis (although the disaffection-related items are still included in the overall measure of student engagement).

Students' Summer Activities

In the fall 2012 survey, students reported on the frequency with which they engaged in seven different types of summer activity: wrote a letter, poem, or story; played math games or did math problems; went to the library; read a book during free time; watched TV during weekdays; did activities at a club, community center, church, or day camp; and played in a sports program. These seven survey items are on a 5-point response scale: 1 = "Never"; 2 = "Hardly ever (1 or 2 times)"; 3 = "Not very often (once a month)"; 4 = "Sometimes (about once a week)"; 5 = "Pretty often (a couple times or more a week)." Students' responses to these items were converted to number of times per summer as follows:

- Never = 0 times per summer
- Hardly ever (1 or 2 times) = 1.5 times per summer
- Not very often (once a month) = 3 times per summer[4]
- Sometimes (about once a week) = 13 times per summer[5]
- Pretty often (a couple times or more a week) = 26 times per summer[6]

If a student answered at least one of the seven items, then missing values on any of the remaining items are imputed as zero. Composite survey measures representing the total number of times that students participated in two categories of summer activities were then created:

- **"Academic" summer activities.** Sum of a student's responses across the two academic summer activities that one would expect BELL to increase (wrote a letter, poem, or story; played math games or did math problems).

[4]This is based on the assumption that there are 3 months in the summer.

[5]This is based on the assumption that there are 13 weeks in the summer.

[6]This is based on the assumptions that there are 13 weeks in the summer and that students participated in the activity twice a week.

- **"Typical" summer activities.** Sum of a student's responses across the five typical summer activities (went to the library; read a book during free time; watched TV during weekdays; did activities at a club, community center, church, or day camp; and played in a sports program).

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table B.4

Student Engagement Scales and Subscales: Fall 2012 Analysis Sample

Survey Items	Cronbach's Alpha
All items in the student engagement scale	0.84
Behavioral engagement subscale	0.75
I try hard to do well in school	
In class, I work as hard as I can	
When I'm in class, I participate in class discussions	
I pay attention in class	
When I'm in class, I listen very carefully	
Emotional engagement subscale	0.77
When I'm in class, I feel good	
When we work on something in class, I feel interested	
Class is fun	
I enjoy learning new things in class	
When we work on something in class, I get involved	
Behavioral disaffection subscale	0.61
When I'm in class, I just act like I'm working (Reverse-coded)	
I don't try very hard at school (Reverse-coded)	
In class, I do just enough to get by (Reverse-coded)	
When I'm in class, I think about other things (Reverse-coded)	
When I'm in class, my mind wanders (Reverse-coded)	
Other items included in overall scale	
When we work on something in class, I feel bored (Reverse-coded)	--

NOTES: The analyses reported in this table are based on the sample of students who took the GRADE and GMADE assessments and who responded to the student survey in fall 2012 (Fall 2012 Analysis Sample). All items use a 4-point truth scale: 1 = "not at all true"; 2 = "not very true"; 3 = "sort of true"; and 4 = "very true." The student engagement scale and subscales are based on Skinner, Furrer, Marchand, and Kindermann (2008). The behavioral disaffection subscale is not included as an outcome in the impact analysis due to its lower reliability.

Teacher Survey Scales

In summer 2012, BELL administered a survey to its teachers as part of regular program monitoring and evaluation activities. The target population for the survey includes all teachers — academic (English Language Arts [ELA] or math or both) and enrichment teachers — as well as mentors (teaching assistants). Given the academic focus of this evaluation, analyses from the teacher survey in this report are based on the responses of academic teachers[7] who taught students in the Fall 2012 Analysis Sample. The response rate among these teachers is 85 percent in the average study district (80 percent in District A, 100 percent in District B, and 75 percent in District C).

The BELL teacher survey asks teachers to rate (1) their experience and satisfaction with various aspects of the BELL program (training, materials, staffing, and such), (2) their own performance in the classroom, and (3) their students' performance and behavior. The items in the teacher survey are grouped into sections based on these topics. Factor analysis was used to verify and confirm that the items under a particular topic were sufficiently correlated to be combined into a survey scale. When an item did not correlate with the other items under a particular topic, it was excluded from the scale; this happened for three items. Appendix Table B.5 summarizes which survey items were used to construct the teacher scales used in this evaluation and gives the reliability of each scale. As shown, the reliability of the teacher survey scales ranges from 0.80 to 0.95.

[7]This includes "dual teachers" who taught academics in the morning and enrichment activities in the afternoon. Teachers who taught only enrichment are excluded, and so are mentors.

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table B.5

Teacher Survey Scales: BELL Academic Teachers

Survey Scale and Items	Cronbach's Alpha
Usefulness and adequacy of BELL preparation and training	0.91
E-learning/online training was user-friendly and structured in a way that was easy to understand	
E-learning prepared me to use the BELL Reading Club/multicultural library resources	
After completing the classroom training with my summer colleagues, I felt prepared to work as a collaborative team	
E-learning and classroom training prepared me to use assessment data to impact scholars' academic development	
E-learning and classroom training prepared me to effectively implement the curriculum in my BELL classroom	
After completing both e-learning and classroom training, I felt prepared to be a role model for my scholars	
After completing both e-learning and classroom training, I felt prepared to manage behavior in my BELL classroom	
After completing both e-learning and classroom training, I felt inspired to implement BELL's mission	
BELL's training, both e-learning and classroom, was of a high quality	
I found the content-specific afternoon sessions (Elementary Academic, Enrichment, Middle School Math, Middle School ELA) useful in preparing me to implement/support instruction	
Usefulness and adequacy of BELL academic resources and materials	0.80
The Stanford Diagnostic Test results were useful in planning instruction	
The skill-based quiz results were useful in planning instruction	
Scholars were engaged in the literacy curriculum	
BELL's literacy resources (including the BELL Reading Club) effectively prepared scholars for school in the fall	
Scholars were engaged in the mathematics curriculum	
BELL's literacy resources (including the BELL Reading Club) fostered a love of reading in my scholars	
BELL's math resources effectively prepared scholars for school in the fall	
BELL's math resources fostered a love of math in my scholars	
Supplies in the cluster bins were age- and grade-appropriate	
Supplies in the cluster bins were adequate for my site	
The behavior management effectively managed scholar behavior in the classroom	
The behavior system allowed me to use positive discipline	
The behavior system allowed for consistent management of behavior	
The behavior system allowed for fair treatment of all scholars	
The behavior system allowed for scholars to be treated respectfully	
The behavior system allowed scholars to learn self-management	
The behavior system is consistent with the behavior management system I use during the school year	

(continued)

Appendix Table B.5 (continued)

Survey Scale and Items	Cronbach's Alpha
Level of support from BELL leadership and management team	0.95
My Summer Program manager clearly and regularly communicated the expectations for academic teachers	
My Summer Program manager clearly and regularly communicated the expectations for enrichment teachers	
My Summer Program manager clearly and regularly communicated the expectations for teaching assistants	
My Summer Program manager clearly and regularly communicated the expectations for site administrators	
I received the tools and resources I needed from the program's leadership structure to do my job well	
The policies for BELL staff were clearly communicated to me by my program manager	
BELL's payroll process was clearly explained to me by the program administrators	
My site administrators helped me to develop my skills in managing scholar behavior	
My site administrators promoted team work at my site	
I regularly met with my site administrators and/or other site staff to communicate site information (for example, upcoming events)	
I regularly met with my site administrators and/or other site staff to discuss teaching, mentoring, and/or child development strategies	
The Lead Teacher at my site gave me feedback on my instructional plans and/or delivery of instruction	
Quality of teacher's relationship with students	0.92
There are too many youth in this class for me to build a relationship with each one (Reverse-coded)	
I know all of the students in this activity by first name	
The class period is too short for me to really get to know the students (Reverse-coded)	
I feel like the youth in this class trust and respect me	
I typically look forward to spending time with the youth in this activity	
I feel very close to my BELL students	
I interact with each student (call on them or talk to them individually)	
I try to give some feedback every class to each student	

(continued)

Appendix Table B.5 (continued)

Survey Scale and Items	Cronbach's Alpha
Quality of teacher's classroom management	0.82
I rarely have behavior problems with the youth in this group	
If youth misbehave, I am comfortable dealing with it myself	
If youth misbehave, I am comfortable calling on other BELL staff to help	
Youth in this class know that there will be consequences if they act out	
I feel like I spend a lot of time trying to get youth to settle down and stop talking (Reverse-coded)	
Most youth in this class are good at following instructions	
This class often gets out of control (Reverse-coded)	
Student engagement in the program	0.92
Scholars were engaged in the literacy curriculum	
Scholars were engaged in the mathematics curriculum	
Scholars developed new skills from the afternoon enrichment classes	
Scholars enjoyed the field trips	
Field trips, guest speakers, and cultural activities enhanced the program	

NOTES: This analysis is based on teachers who responded to the BELL teacher survey and who taught students in the study sample. All items, except teacher satisfaction, use a 5-point agreement scale: 1 = "strongly disagree," 2 = "disagree," 3 = "undecided," 4 = "agree," and 5 = "strongly agree." Teacher satisfaction is based on a 10-point rating scale.

Survey Instrument: Fall 2012 Student Survey

First Name*:* _____

Last Name*:* _____

Date: __ __ / __ __ / 2012

Welcome to the BELL Evaluation Student Survey! We would like to ask you some questions about your summer, June to August 2012 activities and about your school work.

- This is not a test. There are no right or wrong answers.
- We hope that you will answer all of the questions. You do not have to answer any questions you do not want to.
- No one at your school or in your family will see your answers. Your answers will be kept secret.
- Please listen to the full question before answering.
- You will be instructed at a specific time to tear off this cover sheet. Your name will <u>not</u> be connected to any of your answers and your answers will not be seen by anyone except the researchers, not your parents or any summer or regular school staff.

INSTRUCTIONS: For each question please choose only 1 answer. Mark your answer by *circling* the number that shows how you feel OR by *filling in the box* with your pencil like this ■.
.

	(Circle One)			
	Not At All True	**Not Very True**	**Sort Of True**	**Very True**
I love chocolate ice cream.	1	2	3	4

This last summer, did you watch TV during the weekend?

■$_1$ Yes
❑$_2$ No

Thank you for helping us to learn more about students and their school work!

YOUTH SURVEY

Name of the school you are going to now: _____

Name of your current math teacher:_____

Date Youth Survey completed: __ __ / __ __ / 2012

In the next few questions, we'd like to know about the kinds of things you did during your SUMMER BREAK, June to August 2012. Think about HOW OFTEN you did each of these activities and circle the answer that is the closest to how often you did it.

- If it was NEVER, circle "0".
- If it was HARDLY EVER (meaning only once or twice during the summer), circle "1".
- If it was NOT VERY OFTEN (meaning something like once a month), circle "2".
- If it was SOMETIMES (meaning something more like once a week), circle "3".
- If it was PRETTY OFTEN (meaning a couple time or more a week), circle "4".

How often…...	**(Circle One)**				
	Never	Hardly Ever (1 or 2 times)	Not Very Often (once a month)	Some-Times (about once a week)	Pretty Often (a couple times or more a week)
1. did you go to the library?	0	1	2	3	4
2. did you write something like a letter, poem, or a story?	0	1	2	3	4
3. did you play math games or do math problems?	0	1	2	3	4
4. did you read a book during your free time?	0	1	2	3	4
5. during the day, Monday through Friday, did you just hang around watching TV?	0	1	2	3	4
6. did you do activities at a Boys and Girls club, YMCA, a community center, church or a day camp?	0	1	2	3	4
7. did you play in a sports program?	0	1	2	3	4

8. This summer, June to August 2012, did you go to a program that did reading and/or math activities or a summer school program? (PLEASE CHOOSE ONE ANSWER)

❑1 Yes, I went to the BELL program or some other program that did BOTH math and reading activities.
❑2 Yes, I went to a program that did mostly reading activities.
❑3 Yes, I went to a program that did mostly math activities.
❑4 No, I did NOT go to a summer program that did math or reading.

9. This summer, June to August 2012, on an average day Monday through Friday, what did you do?

_____
_____
_____
_____
_____

Kids can't always come every day to the programs they sign up for. We would like to find out what some of the reasons are that you may have missed days in your main summer program.

- If you NEVER MISSED a day, check the first response.
- If you DIDN'T GO to any program over the summer, check the last response for this set of questions.

10. When you missed days of your main summer program, it was because you had another activity you wanted to go to more. (PLEASE CHOOSE ONE ANSWER)

❑8 No, I never missed a day all summer.
❑0 No, this was NEVER the reason.
❑1 This was HARDLY EVER the reason.
❑2 This was NOT VERY OFTEN the reason.
❑3 This was SOMETIMES the reason.
❑4 This was the reason PRETTY OFTEN.
❑9 I didn't go to a summer program.

11. When you missed days of your main summer program, it was because you couldn't get to or get home from the program. (PLEASE CHOOSE ONE ANSWER)

❑8 No, I never missed a day all summer.
❑0 No, this was NEVER the reason.
❑1 This was HARDLY EVER the reason.
❑2 This was NOT VERY OFTEN the reason.
❑3 This was SOMETIMES the reason.
❑4 This was the reason PRETTY OFTEN.
❑9 I didn't go to a summer program.

12. When you missed days of your main summer program, it was because you didn't like what they did at the program.
(PLEASE CHOOSE ONE ANSWER)

- ❑$_8$ No, I never missed a day all summer.
- ❑$_0$ No this was NEVER the reason.
- ❑$_1$ This was HARDLY EVER the reason.
- ❑$_2$ This was NOT VERY OFTEN the reason.
- ❑$_3$ This was SOMETIMES the reason.
- ❑$_4$ This was the reason PRETTY OFTEN.
- ❑$_9$ I didn't go to a summer program.

13. When you missed days of your main summer program, it was because your parent made you do something else. (PLEASE CHOOSE ONE ANSWER)

- ❑$_8$ No, I never missed a day all summer.
- ❑$_0$ No this was NEVER the reason.
- ❑$_1$ This was HARDLY EVER the reason.
- ❑$_2$ This was NOT VERY OFTEN the reason.
- ❑$_3$ This was SOMETIMES the reason.
- ❑$_4$ This was the reason PRETTY OFTEN.
- ❑$_9$ I didn't go to a summer program.

14. When you missed days of your main summer program, it was because your family went on vacation. (PLEASE CHOOSE ONE ANSWER)

- ❑$_8$ No, I never missed a day all summer.
- ❑$_0$ No this was NEVER the reason.
- ❑$_1$ This was HARDLY EVER the reason.
- ❑$_2$ This was NOT VERY OFTEN the reason.
- ❑$_3$ This was SOMETIMES the reason.
- ❑$_4$ This was the reason PRETTY OFTEN.
- ❑$_9$ I didn't go to a summer program.

We'd like to know how school is going for you. I will read statements about how you might act or feel about school. For each statement, decide how true the statement is for you. Then circle one number that fits best.

- If you think the statement is NOT AT ALL TRUE, circle "1".
- If you think the statement is NOT VERY TRUE, circle "2".
- If the statement is SORT OF TRUE, circle "3".
- If you think the statement is VERY TRUE, circle "4."

	(Circle One)			
	Not At All True	**Not Very True**	**Sort Of True**	**Very True**
15. When I'm in class, I just act like I'm working.	1	2	3	4
16. When I'm in class, I participate in class discussions.	1	2	3	4
17. I try hard to do well in school.	1	2	3	4
18. When I'm in class, my mind wanders.	1	2	3	4
19. When we work on something in class, I feel interest-ed.	1	2	3	4
20. When I'm in class, I listen very carefully.	1	2	3	4
21. I don't try very hard at school.	1	2	3	4
22. When we work on something in class, I feel bored.	1	2	3	4
23. I enjoy learning new things in class.	1	2	3	4
24. When I'm in class, I think about other things.	1	2	3	4
25. I always finish all my homework for school.	1	2	3	4
26. When we work on something in class, I get involved.	1	2	3	4
27. I pay attention in class.	1	2	3	4
28. Class is fun.	1	2	3	4
29. In class, I do just enough to get by.	1	2	3	4
30. In class, I work as hard as I can.	1	2	3	4
31. When I'm in class, I feel good.	1	2	3	4

THANK YOU FOR HELPING US TO LEARN MORE ABOUT STUDENTS AND THEIR SCHOOLWORK!

Appendix C

Characteristics of Students in the Study and Response Analysis

Appendix C examines the baseline characteristics of students who participated in the Building Educated Leaders for Life (BELL) evaluation. Both the BELL group and the non-BELL group are examined, in order to verify that random assignment produced two groups of students who were statistically equivalent at baseline (spring 2012). The appendix also looks at response rates on the achievement tests and student surveys that were administered in fall 2012, to measure student outcomes, and it compares the characteristics of students who did and who did not complete the achievement tests and the survey. As in the analyses presented in Chapters 2 and 3, the three study districts (Districts A, B, and C) are weighted equally in all the tables presented in this appendix; therefore, the pooled results should be interpreted as the findings for the average study district.

Most tables in this appendix compare the baseline characteristics of two groups of students — for example, the BELL group and the non-BELL group or students in the Fall 2012 Analysis Sample and students excluded from the sample. Because many hypothesis tests are conducted in these tables (one for each baseline characteristic), there is an increased probability of concluding that a particular baseline difference is statistically significant when, in fact, it is not; this is a Type I error, or a "false positive."[1] For this reason, an omnibus (or joint) test is used to look for a systematic or *overall* difference between the characteristics of the BELL group and the non-BELL group. This test is reported at the bottom of tables. If the joint test is *not* statistically significant, then this means that a statistically significant difference for any individual baseline characteristic may be due to chance.

Appendix C first compares the baseline characteristics of BELL and non-BELL students in the full study sample. Then it discusses the response rates for BELL and non-BELL students on the fall 2012 achievement tests and student surveys. Next, it examines the baseline characteristics of BELL and non-BELL students who completed the achievement tests and survey (the Fall 2012 Analysis Sample). Finally, the baseline characteristics of students for whom fall data were *not* collected are compared with the characteristics of students for whom these data were available.

Characteristics of Students in the Full Study Sample

Appendix Table C.1 compares the baseline characteristics of BELL and non-BELL students in the full study sample — that is, all students who agreed to participate in the study and were randomly assigned in spring 2012. Appendix Tables C.2, C.3, and C.4 show this comparison for each study district: Districts A, B, and C.

[1] In particular, for a statistical significance level of 10 percent, one would expect to see a "false positive" for every 10 hypothesis tests conducted.

- **In the average study district, students in the BELL and non-BELL groups do not systematically differ in terms of baseline characteristics. This is also true for each of the three study districts.**

In the average district (Appendix Table C.1), BELL and non-BELL students are not statistically different from each other on any individual characteristic, and the magnitude of the differences is small (at most, 0.13 in effect size). Moreover, a joint test of the difference between the two groups across all characteristics is not statistically significant. This indicates that random assignment was successful in creating two equivalent research groups at baseline.

This conclusion also holds for each of the three study districts. Although BELL and non-BELL students differ from each other at baseline on a few characteristics in Districts A and C, there is no systematic difference across the two groups in either district based on a joint test, which indicates that these differences are likely due to chance.

Of note is the fact that, in District A, students in the BELL group scored statistically and substantially lower on state assessments before the start of the program (effect size in reading = –0.34). This preexisting difference introduces the risk that impact estimates in District A could be biased downward (too small), because BELL students were lower performing at baseline. However, sensitivity analyses conducted for this district indicate that controlling for students' baseline state test scores in the analysis is able to remove this bias. (See Appendix E.)

Response Rates and Creation of the Analysis Sample

The main impact findings for the BELL evaluation (Chapter 3) are based on the Fall 2012 Analysis Sample, which includes the subset of students for whom fall outcome data were available. Appendix Figure C.1 illustrates the creation of the analysis sample from the full study sample, and it describes the reasons why students were excluded from the analysis sample and how many were excluded. As shown in this figure, students were excluded if they did not have a score on both the Group Reading Assessment and Diagnostic Examination (GRADE) and the Group Mathematics Assessment and Diagnostic Examination (GMADE) or if they did not respond to the student survey. Also excluded are students in random assignment blocks that did not have at least one BELL student and one non-BELL student who had outcome data.

Appendix Table C.5 presents response rates for each data source, for both the BELL and the non-BELL group. Appendix Table C.6 presents response rates for each of the three districts in the study. Because the student survey and the GRADE and GMADE were all administered at the same time, the response rate across these three data sources is very similar.

- **In the average study district, the average response rate for BELL and non-BELL students is 92 percent for each data source, and the difference in rates between the two groups is not statistically significant.[2]**
- **For two study districts (Districts A and B), response rates are high (about 93 percent), and they do not differ by a statistically significant amount across the BELL and non-BELL students.**
- **In District C, response rates are lower (about 85 to 86 percent, on average, across BELL and non-BELL students), and there is a statistically significant difference between the two groups, with response rates being about 8 percent to 9 percent higher for BELL students.**

Referring to Appendix Table C.5, one can see that, in the average study district, the percentage of students included in the Fall 2012 Analysis Sample is high (about 91 percent across BELL and non-BELL students) and that the percentage does not differ statistically between the two groups (difference = 0.2 percent; p-value = 0.359).[3] The results are similar in Districts A and B (Appendix Table C.6).

In District C, a statistically greater percentage of students in the BELL group is included in the analysis sample, compared with the non-BELL group (88 percent and 81 percent, respectively; p-value = 0.011). Based on What Works Clearinghouse standards, this combination of overall attrition and differential attrition is considered "moderate attrition";[4] therefore, it must be demonstrated that baseline equivalence is maintained in the analysis sample. (The next section demonstrates this for District C.)

Characteristics of Students in the Analysis Sample

Comparison of the BELL and Non-BELL Groups in the Analysis Sample

When response rates are less than 100 percent, an important question is whether the "balance" of the experiment is preserved in the analysis sample. Accordingly, Appendix Table C.7 compares the baseline characteristics of BELL and non-BELL students in the Fall 2012

[2]As noted in Chapter 1, across all three districts, there are 1,032 students in the study sample, of whom 919 (89 percent) are in the Fall 2012 Analysis Sample. The response rate reported in Table C.5 for the Fall 2012 Analysis Sample is higher (91 percent) because it represents the response rate for the average study district; that is, the three study districts are weighted equally.

[3]Based on What Works Clearinghouse standards, this combination of overall attrition and differential attrition is considered "low attrition" (What Works Clearinghouse, 2014).

[4]What Works Clearinghouse (2014).

Analysis Sample. Appendix Tables C.8 though C.10 present these findings separately for each study district.

- **In the average study district, there is still a high degree of similarity between BELL and non-BELL students in the Fall 2012 Analysis Sample. This is also true for each of the three study districts.**

The baseline equivalence seen in the full study sample (overall and by district) carries over into the analysis sample. A joint test confirms that, overall and by district, the BELL and the non-BELL groups are not systematically different in terms of baseline characteristics. This suggests that the Fall 2012 Analysis Sample preserves the balance that was achieved with random assignment for the full study sample and that differences in fall 2012 outcomes between the two groups reflect the impact of BELL rather than preexisting differences in students' baseline characteristics.

Of note here is the fact that the baseline characteristics of BELL and non-BELL students are not systematically different in District C. This suggests that the moderately sized difference in the response rates on the fall 2012 survey and testing of BELL and non-BELL students in District C is unlikely to bias the impact estimates for this district. This conclusion is further supported by a sensitivity analysis presented in Appendix E, which is based on a sample that excludes the random assignment blocks in District C that have particularly large differential response rates between the two groups of students. Dropping these blocks does not affect the impact estimates for this district, which suggests that the impact estimates for District C represent the causal effect of BELL on student outcomes.

Also of note is the fact that, in District A (Appendix Table C.8), there remains a large and statistically significant difference with respect to students' state test scores at baseline. As noted above, sensitivity analyses conducted for this district indicate that controlling for students' baseline state test scores in the impact model is able to correct the results for this baseline difference. (See Appendix E.) Thus, impact findings for District A are also likely to be unbiased.

Comparison of Students Included and Students Excluded from the Analysis Sample

Another important question is whether the sample of students who are included in the analysis is representative of the full sample of students who were recruited into the study. Accordingly, Appendix Table C.11 compares the baseline characteristics of students in the Fall 2012 Analysis Sample and the characteristics of students who were excluded from the analysis sample due to missing outcome data. As in other analyses in this report, the three study districts are weighted equally in the pooled results.

- **In the average study district, students who were excluded from the Fall 2012 Analysis Sample do not systematically differ from students who are included in the sample.**

Although there are differences between the two groups of students on two characteristics (free or reduced-price lunch status and race/ethnicity), an omnibus test indicates that, overall, there is no systematic difference between students included and students excluded from the analysis sample.

This suggests that the impact findings from this evaluation are generalizable to students who were excluded from the analysis due to missing outcome data. In other words, the impact of BELL would have been similar for the small group of students excluded from the analysis. In addition, because the Fall 2012 Analysis Sample includes almost all students in the full study sample (91 percent of them in the average study district), the findings for this evaluation are also likely to be similar to what the findings would have been had outcome data been available for the full study sample.

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table C.1

Baseline Characteristics of Students in the Study Sample, by Treatment Group

Characteristic in Spring 2012	BELL Group	Non-BELL Group	Estimated Difference	Effect Size	P-Value for Estimated Difference
Grade level (%)					NA
Rising into grade 6	19.6	19.6	0.0	0.00	
Rising into grade 7	41.2	41.2	0.0	0.00	
Rising into grade 8	39.2	39.2	0.0	0.00	
Race/ethnicity (%)					0.986
Hispanic	33.1	32.0	1.1	0.02	
Black, non-Hispanic	42.2	45.0	-2.9	-0.06	
White, non-Hispanic	8.2	7.2	1.0	0.03	
Asian	8.7	9.4	-0.8	-0.03	
Other	7.8	6.3	1.5	0.06	
Female (%)	43.4	45.8	-2.4	-0.05	0.595
Eligible for free/reduced-price lunch (%)	88.4	87.7	0.7	0.02	0.815
English as a Second Language (%)	8.0	9.9	-1.9	-0.06	0.457
Parent education level[a] (%)					0.567
Did not finish high school	17.5	13.8	3.8	0.10	
Has high school diploma or GED certificate	33.8	27.9	5.9	0.13	
Completed some postsecondary education	28.4	34.0	-5.6	-0.12	
Has bachelor's degree or higher	12.5	14.8	-2.3	-0.06	
Other	7.7	9.5	-1.8	-0.06	
Has an individualized education plan (IEP) (%)	18.9	19.4	-0.6	-0.01	0.858
Proficient on state test in spring 2012[b] (%)					
Reading	39.4	38.1	1.3	0.03	0.755
Math	42.2	41.9	0.2	0.00	0.958
Joint test of difference between groups[c] ($\chi 2 = 13.7$)					0.883
Sample size[d] (N = 1,032)	643	389			

(continued)

Appendix Table C.1 (continued)

SOURCES: MDRC calculations based on the BELL baseline intake form administered in spring 2012 and student records obtained from school districts.

NOTES: The analyses reported in this table are based on the sample of students who applied to the BELL middle school program and were recruited into the study (study sample). The estimated differences between the BELL group and the non-BELL group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012. The values in the column labeled "BELL Group" are the observed means for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-BELL group, using the observed distribution of the BELL group across random assignment blocks as the basis for the adjustment. Each of the three study districts is given an equal weight when estimating the results reported in this table. Rounding may cause slight discrepancies in calculating sums and differences.

Effect sizes are calculated by dividing the difference by the standard deviation of the characteristic for students in the study sample who are in the non-BELL group.

A two-tailed t-test was applied to differences between BELL and non-BELL groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]For students with two guardians, this is the maximum education level of the two guardians.

[b]A student's proficiency is based on the standards in the state where he or she is attending school.

[c]A chi-square test was used to determine whether there was a systematic difference between the BELL group and the non-BELL group at baseline, based on the characteristics included in this table as well as indicators of missing data for all relevant student characteristics.

[d]Due to missing values, the number of students included varies by characteristic. The sample size reported here is for the full study sample. The percentage of missing data on any given characteristic does not exceed 10 percent.

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table C.2

Baseline Characteristics of Students in the Study Sample, by Treatment Group: District A

Characteristic in Spring 2012 (%)	BELL Group	Non-BELL Group	Estimated Difference	Effect Size	P-Value for Estimated Difference
Grade level (%)					NA
Rising into grade 7	53.3	53.3	0.0	0.00	
Rising into grade 8	46.7	46.7	0.0	0.00	
Race/ethnicity (%)					0.581
Hispanic	67.0	60.6	6.4	0.14	
Black, non-Hispanic	1.7	1.6	0.2	0.00	
White, non-Hispanic	1.4	0.1	1.2	0.04	
Asian	23.4	26.2	-2.8	-0.11	
Other	6.5	11.5	-5.0	-0.19	
Female (%)	45.6	48.9	-3.3	-0.07	0.609
Eligible for free/reduced-price lunch (%)	88.3	84.6	3.7	0.11	0.375
English as a Second Language (%)	10.0	10.5	-0.4	-0.01	0.921
Parent education level[a] (%)					0.371
Did not finish high school	19.7	19.7	0.0	0.00	
Has a high school diploma or GED certificate	40.1	38.8	1.3	0.03	
Completed some postsecondary education	23.8	15.5	8.3	0.18	
Has a bachelor's degree or higher	5.9	11.1	-5.1	-0.15	
Other	10.4	14.9	-4.5	-0.16	
Has an individualized education plan (IEP) (%)	6.44	5.72	0.72	0.02	0.811
Proficient on state test[b] (%)					
Reading	38.1	42.6	-4.5	-0.10	0.500
Math	35.9	46.6	-10.7	-0.22	0.104
State test scores[c]					
Reading	328.3	347.4	-19.1 **	-0.34	0.022
Math	329.8	355.2	-25.4 **	-0.33	0.015
Joint test of difference between groups[d] ($\chi 2 = 17.9$)					0.531
Sample size[e] (N = 385)	300	85			

(continued)

Appendix Table C.2 (continued)

SOURCES: MDRC calculations based on the BELL baseline intake form administered in spring 2012 and student records obtained from school districts.

NOTES: The analyses reported in this table are based on the sample of students who applied to the BELL middle school program and were recruited into the study (study sample). The estimated differences between the BELL group and the non-BELL group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012. The values in the column labeled "BELL Group" are the observed means for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-BELL group, using the observed distribution of the BELL group across random assignment blocks as the basis for the adjustment.

Effect sizes are calculated by dividing the difference by the standard deviation of the characteristic for students in the study sample who are in the non-BELL group.

A two-tailed t-test was applied to differences between research groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Rounding may cause slight discrepancies in calculating sums and differences.

[a]For students with two guardians, this is the maximum education level of the two guardians.

[b]A student's proficiency is based on the standards in the state where he or she is attending school.

[c]The scale of the test is the one used by the state.

[d]A chi-square test was used to determine whether there was a systematic difference between the BELL group and the non-BELL group at baseline, based on the characteristics included in this table as well as indicators of missing data for all relevant student characteristics.

[e]Due to missing values, the number of students included varies by characteristic. The sample size reported here is for the full study sample. The percentage of missing data is 0 percent (for free or reduced-price lunch), 3 percent (for race), 5 percent (for female), 10 percent (for parent education), 18 percent (for state test scores), 30 percent (for English as a Second Language).

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table C.3

Baseline Characteristics of Students in the Study Sample, by Treatment Group: District B

Characteristic in Spring 2012 (%)	BELL Group	Non-BELL Group	Estimated Difference	Effect Size	P-Value for Estimated Difference
Grade level (%)					NA
Rising into grade 6	58.8	58.8	0.0	0.00	
Rising into grade 7	16.2	16.2	0.0	0.00	
Rising into grade 8	25.0	25.0	0.0	0.00	
Race/ethnicity (%)					0.710
Hispanic	10.3	10.6	-0.4	-0.01	
Black, non-Hispanic	70.6	81.4	-10.8	-0.22	
White, non-Hispanic	10.3	5.7	4.6	0.15	
Asian	0.0	0.0	0.0	0.00	
Other	8.8	2.3	6.5	0.25	
Female (%)	43.5	41.1	2.4	0.05	0.822
Eligible for free/reduced-price lunch (%)	92.5	92.1	0.4	0.01	0.943
English as a Second Language (%)	4.5	6.9	-2.3	-0.08	0.565
Parent education level[a] (%)					0.319
Did not finish high school	15.9	6.1	9.8	0.27	
Has a high school diploma or GED certificate	38.1	20.5	17.6	0.39	
Completed some postsecondary education	30.2	52.5	-22.3	-0.48	
Has a bachelor's degree or higher	12.7	15.8	-3.1	-0.09	
Other	3.2	5.2	-2.1	-0.07	
Has an individualized education plan (IEP) (%)	16.67	25.79	-9.13	-0.22	0.289
Proficient on state test[b] (%)					
Reading	55.9	47.2	8.7	0.19	0.427
Math	44.1	38.7	5.4	0.11	0.628
State test scores[c]					
Reading	606.7	601.0	5.8	0.13	0.576
Math	597.5	592.2	5.3	0.15	0.531
Joint test of difference between groups[d] ($\chi 2 = 10.9$)					0.897
Sample size[e] (N = 127)	68	59			

(continued)

Appendix Table C.3 (continued)

SOURCES: MDRC calculations based on the BELL baseline intake form administered in spring 2012 and student records obtained from school districts.

NOTES: The analyses reported in this table are based on the sample of students who applied to the BELL middle school program and were recruited into the study (study sample). The estimated differences between the BELL group and the non-BELL group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012. The values in the column labeled "BELL Group" are the observed means for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-BELL group, using the observed distribution of the BELL group across random assignment blocks as the basis for the adjustment.

Effect sizes are calculated by dividing the difference by the standard deviation of the characteristic for students in the study sample who are in the non-BELL group.

A two-tailed t-test was applied to differences between research groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Rounding may cause slight discrepancies in calculating sums and differences.

[a]For students with two guardians, this is the maximum education level of the two guardians.

[b]A student's proficiency is based on the standards in the state where he or she is attending school.

[c]The scale of the test is the one used by the state.

[d]A chi-square test was used to determine whether there was a systematic difference between the BELL group and the non-BELL group at baseline, based on the characteristics included in this table as well as indicators of missing data for all relevant student characteristics.

[e]Due to missing values, the number of students included varies by characteristic. The sample size reported here is for the full study sample. The percentage of missing data on any given characteristic does not exceed 12 percent.

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table C.4

Baseline Characteristics of Students in the Study Sample, by Treatment Group: District C

Characteristic in Spring 2012 (%)	BELL Group	Non-BELL Group	Estimated Difference	Effect Size	P-Value for Estimated Difference
Grade level (%)					NA
Rising into grade 7	54.2	54.2	0.0	0.00	
Rising into grade 8	45.8	45.8	0.0	0.00	
Race/ethnicity (%)					0.543
Hispanic	22.1	24.8	-2.7	-0.06	
Black, non-Hispanic	54.2	52.2	2.1	0.04	
White, non-Hispanic	12.9	15.7	-2.8	-0.09	
Asian	2.6	2.1	0.5	0.02	
Other	8.1	5.2	2.9	0.11	
Female (%)	40.9	47.2	-6.3	-0.13	0.178
Eligible for free/reduced-price lunch (%)	84.4	86.4	-2.0	-0.06	0.513
English as a Second Language (%)	9.5	12.3	-2.8	-0.09	0.275
Parent education level[a] (%)					0.923
Did not finish high school	17.0	15.5	1.4	0.04	
Has a high school diploma or GED certificate	23.2	24.4	-1.2	-0.03	
Completed some postsecondary education	31.3	34.0	-2.8	-0.06	
Has a bachelor's degree or higher	18.9	17.5	1.4	0.04	
Other	9.7	8.5	1.2	0.04	
Has an individualized education plan or IEP (%)	33.45	26.79	6.66 *	0.16	0.099
Proficient on state test[b] (%)					
Reading	24.2	24.5	-0.3	-0.01	0.935
Math	46.5	40.5	6.1	0.12	0.159
State test scores[c]					
Reading	348.0	347.7	0.3	0.05	0.556
Math	352.1	351.5	0.6	0.09	0.268
Joint test of difference between groups[d] ($\chi 2 = 14.4$)					0.700
Sample size[e] (N = 520)	275	245			

(continued)

Appendix Table C.4 (continued)

SOURCES: MDRC calculations based on the BELL baseline intake form administered in spring 2012 and student records obtained from school districts.

NOTES: The analyses reported in this table are based on the sample of students who applied to the BELL middle school program and were recruited into the study (study sample). The estimated differences between the BELL group and the non-BELL group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012. The values in the column labeled "BELL Group" are the observed means for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-BELL group, using the observed distribution of the BELL group across random assignment blocks as the basis for the adjustment.

Effect sizes are calculated by dividing the difference by the standard deviation of the characteristic for students in the study sample who are in the non-BELL group.

A two-tailed t-test was applied to differences between research groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Rounding may cause slight discrepancies in calculating sums and differences.

[a]For students with two guardians, this is the maximum education level of the two guardians.

[b]A student's proficiency is based on the standards in the state where he or she is attending school.

[c]The scale of the test is the one used by the state.

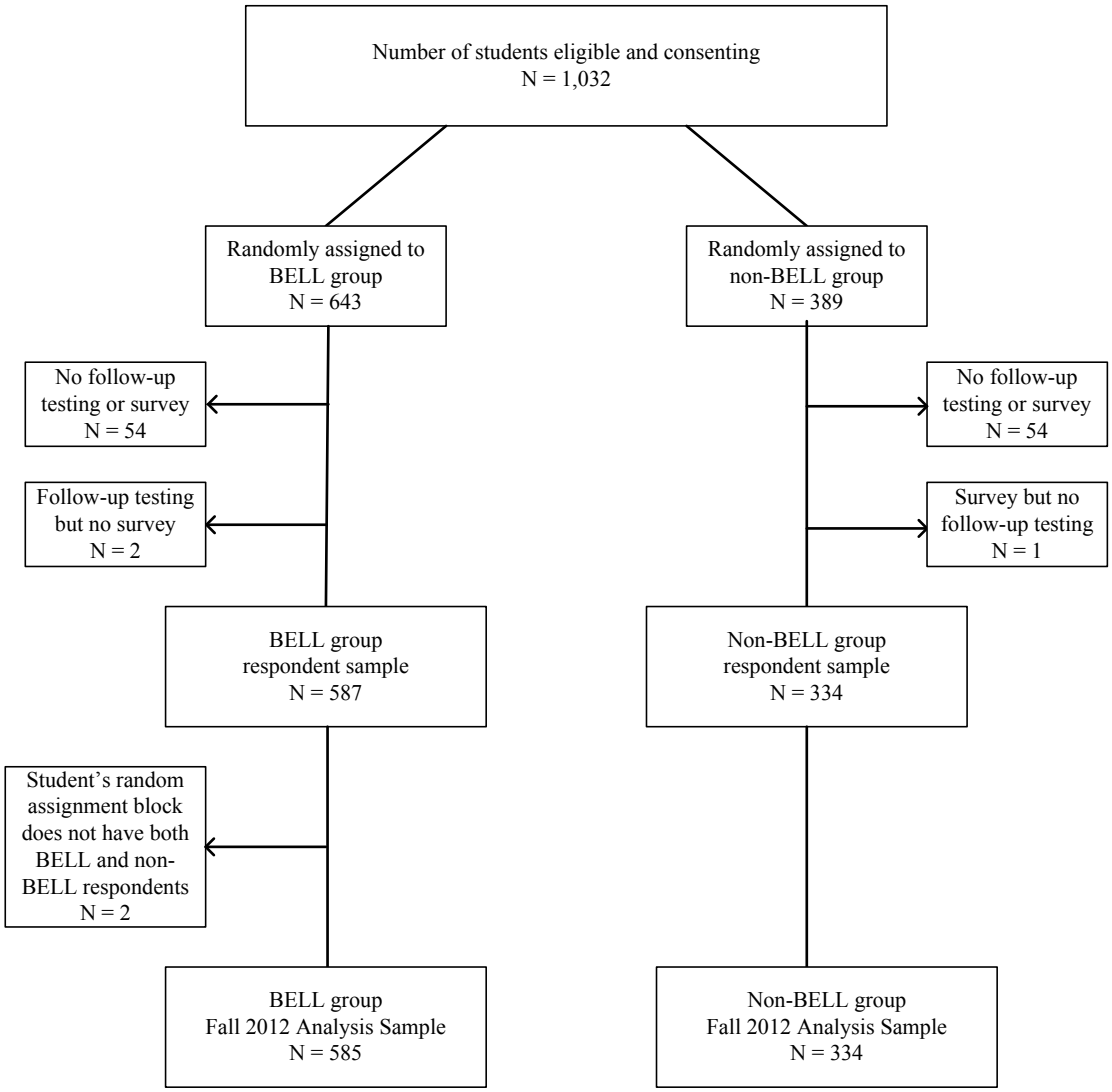
[d]A chi-square test was used to determine whether there was a systematic difference between the BELL group and the non-BELL group at baseline, based on the characteristics included in this table as well as indicators of missing data for all relevant student characteristics.

[e]Due to missing values, the number of students included varies by characteristic. The sample size reported here is for the full study sample. The percentage of missing data on any given characteristic does not exceed 7 percent.

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Figure C.1

CONSORT Chart



The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table C.5

Response Rates, by Data Source and Treatment Group

Data Source	BELL Group	Non-BELL Group	Estimated Difference	P-Value for Estimated Difference
Fall 2012 testing (%)				
GRADE assessment	91.82	91.79	0.03	0.277
GMADE assessment	91.82	91.79	0.03	0.277
Fall 2012 student survey (%)	91.58	91.56	0.02	0.352
Fall 2012 Analysis Sample[a] (%)	91.34	91.31	0.02	0.359
Sample size (N = 1,032)	643	389		

SOURCES: MDRC calculations based on the GRADE and GMADE assessments administered in fall 2012 and the student survey administered in fall 2012.

NOTES: The analyses reported in this table are based on the sample of students who applied to the BELL middle school program and were recruited into the study (study sample). The estimated differences between the BELL group and the non-BELL group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012. The values in the column labeled "BELL Group" are the observed means for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-BELL group, using the observed distribution of the BELL group across random assignment blocks as the basis for the adjustment. Each of the three study districts is given an equal weight when estimating the results reported in this table. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to differences between BELL and non-BELL groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]The Fall 2012 Analysis Sample includes students in the study sample who took the GRADE and GMADE assessments and who completed the fall 2012 student survey.

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table C.6

Study District Response Rates, by Data Source and Treatment Group

Data Source	BELL Group	Non-BELL Group	Estimated Difference	P-Value for Estimated Difference
District A				
Fall 2012 testing (%)				
GRADE assessment	93.0	92.5	0.5	0.880
GMADE assessment	93.0	92.5	0.5	0.880
Fall 2012 student survey (%)	93.0	92.5	0.5	0.880
Fall 2012 Analysis Sample[a] (%)	93.0	92.5	0.5	0.880
Sample size (N = 385)	300	85		
District B				
Fall 2012 testing (%)				
GRADE assessment	92.6	93.8	-1.2	0.835
GMADE assessment	92.6	93.8	-1.2	0.835
Fall 2012 student survey (%)	92.6	93.8	-1.2	0.835
Fall 2012 Analysis Sample[a] (%)	92.6	93.8	-1.2	0.835
Sample size (N = 127)	68	59		
District C				
Fall 2012 testing (%)				
GRADE assessment	89.8	80.7	9.1 ***	0.003
GMADE assessment	89.8	80.7	9.1 ***	0.003
Fall 2012 student survey (%)	89.1	81.2	7.9 ***	0.010
Fall 2012 Analysis Sample[a] (%)	88.4	80.6	7.8 **	0.011
Sample size (N = 520)	275	245		

SOURCES: MDRC calculations based on the GRADE and GMADE assessments administered in fall 2012 and the student survey administered in fall 2012.

NOTES: The analyses reported in this table are based on the sample of students who applied to the BELL middle school program and were recruited into the study (study sample). The estimated differences between the BELL group and the non-BELL group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012. The values in the column labeled "BELL Group" are the observed means for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-BELL group, using the observed distribution of the BELL group across random assignment blocks as the basis for the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to differences between research groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]The Fall 2012 Analysis Sample includes students in the study sample who took the GRADE and GMADE assessments and who completed the fall 2012 student survey.

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table C.7

Baseline Characteristics of Students in the Fall 2012 Analysis Sample, by Treatment Group

Characteristic in Spring 2012 (%)	BELL Group	Non-BELL Group	Estimated Difference	Effect Size	P-Value for Estimated Difference
Grade level					NA
Rising into grade 6	19.6	19.6	0.0	0.00	
Rising into grade 7	41.6	41.6	0.0	0.00	
Rising into grade 8	38.8	38.8	0.0	0.00	
Race/ethnicity					1.000
Hispanic	33.9	34.3	-0.4	-0.01	
Black, non-Hispanic	44.1	45.4	-1.4	-0.03	
White, non-Hispanic	6.2	4.7	1.6	0.06	
Asian	8.6	9.3	-0.7	-0.03	
Other	7.2	6.3	0.9	0.04	
Female	43.0	46.2	-3.2	-0.06	0.492
Eligible for free/reduced-price lunch	89.1	90.1	-1.0	-0.03	0.720
English as a Second Language	8.4	11.0	-2.6	-0.09	0.319
Parent education level[a]					0.636
Did not finish high school	17.7	15.5	2.2	0.06	
Has high school diploma or GED certificate	34.8	27.6	7.3	0.16	
Has some postsecondary education	27.0	33.1	-6.1	-0.13	
Has bachelor's degree or higher	12.5	14.6	-2.1	-0.06	
Other	7.9	9.2	-1.3	-0.05	
Has an individualized education plan (IEP)	18.1	19.5	-1.4	-0.03	0.667
Proficient on state test in spring 2012[b]					
Reading	39.5	37.1	2.4	0.05	0.568
Math	42.3	40.6	1.6	0.03	0.715
Joint test of difference between groups[c] ($\chi 2 = 12.3$)					0.950
Sample size[d] (N = 919)	585	334			

(continued)

Appendix Table C.7 (continued)

SOURCES: MDRC calculations based on the BELL baseline intake form administered in spring 2012 and student records obtained from school districts.

NOTES: The analyses reported in this table are based on the sample of students who took the GRADE and GMADE assessments and who responded to the student survey in fall 2012 (Fall 2012 Analysis Sample). The estimated differences between the BELL group and the non-BELL group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012. The values in the column labeled "BELL Group" are the observed means for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-BELL group, using the observed distribution of the BELL group across random assignment blocks as the basis for the adjustment. Each of the three study districts is given an equal weight when estimating the results reported in this table. Rounding may cause slight discrepancies in calculating sums and differences.

Effect sizes are calculated by dividing the difference by the standard deviation of the characteristic for students in the Fall 2012 Analysis Sample who are in the non-BELL group.

A two-tailed t-test was applied to differences between BELL and non-BELL groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]For students with two guardians, this is the maximum education level of the two guardians.

[b]A student's proficiency is based on the standards in the state where he or she is attending school.

[c]A chi-square test was used to determine whether there was a systematic difference between the BELL group and the non-BELL group at baseline, based on the characteristics included in this table as well as indicators of missing data for all relevant student characteristics.

[d]Due to missing values, the number of students included varies by characteristic. The sample size reported here is for the full Fall 2012 Analysis Sample. The percentage of missing data on any given characteristic does not exceed 10 percent.

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table C.8

Baseline Characteristics of Students in the Fall 2012 Analysis Sample, by Treatment Group: District A

Characteristic in Spring 2012 (%)	BELL Group	Non-BELL Group	Estimated Difference	Effect Size	P-Value for Estimated Difference
Grade level (%)					NA
Rising into grade 7	52.7	52.7	0.0	0.00	
Rising into grade 8	47.3	47.3	0.0	0.00	
Race/ethnicity (%)					0.744
Hispanic	67.9	63.3	4.6	0.10	
Black, non-Hispanic	1.5	0.2	1.3	0.03	
White, non-Hispanic	0.7	0.2	0.6	0.02	
Asian	23.2	25.2	-1.9	-0.08	
Other	6.6	11.2	-4.5	-0.18	
Female (%)	45.1	49.3	-4.2	-0.08	0.530
Eligible for free/reduced-price lunch (%)	89.2	87.2	2.0	0.07	0.623
English as a Second Language (%)	10.8	11.4	-0.5	-0.02	0.907
Parent education level[a] (%)					0.442
Did not finish high school	19.9	21.1	-1.2	-0.03	
Has high school diploma or GED certificate	41.4	38.8	2.6	0.06	
Completed some postsecondary education	22.7	15.1	7.6	0.16	
Has bachelor's degree or higher	5.2	10.1	-5.0	-0.14	
Other	10.8	14.8	-4.0	-0.14	
Has an individualized education plan (IEP) (%)	6.52	4.91	1.61	0.04	0.605
Proficient on state test[b] (%)					
Reading	37.8	40.7	-2.9	-0.06	0.669
Math	36.4	45.0	-8.7	-0.18	0.197
State test scores[c]					
Reading	327.5	346.0	-18.5 **	-0.33	0.032
Math	329.5	353.4	-23.9 **	-0.31	0.026
Joint test of difference between groups[d] ($\chi 2 = 15.4$)					0.697
Sample size[e] (N = 358)	279	79			

(continued)

Appendix Table C.8 (continued)

SOURCES: MDRC calculations based on the BELL baseline intake form administered in spring 2012 and student records obtained from school districts.

NOTES: The analyses reported in this table are based on the sample of students who took the GRADE and GMADE assessments and who responded to the student survey in fall 2012 (Fall 2012 Analysis Sample). The estimated differences between the BELL group and the non-BELL group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012. The values in the column labeled "BELL Group" are the observed means for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-BELL group, using the observed distribution of the BELL group across random assignment blocks as the basis for the adjustment.

Effect sizes are calculated by dividing the difference by the standard deviation of the characteristic for students in the Fall 2012 Analysis Sample who are in the non-BELL group.

A two-tailed t-test was applied to differences between research groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Rounding may cause slight discrepancies in calculating sums and differences. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the baseline characteristic for students in the Fall 2012 Analysis Sample who are in the non-BELL group.

[a]For students with two guardians, this is the maximum education level of the two guardians.

[b]A student's proficiency is based on the standards in the state where he or she is attending school.

[c]The scale of the test is the one used by the state.

[d]A chi-square test was used to determine whether there was a systematic difference between the BELL group and the non-BELL group at baseline, based on the characteristics included in this table as well as indicators of missing data for all relevant student characteristics.

[e]Due to missing values, the number of students included varies by characteristic. The sample size reported here is for the full Fall 2012 Analysis Sample. The percentage of missing data is 0 percent (for free or reduced-price lunch), 3 percent (for race), 5 percent (for female), 10 percent (for parent education), 17 percent (for state test scores), 30 percent (for English as a Second Language).

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table C.9

Baseline Characteristics of Students in the Fall 2012 Analysis Sample, by Treatment Group: District B

Characteristic in Spring 2012 (%)	BELL Group	Non-BELL Group	Estimated Difference	Effect Size	P-Value for Estimated Difference
Grade level (%)					NA
Rising into grade 6	58.7	58.7	0.0	0.00	
Rising into grade 7	17.5	17.5	0.0	0.00	
Rising into grade 8	23.8	23.8	0.0	0.00	
Race/ethnicity (%)					0.833
Hispanic	11.1	11.4	-0.3	-0.01	
Black, non-Hispanic	73.0	81.8	-8.8	-0.18	
White, non-Hispanic	7.9	4.3	3.6	0.14	
Asian	0.0	0.0	0.0	0.00	
Other	7.9	2.5	5.5	0.21	
Female (%)	43.9	39.1	4.8	0.10	0.664
Eligible for free/reduced-price lunch (%)	93.5	92.3	1.2	0.04	0.821
English as a Second Language (%)	4.9	7.3	-2.4	-0.08	0.586
Parent education level[a] (%)					0.319
Did not finish high school	15.5	7.7	7.8	0.21	
Has a high school diploma or GED certificate	39.7	19.6	20.0	0.44	
Completed some postsecondary education	27.6	50.4	-22.8	-0.49	
Has a bachelor's degree or higher	13.8	16.8	-3.0	-0.09	
Other	3.4	5.5	-2.0	-0.07	
Has an individualized education plan (IEP) (%)	14.75	26.43	-11.68	-0.29	0.181
Proficient on state test[b] (%)					
Reading	57.9	48.9	9.0	0.20	0.425
Math	45.6	37.8	7.8	0.16	0.489
State test scores[c]					
Reading	608.2	601.2	7.0	0.16	0.510
Math	598.6	591.8	6.8	0.19	0.435
Joint test of difference between groups[d] ($\chi 2 = 11$)					0.892
Sample size[e] (N = 117)	63	54			

(continued)

Appendix Table C.9 (continued)

SOURCES: MDRC calculations based on the BELL baseline intake form administered in spring 2012 and student records obtained from school districts.

NOTES: The analyses reported in this table are based on the sample of students who took the GRADE and GMADE assessments and who responded to the student survey in fall 2012 (Fall 2012 Analysis Sample). The estimated differences between the BELL group and the non-BELL group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012. The values in the column labeled "BELL Group" are the observed means for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-BELL group, using the observed distribution of the BELL group across random assignment blocks as the basis for the adjustment.

Effect sizes are calculated by dividing the difference by the standard deviation of the characteristic for students in the Fall 2012 Analysis Sample who are in the non-BELL group.

A two-tailed t-test was applied to differences between research groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Rounding may cause slight discrepancies in calculating sums and differences. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the baseline characteristic for students in the Fall 2012 Analysis Sample who are in the non-BELL group.

[a]For students with two guardians, this is the maximum education level of the two guardians.

[b]A student's proficiency is based on the standards in the state where they are attending school.

[c]The scale of the test is the one used by the state.

[d]A chi-square test was used to determine whether there was a systematic difference between the BELL group and the non-BELL group at baseline, based on the characteristics included in this table as well as indicators of missing data for all relevant student characteristics.

[e]Due to missing values, the number of students included varies by characteristic. The sample size reported here is for the full Fall 2012 Analysis Sample. The percentage of missing data on any given characteristic does not exceed 9 percent.

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table C.10

Baseline Characteristics of Students in the Fall 2012 Analysis Sample, by Treatment Group: District C

Characteristic in Spring 2012 (%)	BELL Group	Non-BELL Group	Estimated Difference	Effect Size	P-Value for Estimated Difference
Grade level (%)					NA
Rising into grade 7	54.7	54.7	0.0	0.00	
Rising into grade 8	45.3	45.3	0.0	0.00	
Race/ethnicity (%)					0.649
Hispanic	22.6	28.2	-5.6	-0.12	
Black, non-Hispanic	57.7	54.4	3.4	0.07	
White, non-Hispanic	10.0	9.5	0.5	0.02	
Asian	2.5	2.7	-0.2	-0.01	
Other	7.1	5.2	1.9	0.07	
Female (%)	39.9	50.1	-10.2 **	-0.20	0.049
Eligible for free/reduced-price lunch (%)	84.4	90.6	-6.2 *	-0.21	0.050
English as a Second Language (%)	9.5	14.5	-5.0 *	-0.16	0.082
Parent education level[a] (%)					0.788
Did not finish high school	17.7	17.8	0.0	0.00	
Has a high school diploma or GED certificate	23.4	24.3	-0.9	-0.02	
Completed some postsecondary education	30.7	33.7	-2.9	-0.06	
Has a bachelor's degree or higher	18.6	16.9	1.7	0.05	
Other	9.5	7.4	2.1	0.08	
Has an individualized education plan (IEP) (%)	32.92	27.13	5.79	0.14	0.196
Proficient on state test[b] (%)					
Reading	22.8	21.7	1.2	0.03	0.771
Math	44.8	39.0	5.8	0.12	0.217
State test scores[c]					
Reading	347.7	347.1	0.6	0.10	0.352
Math	351.9	351.3	0.5	0.08	0.372
Joint test of difference between groups[d] ($\chi 2 = 17.8$)					0.470
Sample size[e] (N = 444)	243	201			

(continued)

Appendix Table C.10 (continued)

SOURCES: MDRC calculations based on the BELL baseline intake form administered in spring 2012 and student records obtained from school districts.

NOTES: The analyses reported in this table are based on the sample of students who took the GRADE and GMADE assessments and who responded to the student survey in fall 2012 (Fall 2012 Analysis Sample). The estimated differences between the BELL group and the non-BELL group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012. The values in the column labeled "BELL Group" are the observed means for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-BELL group, using the observed distribution of the BELL group across random assignment blocks as the basis for the adjustment.

Effect sizes are calculated by dividing the difference by the standard deviation of the characteristic for students in the Fall 2012 Analysis Sample who are in the non-BELL group.

A two-tailed t-test was applied to differences between research groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Rounding may cause slight discrepancies in calculating sums and differences. Effect sizes are calculated by dividing the impact estimate by the standard deviation of the baseline characteristic for students in the Fall 2012 Analysis Sample who are in the non-BELL group.

[a]For students with two guardians, this is the maximum education level of the two guardians.

[b]A student's proficiency is based on the standards in the state where they are attending school.

[c]The scale of the test is the one used by the state.

[d]A chi-square test was used to determine whether there was a systematic difference between the BELL group and the non-BELL group at baseline, based on the characteristics included in this table as well as indicators of missing data for all relevant student characteristics.

[e]Due to missing values, the number of students included varies by characteristic. The sample size reported here is for the full Fall 2012 Analysis Sample. The percentage of missing data on any given characteristic does not exceed 6 percent.

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table C.11

Baseline Characteristics of Students in the Fall 2012 Analysis Sample, Relative to Students Excluded from the Analysis Sample (Nonrespondents)

Characteristic in Spring 2012	Analysis Sample	Non-Respondents	Estimated Difference		P-Value for Estimated Difference
Race/ethnicity (%)				*	0.061
Hispanic	33.6	22.0	11.6		
Black, non-Hispanic	43.9	31.9	12.0		
White, non-Hispanic	6.6	26.8	-20.1		
Asian	8.6	10.4	-1.8		
Other	7.2	9.0	-1.8		
Female (%)	43.6	42.9	0.7		0.922
Eligible for free/reduced-price lunch (%)	90.0	75.7	14.3	***	0.001
English as a Second Language (%)	8.6	3.0	5.7		0.157
Parent education level[a] (%)					0.836
Did not finish high school	16.4	13.0	3.4		
Has a high school diploma or GED certificate	33.1	29.0	4.1		
Completed some postsecondary education	30.1	39.4	-9.2		
Has a bachelor's degree or higher	12.4	11.6	0.8		
Other	7.9	7.0	0.9		
Has an individualized education plan (IEP) (%)	19.1	15.7	3.5		0.506
Proficient on state test in spring 2012[b] (%)					
Reading	36.7	30.4	6.3		0.404
Math	41.3	41.4	0.0		0.996
Joint test of difference between groups[c] ($\chi 2 = 26.7$)					0.221
Sample size[d] (N = 1,032)	919	113			

(continued)

Appendix Table C.11 (continued)

SOURCES: MDRC calculations based on the BELL baseline intake form administered in spring 2012 and student records obtained from school districts.

NOTES: The Fall 2012 Analysis Sample includes students in the study sample who took the GRADE and GMADE assessments and who completed the fall 2012 student survey. Nonrespondents are students who are excluded from the analysis due to missing outcomes data. The estimated differences between the analysis sample and nonrespondents are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012. The values in the column labeled "Analysis Sample" are the observed means for students in the analysis sample. The "Non-Respondents" values in the next column are the regression-adjusted means for students excluded from the analysis sample, using the observed distribution of the BELL group across random assignment blocks as the basis for the adjustment. Each of the three study districts is given an equal weight when estimating the results reported in this table. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to differences between respondents and nonrespondents. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]For students with two guardians, this is the maximum education level of the two guardians.

[b]A student's proficiency is based on the standards in the state where he or she is attending school.

[c]A chi-square test was used to determine whether there was a systematic difference between students in the analysis sample and nonrespondents at baseline, based on the characteristics included in this table as well as indicators of missing data for all relevant student characteristics.

[d]Due to missing values, the number of students included varies by characteristic. The sample size reported here is for the full study sample. The percentage of missing data on any given characteristic does not exceed 11 percent.

Appendix D

Data Collection During the Site Visits

As explained in Chapter 1, the team evaluating the middle school summer program offered by Building Educated Leaders for Life (BELL) visited each of the three study districts during the third and fourth weeks of the BELL program in summer 2012. The research team visited all five program schools attended by students in the study (one school in District A, one school in District B, and three schools in District C). During these visits, interviews were conducted with school program leaders (the program manager, the assistant program manager, and the lead teacher at each school); with BELL regional leaders (one per study district); and with school district liaisons (two of the three study districts). In addition, focus groups were conducted with teachers (including both academic and enrichment teachers) and mentors (teaching assistants). The protocols used by the evaluation team to conduct the interviews and focus groups follow.

Appendix Table D.1 shows the number and percentage of relevant BELL staff who were interviewed. As shown, the evaluation team was able to talk to all program leaders and all regional leaders. In addition, across the program schools, the study team interviewed about 46 percent of teachers and 51 percent of mentors, on average.

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table D.1

Number of Staff Interviewed or Included in Focus Groups, by District (As a Percentage of Total Staff of That Type at the School or District)

	District A (1 School)	District B (1 School)	District C		
			School 1	School 2	School 3
Number of Program Leader Interviews					
Program manager	1 of 1 (100%)	1 of 1 (100%)	1 of 1 (100%)	1 of 1 (100%)	1 of 1 (100%)
Assistant program manager	NA	1 of 1 (100%)	1 of 1 (100%)	NA	1 of 1 (100%)
Lead teacher	1 of 1 (100%)	1 of 1 (100%)	1 of 1 (100%)	1 of 1 (100%)	1 of 1 (100%)
Regional leader	1 of 1 (100%)	1 of 1 (100%)		2 of 2 (100%)	
District liaison	NA	1 of 1 (100%)		1 of 1 (100%)	
Number of Teachers and Mentors in Focus Groups[a]					
Teachers (academic and enrichment teachers)	7 of 15 (47%)	12 of 19 (63%)	8 of 18 (44%)	8 of 19 (42%)	7 of 22 (32%)
Mentors	6 of 15 (40%)	7 of 8 (88%)	4 of 8 (50%)	4 of 12 (33%)	5 of 11 (45%)

NOTES: The first number in each set is the number of staff in a particular category who were interviewed, and the second number is the total number of staff in a particular position.

[a]At the school in District A, there was one focus group with teachers and one with mentors. In Districts B and C, there was one focus group per school with mentors and two focus groups per school with teachers (one focus group for academic teachers and one for enrichment teachers). The sizes of the academic teacher focus groups in these two districts are the following: six teachers at the school in District B, four teachers in School 1 in District C, four teachers in School 2 in District C, and three teachers in School 3 in District C.

Interview Protocol for Program Managers

(Also Used for Assistant Program Managers)

"Thank you for agreeing to take part in the BELL Summer Program Implementation Study interview. Your participation will help BELL and the program funders understand how the program is being implemented. You do not have to answer any question that makes you feel uncomfortable and may terminate or leave the interview at any point. If you complete only part of the discussion, please note that MDRC may use whatever information was collected about you before that point.

You will not be identified in any of the papers that are written from this interview. Note, however, if keeping your answers confidential would put you or someone else in danger, then we will have to tell the appropriate agencies in order to protect you or the other person. Your comments may be repeated or quoted in some documents, but the names of students, teachers, and administrators will not be used in published reports.

The interview will last approximately 60-90 minutes.

Do you give permission for me to type or write notes during the interview? Notes prepared from this interview will not include any identifying information such as your name.

Additionally, do you give us permission to record the interview? The recordings are for the interviewers' use only and will be stored securely until they are reviewed to confirm the accuracy of our notes, after which time they will be destroyed. Also, they will not be shared with anyone outside of MDRC."

<u>Administrator Background</u>

1. How did you end up coming to work in the BELL program?
    a. What were you doing before you came to BELL? How long have you been working with BELL?
    b. Previous experience in K-12 education; ties to local school system and community
2. What is your role at BELL?
    a. Has your role changed over time?

<u>Program local context</u>

3. Are the BELL students very similar or different from the average [insert locality name] middle school student in this area?
    a. In what ways besides testing below grade level?

4. Please confirm the criteria used for identifying students who are eligible to attend the BELL summer program.
    a. How did you select students to participate in the BELL program in this area?
        i. Any mandatory students?
        ii. What was the level of interest among students eligible for the program (difficulty to recruit vs. waiting list)?
5. Are there other summer academic programs offered to 5th-7th grade students in this area?
    a. If so, what are the following characteristic of the those programs:
        i. Length
        ii. Curriculum
        iii. Topics covered
        iv. Staff qualifications
        v. Key differences from BELL program
    b. What percentage of students in the district attend one of these programs?
6. Do the BELL students in [insert locality name] face unique challenges relative to other 5th-7th graders? If so, what are they?
7. How does the BELL program interface with the school district (central and school level staff) in the administration of the summer program?
    a. How does BELL work with the school district in managing the summer program?
    b. How does BELL work with the district in implementing the summer program?
    c. Does BELL meet regularly with a district representative?
        i. If so, Who?
        ii. How often?
        iii. What is typically discussed at these meetings?
    d. Does the district and BELL have defined roles in the administration of the summer program?
8. Does the BELL program face any challenges specific to this region?
    a. If so, what are they?
9. What is your assessment of the teacher/staff training BELL provided? (Probe: utility, scope, length)
10. What is your assessment of BELL's ability to cover all needed activities with the resources that are allocated for the program in the budget?
    a. How do you or other BELL leaders make decisions around the budget?

Program Characteristics

11. How does BELL implement and manage all of the components of the summer program?
    a. Please discuss roles and responsibilities of key staff for each of the following program components:
        - Community Time
        - Classroom set up/management
        - Behavioral management
        - Teaching and learning: Academics
        - Teaching and learning: Enrichment
        - Parent engagement
    b. How do you monitor what's going on in the program?
12. Are there other key program components that we have not discussed? If so, how are they managed?
13. What types of enrichment courses are offered?
    a. How were the enrichment courses selected?
    b. How were teachers recruited for teaching enrichment courses?
        i. Recruited to teach specific enrichment classes or recruited to teach one of multiple options?
14. How were students placed into enrichment courses?
    a. What is your assessment of this course selection process? [In districts where students choose the enrichment courses] what happens when there is extra demand?
        i. How frequently were kids not put in their first choices?
15. What is the teacher/TA arrangement for academic and enrichment instruction?
    a. Do TA's stay with the students or teachers during the academic classes?
    b. Do TAs stay with the students or teachers during the enrichment classes?
16. Please describe BELL's attendance policy in this district for 5th-7th graders being implemented in this district.
    a. How is this policy enforced at the site level?
    b. How do is this policy communicated to parents and students?
17. Please describe the behavior management policy being implemented in this district.
    a. How is this policy enforced at the site level?
        i. To what extent is there variation in implementation across teachers/TAs in using the behavioral model?
    b. What types of disciplinary problems does BELL face?
        i. How does BELL address these problems?
    c. What is your assessment of the BELL behavioral model? (Probe: age appropriateness/effectiveness)

18. How do the 5th-7th grade students experience the broader community through the BELL summer program?
    a. What was your level of involvement with the local community prior to BELL?
    b. Did you already have existing community ties or do you have to develop them after taking this position at BELL?
19. How does the BELL program offer positive adult role models and mentors?
20. How does BELL offer opportunities for students to experience success?
21. How does the BELL program engage parents?

Quality and Fidelity to Program Model

22. How has each program component been implemented so far?
    a. How does BELL evaluate program quality and fidelity to the model?
    b. How is the data BELL gathers on program quality used?
    c. How has this district's TA arrangement been working so far?
        i. Is it being implemented consistently across clusters and grades?
        ii. Would you change it going forward?
    d. How has the academic instruction been going?
        i. Is it being implemented consistently across clusters and grades?
        ii. Would you change it going forward?
    e. How has the enrichment implementation been going?
        i. Is it being implemented consistently across clusters and grades?
        ii. Would you change it going forward?
    f. How has community time/community engagement been going?
        i. Is it being implemented consistently across clusters and grades?
        ii. Would you change it going forward?
23. Have any of the summer program components not been implemented as planned?
    a. If so, why?
    b. How have these changes affected how the BELL program operates?
    c. Are there components that need to be enhanced?
        i. If so, what are they?
24. How is BELL doing in its efforts to close the summer learning loss gap?
    a. Why?

Programmatic Relationships

25. How would you describe your management style?
    a. How do you typically communicate with staff?
        i. Regular meetings?
        ii. Electronic communication?
        iii. Topics of interaction — most freq, next most, etc.

26. What is your assessment of your relationship with your staff?
    a. Examples?
27. Do you use a more centralized or decentralized approach?
    a. Examples?
28. What is your assessment of the general BELL organizational/structural model? (National staff/administration, field director, program manager, lead teacher, teacher, TA)
    a. Is it an effective model for this locality?

Advice

29. What advice would you give to other program administrators in terms of operating the summer program?
    a. Probe: What advice would you give to other program administrators related to having a good working relationship with the school district?
30. What advice would you give to BELL in terms of the overall program model?

Interview Protocol for Lead Teachers

"Thank you for agreeing to take part in the BELL Summer Program Implementation Study interview. Your participation will help BELL and the program funders understand how the program is being implemented. You do not have to answer any question that makes you feel uncomfortable and may terminate or leave the focus group/interview at any point. If you complete only part of the discussion, please note that MDRC may use whatever information was collected about you before that point.

You will not be identified in any of the papers that are written from this interview. Note, however, if keeping your answers confidential would put you or someone else in danger, then we will have to tell the appropriate agencies in order to protect you or the other person. Your comments may be repeated or quoted in some documents, but the names of students, teachers, and administrators will not be used in published reports.

The interview will last approximately 60-90 minutes.

Do you give permission for me to type or write notes during the interview? Notes prepared from this interview will not include any identifying information such as your name.

Additionally, do you give us permission to record the interview? The recordings are for the interviewers' use only and will be stored securely until they are reviewed to confirm the accuracy of our notes, after which time they will be destroyed. Also, they will not be shared with anyone outside of MDRC."

<u>Administrator Background</u>

1. How did you end up coming to work in the BELL program?
    a. What were you doing before you came to BELL?
    b. How long have you been working with BELL?
    c. Previous experience in K-12 education; ties to local school system and community?
2. What is your role at BELL?
    a. Probe: Has your role changed over time?

<u>Program local context</u>

3. Are the BELL students very similar or different from the average [insert locality name] middle school student in this area?
    a. In what ways besides testing below grade level?

4. Are there other summer academic programs offered to 5th-7th grade students in this area?
    a. If so, what are the following characteristic of the those programs:
        i. Length
        ii. Curriculum
        iii. Topics covered
        iv. Staff qualifications
        v. Key differences from BELL program
    b. What percentage of students in the district attend one of these programs?
5. Do the BELL students in [insert locality name] face unique challenges relative to other 5th-7th graders? If so, what are they?

<u>Program Characteristics</u>

6. How does BELL implement and manage all of the components of the summer program?
    a. Please discuss roles and responsibilities of key staff for each of the following program components and how you do or do not provide support in each area:
        - Community Time
        - Classroom set up/management
        - Behavioral management
        - Teaching and learning: Academics
        - Teaching and learning: Enrichment
        - Parent engagement
7. Are there other key program components that we have not discussed?
    a. If so, what are they?
    b. Do you provide support for teachers and staff in implementing these components?
8. How do you use EduSoft Data?
    a. Is it a helpful tool for supporting teachers?
    b. How is the data gathered through EduSoft used?

<u>Training</u>

9. What type of preparation did you receive for assuming the Lead Teacher role in the BELL summer program?
    a. How do you feel about the training you received?
10. How would you assess the training provided by BELL for Lead Teachers, TAs and Teachers for the previously discussed program components?
    a. Usefulness, scope, length?
    b. Things you would not change?
    c. Ideas for improvement?

Quality and Fidelity to Program Model

11. How has each program component been implemented so far?
    a. How does BELL evaluate program quality and fidelity to the model?
    b. How is the data BELL gathers on program quality used?
    c. How has this district's TA arrangement been working so far?
        i. Is it being implemented consistently across clusters and grades?
        ii. Would you change it going forward?
    d. How has the academic instruction been going?
        i. Is it being implemented consistently across clusters and grades?
        ii. Would you change it going forward?
    e. How has the enrichment implementation been going?
        i. Is it being implemented consistently across clusters and grades?
        ii. Would you change it going forward?
    f. How has community time/community engagement been going?
        i. Is it being implemented consistently across clusters and grades?
        ii. Would you change it going forward?
12. Have any of the summer program components not been implemented as planned?
    a. If so, why?
    b. How have these changes affected how the BELL program operates?
    c. Are there components that need to be enhanced?
        i. If so, what are they?
13. Have you had to respond to issues that have arisen in implementing the reading and math curriculum?
    a. If so, what issues?
    b. How were you made aware of the issue?
    c. What types if staff were involved?

Programmatic Relationships

14. How would you describe the management style of your site director?
    a. One word or short descriptive phrase?
    b. Do you feel this style is generally effective with the teachers and TAs at this site?
    c. Did you work with the PM prior to working in BELL?
        i. If so, what was your previous working relationship like?
15. What is your relationship with the teachers?
    a. How do you support them?
    b. How do you typically communicate with them?
16. What is your relationship with the TAs?
    a. How do you support them?
    b. How do you typically communicate with them?

c. At some BELL sites, TAs remain with the same students all day. At other sites, the TA remains with the same teacher all day. What is your assessment of each approach?

17. What type of relationships do the TAs and Teachers have with the students?
    a. How do the student relate to them?
18. How would you assess student engagement, overall?
    a. In what types of classes and with what types of teachers are students most engaged?
    b. What key factors influence student engagement?
        i. Self-selection of courses?
        ii. Mandatory status?
        iii. Parental support?
        iv. Individual motivation?
        v. Teacher investment?
19. Are teachers' and or TAs' engagement with mandatory students any different than with students who were not mandated to participate in the program?
    a. Do you even know who the mandatory students are?
    b. Do you find that classes with a majority of mandatory students are different in some ways from classes with none or only a few?
    c. Do you or the program manager provide any specialized assistance related to mandatory students?
20. Are teachers' or TAs' engagement with students who have severe IEPs any different than with other students who do not have severe IEPs?
    a. At what point in the program were you made aware of students in your classes who have severe IEPs?
    b. Do you find that classes with more than a few students with severe IEPs are different in some ways from classes with none or only a few?
    c. Do you or the program manager provide any specialized assistance related to students with severe IEPs?

Behavior management

21. Please describe the behavior management policy being implemented in this district.
    a. How is this policy enforced at your BELL site?
        i. To what extent is there variation in implementation across teachers/TAs in using the behavioral model?
    b. What are the most prevalent disciplinary problems do you face?
        i. How do you address these problems?
    c. What is your assessment of the BELL behavioral model?
        i. Age appropriateness?
        ii. Effectiveness?

Resources/Support

22. Do you feel the BELL summer program provides sufficient funding for the programmatic components?
23. What additional resources or support are needed?
24. What are your biggest challenges as a BELL summer program lead teacher?

Thoughts on BELL Model

25. What is your overall assessment of the BELL programmatic structure and philosophy? (Probe: congruence/disconnect between concept and implementation)
26. How is BELL doing in its efforts to close the summer learning loss gap?
    a. Why?

Advice

27. What advice would you give to other lead teachers to be successful in this role?
28. What advice would you give to other lead teachers in terms of having a good working relationship with the program manager and teachers?
29. What advice would you give to BELL in terms of the overall program model?

Interview Protocol for Regional Leaders

"Thank you for agreeing to take part in the BELL Summer Program Implementation Study interview. Your participation will help BELL and the program funders understand how the program is being implemented. You do not have to answer any question that makes you feel uncomfortable and may terminate or leave the interview at any point. If you complete only part of the discussion, please note that MDRC may use whatever information was collected about you before that point.

You will not be identified in any of the papers that are written from this interview. Note, however, if keeping your answers confidential would put you or someone else in danger, then we will have to tell the appropriate agencies in order to protect you or the other person. Your comments may be repeated or quoted in some documents, but the names of students, teachers, and administrators will not be used in published reports.

The interview will last approximately 60-90 minutes.

Do you give permission for me to type or write notes during the interview? Notes prepared from this interview will not include any identifying information such as your name.

Additionally, do you give us permission to record the interview? The recordings are for the interviewers' use only and will be stored securely until they are reviewed to confirm the accuracy of our notes, after which time they will be destroyed. Also, they will not be shared with anyone outside of MDRC."

<u>Administrator Background</u>

1. How did you end up coming to work in the BELL program?
    a. What were you doing before you came to BELL? How long have you been working with BELL?
    b. Previous experience in K-12 education; ties to local school system and community
2. What is your role at BELL?
    a. Probe: Has your role changed over time?

<u>Program local context</u>

3. What are the district's goals for its summer programming?
4. Please confirm the criteria used for identifying students who are eligible to attend the BELL summer program.
    a. How did you select students to participate in the BELL program in this area?

        i. Any mandatory students?
        ii. What was the level of interest among students eligible for the program (difficulty to recruit vs. waiting list)?
5. Is BELL the only school district sponsored academic summer program for 5th-7th grade students?
    a. If not, what are the following characteristic of the those programs:
        i. Length
        ii. Curriculum
        iii. Topics covered
        iv. Staff qualifications
        v. Key differences from BELL program
6. Are you aware of any non-district sponsored academic summer programs offered in the area?
    a. If so, what are the following characteristic of the those programs:
        i. Length
        ii. Curriculum
        iii. Topics covered
        iv. Staff qualifications
        v. Key differences from BELL program
7. Are the BELL students very similar or different from the average [insert locality name] middle school student?
    a. In what ways besides testing below grade level?
8. Do the BELL students in [insert locality name] face unique challenges relative to other 5th-7th graders?
    a. If so, what are they?
9. How does the BELL program interface with the school district (central and school level staff) in the administration of the summer program?
    a. Does BELL meet regularly with a district representative?
        i. If so, Who?
        ii. How often?
        iii. What is typically discussed at these meetings?
    b. Does the district and BELL have defined roles in the administration of the summer program?
10. If so, how are roles divided? Does the BELL program face any challenges specific to this region?
    a. If so, what are they?
11. We'd like to discuss the staff selection process you used to select site directors and teachers. What process did you use?
    a. How would you assess the process?

12. What are the components of the training that BELL offers the following staff:
    a. Field directors and other region manager level staff
    b. Program managers
    c. Lead teachers
    d. Teachers
    e. Teachers Assistants
13. What is your assessment of the staff training BELL provided? (Probe: utility, scope, length)
14. How would you describe your management style and how do you typically communicate with staff?
    a. Regular meetings?
    b. Electronic communication?
    c. Topics of interaction — most freq, next most, etc.
15. Do you see management differences among program managers in this district that may be important in terms of program implementation?
16. What is your assessment of your relationship with your staff?
    a. Examples?
17. Do you use a more centralized or decentralized approach?
    a. Examples?
18. What is your assessment of the general BELL organizational/structural model? (National staff/administration, field director, program, lead teacher, teacher, TA)
    a. Is it an effective model for this locality?
19. What is your assessment on the program's ability to cover all needed activities with the resources that are allocated for the program in the budget?
    a. How do you or other BELL leaders make decisions around the budget?

<u>Program Characteristics</u>

20. How does BELL implement and manage all of the components of the summer program?
    a. Please discuss roles and responsibilities of key staff for each of the following program components:
        - Community Time
        - Classroom set up/management
        - Behavioral management
        - Teaching and learning: Academics
        - Teaching and learning: Enrichment
        - Parent engagement
    b. How do you monitor what's going on in the program?
21. Are there other key program components that we have not discussed? If so, how are they implemented and managed?

22. What types of enrichment courses are offered?
    a. How were the enrichment courses selected?
    b. How were teachers recruited for teaching enrichment courses?
        i. Recruited to teach specific enrichment classes or recruited to teach one of multiple options?
23. How were students placed into enrichment courses?
    a. What is your assessment of this course selection process? [In districts where students choose the enrichment courses] what happens when there is extra demand?
        i. How frequently were kids not put in their first choices?
24. What is the teacher/TA arrangement for academic and enrichment instruction?
    a. Do TAs stay with the students or teachers during the academic classes?
    b. Do TAs stay with the students or teachers during the enrichment classes?
25. Please describe BELL's attendance policy in this district for 5th-7th graders being implemented in this district.
    a. How is this policy enforced at the site level?
    b. How do is this policy communicated to parents and students?
26. Please describe the behavior management policy being implemented in this district.
    a. How is this policy enforced at the site level?
        i. To what extent is there variation in implementation across teachers/TAs in using the behavioral model?
    b. What types of disciplinary problems does BELL face?
        i. How does BELL address these problems?
    c. What is your assessment of the BELL behavioral model? (Probe: age appropriateness/effectiveness)
27. How do the 5th-7th grade students experience the broader community through the BELL summer program?
    a. What was your level of involvement with the local community prior to Bell?
    b. Did you already have existing community ties or do you have to develop them after taking this position at BELL?
28. How does the BELL program offer positive adult role models and mentors?
29. How does BELL offer opportunities for students to experience success?

<u>Quality and Fidelity to Program Model</u>

30. How has each program component been implemented so far?
    a. How does BELL evaluate program quality and fidelity to the model?
    b. How is the data they gather used?
    c. How has this district's TA arrangement been working so far?
        i. Would you change it going forward?
    d. How has the academic instruction been going?

        i. Would you change it going forward?
    e. How has the enrichment implementation been going?
        i. Would you change it going forward?
    f. How has community time/community engagement been going?
        i. Would you change it going forward?

31. Have any of the summer program components not been implemented as planned?
    a. If so, why?
    b. How have these changes affected how the BELL program operates?
    c. Are there components that need to be enhanced?
        i. If so, what are they?
32. Are instructors and TAs across all of the sites consistently implementing the curriculum? What are the differences among the sites in program implementation?
33. Are enrichment providers across all of the sites consistently implementing enrichment activities? What are the differences in enrichment implementation across sites?
34. How is BELL doing in its efforts to close the summer learning loss gap?
    a. Why?

<u>Advice</u>

35. What advice would you give to other program administrators in terms of operating the BELL summer program?
    a. Probe: What advice would you give to other program administrators related to having a good working relationship with the school district?
36. What advice would you give to BELL in terms of the overall program model?

Interview Protocol for School District Liaisons

"Thank you for agreeing to take part in the BELL Summer Program Implementation Study interview. Your participation will help BELL and the program funders understand how the program is being implemented. You do not have to answer any question that makes you feel uncomfortable and may terminate or leave the interview at any point. If you complete only part of the discussion, please note that MDRC may use whatever information was collected about you before that point.

You will not be identified in any of the papers that are written from this interview. Note, however, if keeping your answers confidential would put you or someone else in danger, then we will have to tell the appropriate agencies in order to protect you or the other person. Your comments may be repeated or quoted in some documents, but the names of students, teachers, and administrators will not be used in published reports.

The interview will last approximately 60 minutes.

Notes prepared from this interview would not include any identifying information such as your name. Do you give permission for me to type or write notes during the interview?

We would like to record this interview. The recordings are for the interviewers' use only and will be stored securely until they transcribed and reviewed to confirm the accuracy of our notes, after which time they will be destroyed. Also, they will not be shared with anyone outside of MDRC. Do you give us permission to record the interview?"

<u>Administrator Background</u>

1. How long have you been working for [insert school district]? What role do you play for the district in planning and operating summer programming?
    a. Has your role in the district changed over time? If so, how?
2. Besides the role you play for the summer program, what other roles do you currently play in the district?
3. What percentage of your time is devoted to the summer program?

<u>Local Program/District Context</u>

4. What are the district's goals for its summer programming?
5. How does the district set priorities for which students to serve through its summer programming?
6. Please confirm the criteria used for identifying students who are eligible to attend the BELL summer program.
7. Is BELL the only school district sponsored academic summer program for 5th-7th grade students?

a. If not, what are the following characteristic of the those programs:
    i. Length
    ii. Curriculum
    iii. Topics covered
    iv. Staff qualifications
    v. Key differences from BELL program

8. Are you aware of any non-district sponsored academic summer programs offered in the area?
    a. If so, what are the following characteristic of the those programs:
        i. Length
        ii. Curriculum
        iii. Topics covered
        iv. Staff qualifications
        v. Key differences from BELL program

<u>District Relationship with and assessment of BELL</u>

9. How did the school district select BELL as the summer program provider? Why?
10. What is the structure of the relationship between the district and BELL in terms of planning for and operating summer programming?
    a. How involved are you or any of our colleagues at the district in the following activities related to the BELL summer program:
        i. Curriculum development
        ii. Selection
        iii. Staff training
        iv. Day-to-day program operations
11. What are your thoughts on BELL's program implementation for the targeted students?
    a. Do you think BELL has been successful? If so, in what ways?
12. How does the district assess BELL's success in achieving the districts goals for summer programming?
    a. How would you compare BELL to other summer programs?

<u>Advice</u>

13. What advice would you give other districts that are considering partnering with an outside summer program provider?
    a. What have been the greatest lessons you've learned through working with BELL?
    b. What advice would you give school districts considering working with BELL for their summer programming?

Questionnaire for Teachers (Focus Group)

"Thank you for agreeing to take part in the BELL Summer Program Implementation Study focus group/interview. Your participation will help BELL and the program funders understand how the program is being implemented. You do not have to answer any question that makes you feel uncomfortable and may terminate or leave the focus group/interview at any point. If you complete only part of the discussion, please note that MDRC may use whatever information was collected about you before that point.

You will not be identified in any of the papers that are written from this focus group/interview. Note, however, if keeping your answers confidential would put you or someone else in danger, then we will have to tell the appropriate agencies in order to protect you or the other person. Your comments may be repeated or quoted in some documents, but the names of students, teachers, and administrators will not be used in published reports. Additionally, we ask that you respect your fellow participants and keep our conversation today confidential.

The focus group/interview will last approximately 60 minutes. At the end of the focus group/interview, you will receive $50 to compensate you for your time

Do you give permission for me to type or write notes during the focus group/interview? Notes prepared from this interview will not include any identifying information such as your name.

Additionally, do you give us permission to record the focus group/interview? The recordings are for the interviewers' use only and will be stored securely until they are reviewed to confirm the accuracy of our notes, after which time they will be destroyed. Also, they will not be shared with anyone outside of MDRC."

Introduction

Thank you for participating in this focus group. Our names are _____ and ____ and we are part of the MDRC team that is evaluating the BELL summer program. We'd like to get your perspectives on a few key areas within the BELL summer program involving the program implementation, as well as progress and challenges you've experienced so far. The findings from the information you provide us today will be used to develop a report about the BELL summer program nationally — that we will share with BELL national and others interested in summer programs. However, everything you say here will be kept confidential; nothing you say will be attributed to you by name. To ensure that we capture what you say correctly we will tape record this interview; however, again we will not identify anyone by name. Do you have questions before we begin?

First, we'd like to have everyone introduce themselves and their role within the BELL summer program (e.g., how long they've served as a BELL summer instructor; any other involvement with BELL; academic background and training). Also, please tell us about your job during the regular academic year.

Background — each teacher answers these individually

1. As you introduce yourself, please provide us with information about your background. For example, what do you do during the regular school year, what is your background in K-12 education more generally and have you taught summer school before for middle school students?
2. Why did you decide to become a BELL summer program teacher?

Training

3. What type of preparation did you receive for teaching in the BELL summer program?
4. How would you assess the training provided by BELL for teachers?
    a. Usefulness, scope, length?
    b. Things you would not change?
    c. Ideas for improvement?

Programmatic Components (ask as appropriate depending upon subject area composition of the focus group)

We'd like to gain your perspective on four specific aspects of the BELL summer program: reading, math, enrichment and parent engagement. So, let's consider each of these individually.

Reading

5. Ho w would you assess the reading component of the BELL program?
    a. Quality/appropriateness of resources
    b. Length of instructional time
6. Are there any aspects of the reading component that really stand out in terms of being very effective?
7. Are there any aspects of the reading component that really stand out as needing to be changed?
    a. Are the curricula addressing the right topics given the needs of your BELL students?
        i. If not, what student needs are not covered by the writing curricula?
8. Do you make adaptations to the curricula?

9. Do you use additional supplemental materials beyond the BELL materials?
    a. If so, what type?
    b. Why?
10. What are some of the primary reading instructional methods you use?
    a. How do students respond to these approaches?
    b. Have you used these methods before?
11. How do you typically use your teaching assistants in the reading class?
12. Have you noticed any patterns in terms of reading learning by particular demographic groups of students (e.g., gender, race, grade level)

Writing

13. How would you assess the writing component of the BELL program?
    b. Quality/appropriateness of resources
    c. Length of instructional time
14. Are there any aspects of the writing component that really stand out in terms of being very effective?
15. Are there any aspects of the writing component that really stand out as needing to be changed?
    a. Are the curricula addressing the right topics given the needs of your BELL students?
        i. If not, what student needs are not covered by the writing curricula?
16. Do you make adaptations to the curricula?
    a. If so, why?
17. Do you use supplemental materials beyond the BELL materials?
    a. If so, what type?
    b. Why?
18. What are some of the primary writing instructional methods you use?
    a. How do students respond to these approaches?
    b. Have you used these methods before?
19. How do you typically use your teaching assistants in the writing class?
20. Have you noticed any patterns in terms of writing learning by particular demographic groups of students (e.g., gender, race, grade level)

Math

21. How would you assess the math component of the BELL program?
    a. Quality/appropriateness of resources?
    b. Length of instructional time?
22. Are there any aspects of the writing component that really stand out in terms of being very effective?

23. Are there any aspects of the writing component that really stand out as needing to be changed?
    a. Are the curricula addressing the right topics given the needs of your BELL students?
        i. If not, what student needs are not covered by the writing curricula?
24. Do you make adaptations to the curricula?
    a. Why?
25. Do you use additional supplemental materials beyond the BELL materials? If so, what type?
26. Are there any aspects of the math area that really stand out in terms of being very effective or needing to be changed?
27. What are some of the primary math instructional methods that you use?
    a. How do students respond to these approaches?
    b. Have you used these methods before?
28. How do you typically use your teaching assistants in math instruction?
29. Have you noticed any patterns in terms of math learning by particular groups of students (e.g., gender, race, grade level)?

<u>Enrichment</u>

30. What types of enrichment classes do you teach/lead? How were these social enrichment activities selected?
31. How would you assess the linkage between enrichment activities and academic instruction?
    a. Most effective?
    b. Least effective?
32. How do the students experience the broader community through the enrichment activities?
33. How would you assess student engagement in enrichment?
    a. What important factors influence student engagement?
        i. Self-selection of courses?
        ii. Parental support?
        iii. Individual motivation?
        iv. Teacher investment?
34. How were enrichment courses selected for this summer?
35. How were students placed into enrichment courses?
    a. What is your assessment of this course placement/selection process?
36. How do you typically use your teaching assistants in this area?
37. Have you noticed any learning patterns by particular groups (e.g., gender, race, grade level)

Parental Engagement

38. What is your assessment of the BELL's efforts to engage parents?
    a. How does it compare to other models of parental engagement that are familiar to you?
39. How have you tried to engage parents so far?
    a. How would you assess parental engagement?
40. Are there any "downsides" to parental engagement?
    a. Are there instances in which parental engagement has not been as helpful as you may have hoped?)
41. If any, what suggestions do you have for improving parental engagement?

Programmatic Relationships

42. How would you describe the management style of your site director?
    a. One word or short descriptive phrase?
    b. Do you feel this style is generally effective with the teachers and TAs at this site?
43. What is your relationship with the TAs?
    a. How do you typically utilize their assistance?
    b. How do you typically communicate with them?
    c. Do you have sufficient interaction/planning time with TAs?
    d. At some BELL sites, TAs remain with the same students all day. At other sites, the TA remains with the same teacher all day. What is your assessment of each approach?
44. What type of relationships do the TAs have with the students? How do the student relate to them?
45. How would you describe your relationships with your students?
    a. Do you have a good sense of the students?
    b. Feel like you receive enough background information to effectively educate them?
46. How would you assess student engagement, overall?
    a. What key factors influence student engagement?
        i. Self-selection of courses?
        ii. Mandatory status?
        iii. Parental support?
        iv. Individual motivation?
        v. Teacher investment?
47. Is your engagement with mandatory students any different than with students who were not mandated to participate in the program?
    a. Do you even know who the mandatory students are?

b. Do you find that classes with a majority of mandatory students are different in some ways from classes with none or only a few?
c. Does the Lead Teacher or program manager provide you with any specialized assistance related to mandatory students?

48. Is your engagement with students who have severe IEPs any different than with other students who do not have severe IEPs?
    a. At what point in the program were you made aware of students in your classes who have severe IEPs?
    b. Do you find that classes with more than a few students with severe IEPs are different in some ways from classes with none or only a few?
    c. Does the Lead Teacher or program manager provide you with any specialized assistance related to students with severe IEPs?

<u>Behavioral management</u>

49. Please describe the behavior management policy being implemented in this district.
    a. How is this policy enforced at your BELL site?
        i. To what extent is there variation in implementation across teachers/TAs in using the behavioral model?
    b. What are the most prevalent disciplinary problems that you face?
        i. How do you address these problems?
    c. What is your assessment of the BELL behavioral model?
        i. Age appropriateness?
        ii. Effectiveness?

<u>Resources/Support</u>

50. Do you feel you are provided with the resources you need from program administrators to be a successful instructor in the BELL summer program?
    a. If not, what additional resources or support do you need?
51. What types of support are provided from the Lead Teacher?
    a. Overall assessment
    b. Suggestions for improvement?
52. What types of support are provided from the Program Manager?
    a. Overall assessment
    b. Suggestions for improvement?
53. What are your biggest challenges as a BELL summer program instructor?
54. For those of you who have worked in summer instructional programs in the past, how does the BELL experience compare? (e.g., another summer program; something distinctive)

Thoughts on BELL Model

55. What is your overall assessment of the BELL programmatic structure and philosophy? (Probe: congruence/disconnect between concept and implementation)
56. How do you think BELL doing in its efforts to close the summer learning loss gap?
    a. Why?
    b. How confident are you that the BELL summer program will have a significant impact on the children enrolled?

Advice

57. Reflecting on your experiences so far, what suggestions would you offer to BELL summer program administrators in terms of program design and implementation?
58. What advice would you give to another colleague if s/he was considering serving as a BELL summer instructor?
59. Any other comments?

Questionnaire for Mentors (Focus Group)

"Thank you for agreeing to take part in the BELL Summer Program Implementation Study focus group/interview. Your participation will help BELL and the program funders understand how the program is being implemented. You do not have to answer any question that makes you feel uncomfortable and may terminate or leave the focus group/interview at any point. If you complete only part of the discussion, please note that MDRC may use whatever information was collected about you before that point.

You will not be identified in any of the papers that are written from this focus group/interview. Note, however, if keeping your answers confidential would put you or someone else in danger, then we will have to tell the appropriate agencies in order to protect you or the other person. Your comments may be repeated or quoted in some documents, but the names of students, teachers, and administrators will not be used in published reports. Additionally, we ask that you respect your fellow participants and keep our conversation today confidential.

The focus group/interview will last approximately 90 minutes. At the end of the focus group/interview, you will receive $50 to compensate you for your time

Do you give permission for me to type or write notes during the focus group/interview? Notes prepared from this interview will not include any identifying information such as your name.

Additionally, do you give us permission to record the focus group/interview? The recordings are for the interviewers' use only and will be stored securely until they are reviewed to confirm the accuracy of our notes, after which time they will be destroyed. Also, they will not be shared with anyone outside of MDRC."

<u>Introduction</u>

Thank you for participating in this focus group. Our names are _____ and ____ and we are part of the MDRC team that is evaluating the BELL summer program. We'd like to get your perspectives on a few key areas within the BELL summer program involving the program implementation, as well as progress and challenges you've experienced so far. The information you provide will be used to develop a _____ that we will share with _____.

To ensure that we capture what you say correctly we will tape record this interview; however, we will not identify anyone by name. Do you have questions before we begin? First, we'd like to have everyone introduce themselves and their role within the BELL summer program (e.g., how long they've served as a BELL summer TA; any other

involvement with BELL; academic background and training). Also, please tell us about your job during the regular academic year.

Background

1. As you introduce yourself, please provide us with information about your background. For example, what is your background in K-12 education more generally?
2. Why did you decide to become a BELL summer program TA and what type of preparation did you receive? (probe: assessment of training, ideas for improvement)

Training

3. What type of preparation did you receive for teaching in the BELL summer program?
4. How would you assess the training provided by BELL for teachers?
    a. Usefulness, scope, length?
    b. Things you would not change?
    c. Ideas for improvement?

Programmatic Components (ask as appropriate depending upon subject area composition of group)

To gain your perspective on four specific aspects of the BELL summer program: reading, math, social enrichment and parent engagement. So, let's consider each of these individually.

Reading

5. How would you assess the reading component of the BELL program?
    a. Quality/appropriateness of resources
    b. Length of instructional time
6. Are there any aspects of reading component that really stands out in terms of being very effective?
7. Are there any aspects of the reading component that really stand out as needing to be changed?
8. Describe the types of tasks you complete as a TA?

Writing

9. How would you assess the writing area within BELL? (probe: quality/appropriateness of resources; length of instructional time)
10. Are there any aspects of the writing area that really stand out in terms of being very effective?

11. Are there any aspects of the writing component that really stand out as needing to be changed?
12. Describe the types of tasks you complete as a TA.

Math

13. How would you assess the "math" area within BELL? (probe: quality/appropriateness of resources; length of instructional time)
14. Are there any aspects of the math area that really stand out in terms of being very effective?
15. Are there any aspects of the math component that really stand out as needing to be changed?
16. Describe the types of tasks you complete as a TA.

Enrichment

17. What types of enrichment activities/classes have you assisted with?
    a. How were these social enrichment activities selected?
18. How would you assess the linkage between social enrichment activities and academic instruction?
    a. Most effective?
    b. Least effective?
19. How do the students experience the broader community through the enrichment activities?
20. Describe the types of tasks you complete as a TA.

Parental Engagement

21. What is your assessment of the BELL's efforts to engage parents? How does it compare to other models of parental engagement that are familiar to you?
22. How have you tried to engage parents so far? How would you assess parental engagement?
23. Are there any "downsides" to parental engagement? (Probe: Are there instances in which parental engagement has not been as helpful as you may have hoped?)
24. What suggestions do you have for improving parental engagement?

Programmatic relationships and structure

25. How would you describe the management style of your site director?
    a. One word or short descriptive phrase.
    b. Do you feel this style is generally effective with teachers and TAs?
26. What is your relationship like with the teachers?
    a. How do you typically communicate with them?
    b. Do you have sufficient interaction/planning time with teachers?

c. At some BELL sites, TAs remain with the same students all day. At other sites, the TA remains with the same teacher all day. What is your assessment of each approach?

27. How would you describe your relationships with your students?
    a. Do you have a good sense of the students?
    b. Feel like you receive enough background information to effectively educate them?
28. How would you assess student engagement, overall?
    a. What key factors influence student engagement?
        i. Self-selection of courses?
        ii. Parental support?
        iii. Individual motivation?
        iv. Teacher investment?

Behavioral management

29. Please describe the behavior management policy being implemented in this district.
    a. How is this policy enforced at your BELL site?
        i. To what extent is there variation in implementation across teachers/TAs in using the behavioral model?
    b. What are the most prevalent disciplinary problems do you face?
        i. How do you address these problems?
    c. What is your assessment of the BELL behavioral model?
        i. Age appropriateness?
        ii. Effectiveness?

Resources/Support

30. Do you feel you are provided with the resources you need from program administrators to be a successful TA in the BELL summer program?
31. What types of support are provided from the lead TA?
    a. Overall assessment?
    b. Suggestions?
32. What types of support are provided from the Program Manager?
    a. Overall assessment
    b. Suggestions for improvement?
33. What additional resources or support do you need?
34. What are your biggest challenges as a BELL summer program TA?
35. For those of you who have worked in summer instructional programs in the past, how does the BELL experience compare? (e.g., another summer program; something distinctive)
36. What is your overall assessment of the BELL programmatic structure and philosophy? (Probe: congruence/disconnect between concept and implementation)

Advice

37. Reflecting on your experiences so far, what suggestions would you offer to BELL summer program administrators in terms of program design and implementation?
38. What advice would you give to another colleague if s/he was considering serving as a BELL summer TA?
39. How confident are you that the BELL summer program will have a significant impact on the children enrolled?
40. Any other comments?

Appendix E

Sensitivity Analyses

Appendix E provides the results of sensitivity analyses that were conducted to verify that the estimates of program impacts presented in this report are unbiased and that they can be interpreted as the effect of the Building Educated Leaders for Life (BELL) program in summer 2012. The first sensitivity analysis compares impact estimates that are *adjusted* for students' baseline characteristics and impact estimates that are *not adjusted* for students' baseline characteristics. The second sensitivity analysis examines the impact of BELL in District C, excluding the random assignment blocks that have large differential response rates between BELL and non-BELL students on the fall 2012 survey and testing. As in the analyses presented in Chapters 2 and 3, the three study districts (Districts A, B, and C) are weighted equally in all the tables in this appendix; therefore, the pooled results should be interpreted as the findings for the average study district.

Unadjusted Impact Estimates

The statistical model that was used to estimate impacts (Appendix A) controls for several measures of students' baseline characteristics and prior achievement. In theory, it is not strictly necessary to control for baseline characteristics, because random assignment *should* ensure that students in the BELL group and those in the non-BELL group are similar at baseline with respect to their observed and unobserved characteristics. In the BELL evaluation, however, the main impact analysis controls for students' baseline characteristics, for two reasons:

1. By including highly predictive student characteristics in the model, it is possible to improve the precision of the impact estimates.

2. As discussed in Appendix C, there is no systematic difference in the baseline characteristics of BELL and non-BELL students in any of the three study districts. However, the What Works Clearinghouse recommends that when differences on any baseline characteristic are larger than 0.05 standard deviation in magnitude, the analysis should control for baseline characteristics to help reduce possible bias arising from preexisting differences in student characteristics.[1] As shown in Appendix C, some baseline differences are larger than 0.05 standard deviation in magnitude.

In order to examine whether the main impact findings are sensitive to controlling for students' baseline characteristics, the statistical model was reestimated *without* controlling for

[1]What Works Clearinghouse (2014).

them.[2] Appendix Table E.1 presents impact estimates that are adjusted for students' baseline characteristics; these are the main impact findings from the evaluation. The table also shows impact estimates that are adjusted for blocking only (that are not adjusted for students' characteristics). Finally, the table also shows the standard error of these impact estimates.

As shown in the table, controlling for students' characteristics does not affect the study's general conclusions; impacts are not statistically significant on either math or reading scores. However, adjusting for students' baseline characteristics does make the magnitude of the pooled impact findings slightly larger: It increases the effect size for reading from –0.05 to 0.01, and it increases the effect size for math from 0.02 to 0.07.

This change is due to the fact that controlling for baseline characteristics affects the impact estimates in District A. Appendix C notes that, in District A, there is no systematic difference in the baseline characteristics of BELL and non-BELL students (Appendix Table C.8). Yet students in the BELL group did have statistically and substantially lower state test scores at baseline than students in the non-BELL group (effect size in reading = 0.33 standard deviation). When this baseline difference is not accounted for in the analysis, the estimated impact of BELL in District A is negative, because the effect of the program is confounded with prior differences in achievement. This bias is removed when the impact model controls for students' baseline state test scores, thereby producing estimates of program impacts that are more likely to be causally valid. Had student-level controls *not* been included in the statistical model, the analysis would have produced a downward-biased estimate of program impacts in District A (that is, an impact estimate that would have been too small or negative). As it turns out, the baseline difference in state test scores in District C is mainly due to one random assignment block. Therefore, this block was dropped from the analysis sample as a further sensitivity test. Dropping this block produces impact estimates for District A that are not appreciably different from the impact findings presented in the report. This confirms that the student-level covariates included in the impact model protect the findings against downward bias resulting from baseline achievement differences between BELL and non-BELL students in District A.

More generally, it is also worth noting that, as expected, controlling for students' baseline characteristics improves the precision of the impact estimates; that is, it reduces the standard error.

[2]These sensitivity tests still include random assignment blocks as fixed effects, in order to account for the way in which random assignment was conducted.

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table E.1

Estimated Impacts on Fall Student Outcomes, Adjusted and Unadjusted for Student Characteristics: Fall 2012 Analysis Sample

Outcome	Adjusted for Blocking and Full Set of Student Characteristics[a]			Adjusted for Blocking Only		
	Estimated Impact (S.E.)	Effect Size	P-Value for Estimated Impact	Estimated Impact (S.E.)	Effect Size	P-Value for Estimated Impact
Average across districts						
Reading achievement (standard score)[b]	0.1	0.01	0.929	-0.7	-0.05	0.555
	(0.8)			(1.1)		
Math achievement (standard score)[b]	0.9	0.07	0.286	0.2	0.02	0.854
	(0.9)			(1.2)		
Sample size	919			919		
District A						
Reading achievement (standard score)[b]	-0.6	-0.04	0.604	-3.2 *	-0.26	0.055
	(1.1)			(1.6)		
Math achievement (standard score)[b]	-1.0	-0.08	0.409	-4.6 **	-0.36	0.008
	(1.2)			(1.7)		
Sample size	358			358		
District B						
Reading achievement (standard score)[b]	2.2	0.18	0.225	2.3	0.19	0.388
	(1.8)			(2.7)		
Math achievement (standard score)[b]	3.1	0.24	0.151	4.0	0.31	0.165
	(2.1)			(2.9)		
Sample size	117			117		
District C						
Reading achievement (standard score)[b]	-1.5 *	-0.12	0.070	-1.2	-0.10	0.352
	(0.8)			(1.3)		
Math achievement (standard score)[b]	0.8	0.06	0.417	1.3	0.10	0.335
	(1.0)			(1.4)		
Sample size	444			444		

(continued)

Appendix Table E.1 (continued)

SOURCES: MDRC calculations based on the GRADE and GMADE assessments and the student survey administered in fall 2012.

NOTES: The analyses reported in this table are based on the sample of students who took the GRADE and GMADE assessments and who responded to the student survey in fall 2012 (Fall 2012 Analysis Sample). All estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012. Each of the three study study districts is given an equal weight when estimating the "average across districts" results reported in this table. Rounding may cause slight discrepancies in calculating sums and differences. "S.E" indicates standard error, given in parentheses.

A two-tailed t-test was applied to differences between research groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome measure for students in the Fall 2012 Analysis Sample who are in the non-BELL group.

[a]Estimated impacts are adjusted for blocking and the following variables: a student's score on state reading and math tests taken in spring 2012, whether a student has an individualized education plan (IEP), whether the student has English as a Second Language (ESL), whether a student is eligible for free or reduced-price lunch, parent education, race/ethnicity, and gender, as well as missing data indicators for each covariate.

[b]Students enrolled in fifth grade in spring 2012 were given Level 5 of the GRADE and GMADE; students in sixth grade were given Level 6; and students in seventh grade were given Level M. The national average for GRADE and GMADE standard scores is 100, and the standard deviation is 15.

Impact Findings Excluding the Random Assignment Blocks in District C That Have Large Differences in Response Rates

As discussed in Appendix C, response rates during the fall 2012 data collection in District C are statistically and substantially higher for students in the BELL group than for students in the non-BELL group (difference = 8 percentage points). This difference in response rates does not, however, affect the balance between the BELL and non-BELL groups with respect to baseline characteristics, which means that impact estimates for District C are likely to be unbiased.

To further verify that the impact estimates for District C are not biased by any unobserved preexisting differences between BELL and non-BELL students, an additional sensitivity analysis was conducted. This sensitivity analysis consists of identifying and excluding the two random assignment blocks in District C that have the largest differences in response rates between BELL and non-BELL students. As shown in Appendix Table E.2, excluding these two blocks is sufficient to reduce the differential attrition to 5.7 percentage points, which is considered to be "low attrition" by the What Works Clearinghouse.[3] The characteristics of BELL and non-BELL students in this "restricted" sample are also still balanced at baseline, as shown in Appendix Table E.3.

Appendix Table E.4 presents the impact findings for the restricted sample. In District C, estimated impacts on reading and math scores based on the "restricted" sample are very similar to the impact estimates based on the entire analysis sample. (For example, the effect size for reading is –0.11 in the restricted sample and –0.12 in the full analysis sample, and both estimates are statistically significant.) This indicates that the differential response rates between BELL and non-BELL students in District C are not biasing the impact findings for this district.

[3]The average attrition rate among all students in the restricted sample is 13.5 percent. Given this level of overall attrition, differential attrition must be less than 6 percentage points for the study to be classified as having "low attrition" (What Works Clearinghouse, 2014, p. 13). As shown in Appendix Table E.2, the differential response rate for the restricted sample (5.7 percentage points) is lower than this threshold value.

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table E.2

Response Rate by Data Source in District C, Excluding Random Assignment Blocks with Large Differential Response Rates

Data Source (%)	BELL Group	Non-BELL Group	Estimated Difference	P-Value for Estimated Difference
Fall 2012 testing				
GRADE assessment	88.7	83.0	5.7 *	0.074
GMADE assessment	88.7	83.0	5.7 *	0.074
Fall 2012 student survey	87.9	83.5	4.4	0.170
Fall 2012 Analysis Sample[a]	87.1	82.9	4.2	0.182
Sample size (N = 474)	248	226		

SOURCES: MDRC calculations based on the GRADE and GMADE assessments administered in fall 2012 and the student survey administered in fall 2012.

NOTES: The analyses reported in this table are based on the sample of students who applied to the BELL middle school program and were recruited into the study (study sample), excluding students in the two random assignment blocks in District C with the largest differential response rates. Estimated differences between the BELL group and the non-BELL group are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012. The values in the column labeled "BELL Group" are the observed means for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-BELL group, using the observed distribution of the BELL group across random assignment blocks as the basis for the adjustment. A two-tailed t-test was used to test differences between the BELL and non-BELL groups. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to differences between research groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]The Fall 2012 Analysis Sample includes students in the study sample who took the GRADE and GMADE assessments and who completed the fall 2012 student survey.

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table E.3

Baseline Characteristics of Students in the Fall 2012 Analysis Sample, by Treatment Group in District C, Excluding Random Assignment Blocks with Large Differential Response Rates

Characteristic in Spring 2012	BELL Group	Non-BELL Group	Estimated Difference	Effect Size	P-Value for Estimated Difference
Grade level (%)					NA
Rising into grade 7	53.2	53.2	0.0	0.0	
Rising into grade 8	46.8	46.8	0.0	0.0	
Race/ethnicity (%)					0.700
Hispanic	21.6	26.9	-5.3	-0.11	
Black, non-Hispanic	60.6	57.7	2.8	0.06	
White, non-Hispanic	10.3	9.8	0.5	0.02	
Asian	0.0	0.0	0.0	0.00	
Other	7.5	5.6	1.9	0.07	
Female (%)	38.8	48.1	-9.3 *	-0.19	0.084
Eligible for free/reduced-price lunch (%)	84.3	90.7	-6.5 **	-0.22	0.050
English as a Second Language (%)	6.0	12.3	-6.3 **	-0.20	0.026
Parent education level (%)[a]					0.926
Did not finish high school	14.7	13.0	1.7	0.05	
Has high school diploma or GED certificate	24.0	25.0	-1.0	-0.02	
Completed some postsecondary education	33.3	35.2	-1.9	-0.04	
Has bachelor's degree or higher	19.1	19.0	0.1	0.00	
Other	8.8	7.8	1.0	0.04	
Has an individualized education plan or IEP (%)	31.9	25.8	6.2	0.15	0.181
Proficient on state test (%)[b]					
Reading	23.8	22.5	1.3	0.03	0.758
Math	46.7	39.7	7.1	0.14	0.147
State test scores[c]					
Reading	348.0	347.5	0.5	0.09	0.401
Math	352.2	351.7	0.5	0.08	0.394
Joint test of difference between groups[d] ($\chi 2 = 19.3$)					0.373
Sample size[e] (N = 406)	216	190			

(continued)

Appendix Table E.3 (continued)

SOURCES: MDRC calculations based on the BELL baseline intake form administered in spring 2012 and student records obtained from school districts.

NOTES: The analyses reported in this table are based on the sample of students who took the GRADE and GMADE assessments and who responded to the student survey in fall 2012 (Fall 2012 Analysis Sample), excluding in the two random assignment blocks in District C with the largest differential response rates. Estimated differences between the BELL group and the non-BELL group in this table are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012. The values in the column labeled "BELL Group" are the observed means for students randomly assigned to the BELL group. The "Non-BELL Group" values in the next column are the regression-adjusted means for students randomly assigned to the non-BELL group, using the observed distribution of the BELL group across random assignment blocks as the basis for the adjustment.

Effect sizes are calculated by dividing the difference by the standard deviation of the characteristic for students in the Fall 2012 Analysis Sample who are in the non-BELL group.

A two-tailed t-test was applied to differences between research groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Rounding may cause slight discrepancies in calculating sums and differences.

[a]For students with two guardians, this is the maximum education level of the two guardians.

[b]A student's proficiency is based on the standards in the state where he or she is attending school.

[c]The scale of the test is the one used by the state.

[d]A chi-square test was used to determine whether there is a systematic difference between the BELL group and the non-BELL group at baseline, based on the characteristics included in this table as well as indicators of missing data for all relevant student characteristics.

[e]Due to missing values, the number of students included varies by characteristic. The sample size reported here is for the full Fall 2012 Analysis Sample. The percentage of missing data on any given characteristic does not exceed 7 percent.

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table E.4

Impacts on Academic Achievement for the Fall 2012 Analysis Sample, Excluding Random Assignment Blocks with Large Differential Response Rates

	Fall 2012 Analysis Sample			Excluded Blocks[a]		
Outcome	Estimated Impact	Effect Size	P-Value	Estimated Impact	Effect Size	P-Value
District C						
Reading achievement (standard score)[b]	-1.5 *	-0.12	0.070	-1.4 *	-0.11	0.098
Math achievement (standard score)[b]	0.8	0.06	0.417	0.3	0.03	0.729
Sample size	444			406		
All districts						
Reading achievement (standard score)[b]	0.1	0.01	0.929	0.1	0.01	0.906
Math achievement (standard score)[b]	0.9	0.07	0.286	0.8	0.06	0.367
Sample size	919			881		

SOURCES: MDRC calculations based on the GRADE and GMADE assessments administered in fall 2012.

NOTES: The analyses reported in this table are based on the sample of students who took the GRADE and GMADE assessments and who responded to the student survey in fall 2012 (Fall 2012 Analysis Sample). Estimated impacts are regression-adjusted using ordinary least squares, controlling for the blocking of random assignment by school and grade level in spring 2012, as well as random differences between the BELL and non-BELL groups with respect to the following variables: students' score on state reading and math tests taken in spring 2012, whether a student has an individualized education plan (IEP), whether the student has English as a second language (ESL), whether a student is eligible for free and reduced price lunch, parent education, race/ethnicity, and gender. Each of the three study study districts is given an equal weight when estimating the "All districts" results reported in this table.

Effect sizes are calculated by dividing the impact estimate by the standard deviation of the outcome measure for students in the Fall 2012 Analysis Sample who are in the non-BELL group.

A two-tailed t-test was applied to differences between BELL and non-BELL groups. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]Students enrolled in fifth grade in spring 2012 were given Level 5 of the GRADE and GMADE; students in sixth grade were given Level 6; and students in seventh grade were given Level M. The national average for GRADE and GMADE standard scores is 100, and the standard deviation is 15.

[b]The two random assignment blocks in District C with the largest differential response rates are excluded from the analysis.

Appendix F

Characteristics of the Study Districts and the Nonstudy Districts

As explained in Chapter 1, the three study districts that participated in this evaluation were new partnerships for Building Educated Leaders for Life (BELL) in summer 2012, and they were operating voluntary (not mandatory) programs. Thus, an important question is whether the findings from this study can be generalized to the seven nonstudy middle school districts where BELL operated its middle school program in summer 2012 — especially districts that were more experienced with the program and/or that operated mandatory summer programs. To inform this question, Appendix F presents the results from three analyses that compare the characteristics of the study districts with those of the nonstudy districts. Based on these analyses, the two groups of sites appear to have had similar characteristics in 2012. This does not guarantee that the study's findings are generalizable to nonstudy sites, however, because the two groups of sites might differ in unobserved ways that could affect the magnitude of program impacts.

Gains in Stanford Test Scores in Study Districts and in Other Districts

As explained in the report, BELL administers a reading and math diagnostic test to all students at the start and at the end of the program. In summer 2012, BELL used the Stanford Diagnostic Reading Test and the Stanford Diagnostic Math Test. The same test form was administered in both test sittings, and so the change from pretest to posttest scores may overestimate true growth in student achievement; that is, students may have performed better on the posttest because they remembered questions from the pretest. Yet the change in Stanford test scores can still be used to compare the progress made by students across different school districts.

Thus, the first analysis in this appendix compares the pretest-to-posttest change in scores on this test in summer 2012 for the three study districts (Districts A, B, and C) and the change in scores in three of BELL's more "mature" middle school sites (Districts D, E, and F). Importantly, Districts D, E, and F had partnered with BELL in summers before 2012, and one of them was operating a mandatory program.

As shown in Appendix Table F.1, reading gains in Districts A and B (new sites) are about the same as gains in District E (a more mature site with a mandatory program). Similarly, math gains in all three study districts are similar to those in District D (also a mature site). This suggests that perhaps BELL's impact in the three study districts can be generalized to its effect in other BELL middle school sites that were more experienced and/or that operated mandatory programs in summer 2012.[1]

[1]Also note that the pattern of test score *gains* in the three study districts mirrors the pattern of *impacts:* Both gains and impacts are largest in District B and smallest in District C.

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table F.1

Gains on Stanford Diagnostic Test:
Study Districts Compared with Other BELL Middle School Districts

	Number of Students	Pretest Score (NCE)	Posttest Score (NCE)	Gains (NCE)	Gains (Effect Size)
Reading					
Study districts					
District A	180	41	46	+5	0.24
District B	21	33	40	+6	0.29
District C	126	39	43	+4	0.17
Other BELL middle school districts					
District D	189	26	37	+11	0.52
District E	149	39	44	+5	0.25
District F	410	40	42	+2	0.08
Math					
Study districts					
District A	180	41	47	+6	0.28
District B	21	33	40	+7	0.32
District C	126	35	42	+7	0.32
Other BELL middle school districts					
District D	192	27	33	+6	0.30
District E	167	26	35	+9	0.41
District F	429	36	37	+1	0.06

SOURCE: MDRC calculations based on the Stanford diagnostic assessment given to BELL students at the beginning and end of the BELL program.

NOTES: Test scores are scaled as normal curve equivalents (NCEs). Effect sizes are based on a standard deviation of 21.06, which is the standard deviation for NCEs. Gains for the study districts are based on the subset of students in the Fall 2012 Analysis Sample for whom Stanford test data are available both before and after testing (pretest and posttest).

Characteristics of Students in the Fall 2012 Analysis Sample and All Middle School Students Served by BELL

An important question related to external validity is whether the middle school students who were served by the three study districts are representative of the students served by BELL nationally in summer 2012. Appendix Table F.2 examines this question by comparing the demographic characteristics of BELL students in the Fall 2012 Analysis Sample and the characteristics of middle school students that BELL served nationally in summer 2012. As shown, the majority of students in both groups were either black or Hispanic. (Among students in the study, 78 percent are black or Hispanic, compared with 73 percent of BELL middle school students nationally.) This suggests that, in terms of their demographic characteristics, students in the analysis sample are representative of the students served by BELL nationally in summer 2012.

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table F.2

Characteristics of BELL Students in the Average Study District Compared with Students in BELL Middle School Programs Nationally in Summer 2012

Characteristic in Spring 2012 (%)	Fall 2012 Analysis Sample (BELL Group)	BELL National
By grade level		
Rising into grade 6	19.6	21.1
Rising into grade 7	41.6	24.8
Rising into grade 8	38.8	26.6
Rising into grade 9	--	27.5
By race/ethnicity		
Hispanic	33.9	22.5
Black, non-Hispanic	44.1	50.5
White, non-Hispanic	6.2	4.5
Asian	8.6	4.5
Other	7.2	18.0
Female	43.0	41.2

SOURCES: MDRC calculations based on the BELL baseline intake form administered in spring 2012 and BELL calculations based on its administrative data.

NOTE: The means in the "analysis sample" column are based on all BELL students in the Fall 2012 Analysis Sample (N = 585).

Characteristics of the Study Districts and All School Districts Where BELL Operated a Middle School Program

Another question related to external validity is whether the three study districts operated in a different local context than BELL's other middle school programs in summer 2012. This question can be examined by comparing the characteristics of the three study districts and the characteristics of the seven nonstudy districts where BELL operated its middle school program in summer 2012. Because the purpose of this analysis is to examine the context in which the program was operated, these comparisons are based on all middle schools and all students in the relevant districts, not just on the schools and students where the program was implemented in summer 2012.

Appendix Table F.3 presents the findings. As shown, the characteristics of the three study districts are fairly similar to those of the nonstudy districts. Both groups of districts have a very high percentage of Title I middle schools (about 92 percent), and these schools are located

The Evaluation of Building Educated Leaders for Life (BELL)

Appendix Table F.3

Characteristics of the Study Districts and Other BELL Middle School Districts in Summer 2012

District Characteristic in 2011-2012	Study Districts	Nonstudy Districts	
U.S. Region			
Northeast	0.0	57.1	**
Southeast	66.7	14.3	
West	33.3	14.3	
Midwest	0.0	14.3	
Average characteristics of schools			
Title I status (%)	92.9	91.7	
Location (%)			
City	90.5	97.3	
Town	0.0	0.0	
Rural	9.5	2.7	
School enrollment	702.3	610.7	
Pupil-staff ratio	16.8	16.0	
Average characteristics of students			
Race/ethnicity (%)			
Hispanic	30.9	35.7	
Black, non-Hispanic	26.7	42.8	
White, non-Hispanic	28.3	14.7	
Asian	11.0	5.0	
Other	3.0	1.8	
Female (%)	48.2	49.3	
Eligible for free or reduced-price lunch (%)	58.7	69.0	
Number of districts	3	7	

SOURCE: MDRC calculations based on the Common Core of Data (2011-2012).

NOTES: The values in this table represent the mean characteristics of the average school district in the relevant target population; districts are weighted equally when calculating these means. Averages are based on all schools in these districts (and not just the schools where the BELL program was operated).

A two-tailed t-test was used to test whether the characteristics of the study districts are different from the characteristics of nonstudy districts. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

primarily in cities (not towns or rural areas). Both groups of districts have high middle school enrollment (on average, 600 to 700 students per school) and a high percentage of minority students (72 percent in study districts and 85 percent in nonstudy districts). Both groups of districts have many students who are eligible for free or reduced-price lunch (59 percent in the study districts and 69 percent in nonstudy districts). The primary difference is that most nonstudy districts (57 percent) are located in the Northeast, whereas none of the study districts are located there. With the exception of this last characteristic, none of the differences between study districts and nonstudy districts is statistically significant.

References

Balfanz, Robert. 2009. *Putting Middle Grades Students on the Path to Graduation: A Policy and Practice Brief.* Westerville, OH: National Middle School Association.

Balfanz, Robert, Liza Herzog, and Douglas J. Mac Iver. 2007. "Preventing Student Disengagement and Keeping Students on the Graduation Path in Urban Middle-Grades Schools: Early Identification and Effective Interventions." *Educational Psychologist* 42, 4: 223-235.

Beckett, Megan, Geoffrey Borman, Jeffrey Capizzano, Danette Parsley, Steven Ross, Allan Schirm, and Jessica Taylor. 2009. "Structuring Out-of-School Time to Improve Academic Achievement: A Practice Guide." NCEE 2009-012. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Bloom, Howard. 2006. "The Core Analytics of Randomized Experiments for Social Research." MDRC Working Paper on Research Methodology. New York: MDRC.

Chaplin, Duncan, and Jeffrey Capizzano. 2006. *Impacts of a Summer Learning Program: A Random Assignment Study of Building Educated Leaders for Life (BELL).* Washington, DC: Urban Institute.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cooper, Harris, Kelly Charlton, Jeff C. Valentine, and Laura Muhlenbruck with Geoffrey D. Borman. 2000. "Making the Most of Summer School: A Meta-Analytic and Narrative Review." *Monographs of the Society for Research in Child Development* 65, 1: i-vi and 1-127.

Downey, Douglas B., Paul T. von Hippel, and Beckett A. Broh. 2004. "Are Schools the Great Equalizer? Cognitive Inequality During the Summer Months and the School Year." *American Sociological Review* 69, 5: 613-635.

Eccles, Jacqueline S. 1999. "The Development of Children Ages 6 to 14." *Future of Children* 9, 2: 30-44.

Eccles, Jacqueline S., and Allan Wigfield. 2000. "Schooling's Influences on Motivation and Achievement." Chapter 5 in S. Danziger and J. Waldfogel (eds.), *Securing the Future: Investing in Children from Birth to College.* New York: Russell Sage Foundation.

Entwistle, Doris R., and Karl L. Alexander. 1992. "Summer Setback: Race, Poverty, School Composition, and Mathematics Achievement in the First Two Years of School." *American Sociological Review* 57, 1: 72-84.

Harter, Susan. 1998. "The Development of Self-Representation." Pages 553-618 in W. Damon and N. Eisenberg (eds.), *Handbook of Child Psychology: Social, Emotional, and Personality Development* (5th ed., vol. 3e). New York: Wiley.

Herrera, Carla, Jean Baldwin Grossman, and Leigh L. Linden. 2013. *Staying on Track: Testing Higher Achievement's Long-Term Impact on Academic Outcomes and High School Choice.* Philadelphia: Public/Private Ventures. Website: http://www.mdrc.org/sites/default/files/staying_on_track_testing_higher_achievement.pdf.

Herrera, Carla, Leigh L. Linden, Amy J. A. Arbreton, and Jean Baldwin Grossman. 2011. *Summer Snapshot: Exploring the Impact of Higher Achievement's Year-Round Out-of-School-Time Program on Summer Learning.* Philadelphia: Public/Private Ventures. Website: http://www.wallacefoundation.org/knowledge-center/summer-and-extended-learning-time/summer-learning/Documents/Summer-Snapshot-Exploring-the-Impact-of-Higher-Achievement.pdf.

Heyns, Barbara. 1978. *Summer Learning and the Effects of Schooling.* New York: Academic Press.

Hill, Caroline, Howard Bloom, Alison Rebeck Black, and Mark W. Lipsey. 2007. "Empirical Benchmarks for Interpreting Effect Sizes in Research." MDRC Working Paper on Research Methodology: New York: MDRC.

Holcomb-McCoy, Cheryl. 2007. "Transitioning to High School: Issues and Challenges for African-American Students." *Professional School Counseling* 10, 3: 253-260.

Jacob, Brian A., and Lars Lefgren. 2004. "Remedial Education and Student Achievement: A Regression Discontinuity Analysis." *Review of Economics and Statistics* 86, 1: 226-244.

Juvonen, Jaana, Vi-Nhuan Le, Tessa Kaganoff, Catherine Augustine, and Louay Constant. 2004. *Focus on the Wonder Years: Challenges Facing the American Middle School.* Santa Monica, CA: RAND Education.

Kieffer, Michael J., and William H. Marinell. 2012. *Navigating the Middle Grades: Evidence from New York City.* New York: Research Alliance for NYC Schools.

Kim, James S., and David M. Quinn. 2013. "The Effects of Summer Reading on Low-Income Children's Literacy Achievement from Kindergarten to Grade 8: A Meta-Analysis of Classroom and Home Interventions." *Review of Educational Research* 83, 3: 386-431. Published online April 25; DOI: 10.3102/0034654313483906.

Levin, Henry, Clive Belfield, Peter Muenning, and Cecilia Rouse. 2007. *The Costs and Benefits of an Excellent Education for All of America's Children.* New York: Columbia University, Teachers College.

Mariano, Louis T., and Paco Martorell. 2013. "The Academic Effects of Summer Instruction and Retention in New York City." *Educational Evaluation and Policy Analysis* 35, 1: 96-117.

Matsudaira, Jordan D. 2008. "Mandatory Summer School and Student Achievement." *Journal of Econometrics* 142, 2: 829-850.

Murnane, Richard J., and John B. Willett. 2010. *Methods Matter: Improving Causal Inference in Educational and Social Science Research.* Oxford, UK: Oxford University Press.

National Research Council and Institute of Medicine. 2004. *Engaging Schools: Fostering High School Students' Motivation to Learn.* Committee on Increasing High School Students' Engagement and Motivation to Learn; Board on Children, Youth, and Families; Division of Behavioral and Social Sciences and Education. Washington, DC: National Academies Press.

Pearson Education. 2001. *GRADE: Administration, Scoring and Interpretation (Level 5).* San Antonio, TX: Pearson.

Pearson Education. 2004. *GMADE: Administration, Scoring and Interpretation (Level 5).* San Antonio, TX: Pearson.

Puma, Michael J., Robert B. Olsen, Stephen H. Bell, and Cristofer Price. 2009. "What to Do When Data Are Missing in Group Randomized Controlled Trials." NCEE 2009-0049. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Raudenbush, Stephen W. 2009. "Adaptive Centering with Random Effects: An Alternative to the Fixed Effects Model for Studying Time-Varying Treatments in School Settings." *Journal of Education, Finance and Policy* 4, 4: 468-491.

Reyes, Olga, Karen L. Gillock, Kimberly Kobus, and Bernadette Sanchez. 2000. "A Longitudinal Examination of the Transition into Senior High School for Adolescents from Urban, Low-Income Status, and Predominantly Minority Backgrounds." *American Journal of Community Psychology* 28, 4: 519-544.

Roderick, Melissa R. 1995. "School Transitions and School Dropout." In K. Wong (ed.), *Advances in Educational Policy.* Connecticut: JAI Press.

Skinner, Ellen A., Carrie Furrer, Gwen Marchand, and Thomas Kindermann. 2008. "Engagement and Disaffection in the Classroom: Part of a Larger Motivational Dynamic?" *Journal of Educational Psychology* 100, 4: 765-781.

Skinner, Ellen A., James G. Wellborn, and James P. Connell. 1990. "What It Takes to Do Well in School and Whether I've Got It: The Role of Perceived Control in Children's Engagement and School Achievement." *Journal of Educational Psychology* 82, 1: 22-32.

Sloan McCombs, Jennifer, Catherine H. Augustine, Heather L. Schwartz, Susan J. Bodilly, Brian Mcinnis, Dahlia S. Lichter, and Amanda Brown Cross. 2011. *Making Summer Count: How Summer Programs Can Boost Children's Learning.* Santa Monica, CA: RAND Corporation.

Terzian, Mary, Kristin Anderson Moore, and Kathleen Hamilton. 2009. *Effective and Promising Summer Learning Programs and Approaches for Economically-Disadvantaged Children and Youth.* A White Paper Commissioned by the Wallace Foundation. Washington, DC: Child Trends.

Wellborn, James Guy. 1992. "Engaged and Disaffected Action: The Conceptualization and Measurement of Motivation in the Academic Domain." Unpublished doctoral dissertation. Rochester, NY: University of Rochester, Department of Psychology.

What Works Clearinghouse. 2014. *WWC Procedures and Standards Handbook* (Version 3.0). Washington, DC: U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse.

About MDRC

MDRC is a nonprofit, nonpartisan social and education policy research organization dedicated to learning what works to improve the well-being of low-income people. Through its research and the active communication of its findings, MDRC seeks to enhance the effectiveness of social and education policies and programs.

Founded in 1974 and located in New York City and Oakland, California, MDRC is best known for mounting rigorous, large-scale, real-world tests of new and existing policies and programs. Its projects are a mix of demonstrations (field tests of promising new program approaches) and evaluations of ongoing government and community initiatives. MDRC's staff bring an unusual combination of research and organizational experience to their work, providing expertise on the latest in qualitative and quantitative methods and on program design, development, implementation, and management. MDRC seeks to learn not just whether a program is effective but also how and why the program's effects occur. In addition, it tries to place each project's findings in the broader context of related research — in order to build knowledge about what works across the social and education policy fields. MDRC's findings, lessons, and best practices are proactively shared with a broad audience in the policy and practitioner community as well as with the general public and the media.

Over the years, MDRC has brought its unique approach to an ever-growing range of policy areas and target populations. Once known primarily for evaluations of state welfare-to-work programs, today MDRC is also studying public school reforms, employment programs for ex-offenders and people with disabilities, and programs to help low-income students succeed in college. MDRC's projects are organized into five areas:

- Promoting Family Well-Being and Children's Development
- Improving Public Education
- Raising Academic Achievement and Persistence in College
- Supporting Low-Wage Workers and Communities
- Overcoming Barriers to Employment

Working in almost every state, all of the nation's largest cities, and Canada and the United Kingdom, MDRC conducts its projects in partnership with national, state, and local governments, public school systems, community organizations, and numerous private philanthropies.