AMERICAN INSTITUTES FOR RESEARCH®

I3 BARR Validation Study

Impact Findings: Cohorts 1 and 2

Trisha H. Borman, Johannes M. Bos, Brenna C. O'Brien, So Jung Park, and Feng Liu

JANUARY 2017

I3 BARR Validation Study
Impact Findings: Cohorts 1 and 2

January 2017

Trisha H. Borman, PhD
Johannes M. Bos, PhD
Brenna C. O'Brien, PhD
So Jung Park, PhD
Feng Liu, PhD

AMERICAN INSTITUTES FOR RESEARCH®

1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
202.403.5000

www.air.org

8022_01/17

Contents

Executive Summary

The Building Assets, Reducing Risks (BARR) model is a comprehensive, strength-based approach to education that aims to improve achievement for all students by improving a school's effectiveness at building relationships, leveraging real-time student data, and capitalizing on the strengths of each student. The U.S. Department of Education's Investing in Innovation (i3) program provided the BARR program developers with a validation grant, and American Institutes for Research (AIR) is conducting a program evaluation as part of this grant.

The BARR model divides the incoming ninth grade into distinct groups of students (blocks) who share the same teachers for at least three of their core subjects (English language arts [ELA], mathematics, science, and/or social studies). The teachers of these different subjects work together as a team to promote their students' success. In doing so, they build on the assets that students bring to school and address the academic and nonacademic risks they face as they progress through ninth grade.

The impact evaluation of the BARR model is a within-school randomized control trial (RCT). Individual ninth-grade students in 11 high schools in Maine, California, Minnesota, Kentucky, and Texas were randomly assigned either to implement BARR and receive BARR supports during their ninth-grade year or to not implement the program and receive these supports. When completed, the validation study will follow three cohorts of high schools participating in the program in three sequential school years (2014–17).

This report provides findings from the first two years of the evaluation, including a total of six schools implementing the BARR model with support from the i3 validation grant: three schools in 2014–15 (Cohort 1) and three schools in 2015–16 (Cohort 2). For these two cohorts, we found that assignment to BARR improved students' academic outcomes at the end of ninth grade and their school experiences during ninth grade.

In summary, the main impact findings for Cohorts 1 and 2 are as follows:

- BARR had a positive and statistically significant impact on students' reading and math skills as measured with the Northwest Evaluation Association's Measures of Academic Progress (NWEA MAP)[1] assessment. This impact manifested itself both in students' average achievement scores and in the percentage of students who met or exceeded their projected growth on the NWEA assessment during ninth grade.

- BARR also had a positive and statistically significant impact on the percentage of total core credits that students earned in ninth grade and the percentage of students passing all of their core courses with no failures.

[1] Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP) assessments.

- BARR also had statistically significant and positive impacts on student-reported measures of supportive relationships in school, student engagement, and teacher expectations. BARR students also reported receiving more challenging assignments than control students.

In addition to these student-level impact estimates, the study found that BARR teachers had more favorable perceptions of their students' behavior in the classroom than teachers in the control condition had of their students' behavior. BARR teachers also were more willing to collaborate with their colleagues, were more likely to report using data to inform their instruction, and reported greater self-efficacy. However, these teacher-level differences cannot be conclusively attributed to the BARR model because teachers were not randomly assigned to BARR or the control group.

Introduction

The Building Assets, Reducing Risks (BARR) model is a comprehensive, strength-based approach to education that aims to improve achievement for all students by improving a school's effectiveness at building relationships, leveraging real-time student data, and capitalizing on the strengths of each student. The U.S. Department of Education's Investing in Innovation (i3) program provided the BARR program developers with a validation grant that provided funding to bring the BARR model to more high schools around the country. In addition, the validation grant required and funded an independent evaluation of the impact of the BARR model on student outcomes. American Institutes for Research (AIR) is conducting this evaluation in 11 high schools, divided into three annual cohorts. This report presents preliminary findings for the first two of these cohorts.

The impact evaluation of the BARR model is a randomized controlled trial (RCT), which is the most rigorous research design for estimating the impact of a program. Individual ninth-grade students in the 11 high schools in the evaluation were randomly assigned to implement the BARR model and receive supports from the providers or to not implement the BARR model. That is, students in the BARR group experienced the BARR model, and students in the control group experienced "business as usual."

This report presents impact findings from the first and second cohorts of schools to implement the study and the BARR model, in the 2014–15 and 2015–16 academic years. Data from these six schools will be pooled with the data from the five schools in the third and final cohort (2016–17), and final findings will be available after the five-year study has concluded.

This document first provides a brief description of the BARR model, followed by a methods section that describes our analytical approach and the background and outcome measures included in the study. Subsequent sections describe the sample of schools and students and present impact results. We conclude with a brief summary of findings.

The BARR Model

The BARR model is designed to address developmental, academic, and structural challenges facing students and teachers in ninth grade by combining student asset building, teachers' real-time analysis of student data, and intensive teacher collaboration to prevent course failure and improve student academic achievement, motivation, and engagement. The BARR model was developed in St. Louis Park High School in Minneapolis, Minnesota, to address persistent course failure and achievement gap issues in that school. After the initial success of that application, the model was developed further with support from an i3 Development grant. The present validation study thus represents the next step in the implementation and improvement of this high school reform model.

The primary mechanism through which BARR pursues its academic objectives is by developing and fostering better student-teacher and teacher-teacher relationships. In addition, the BARR model seeks to create the structural and organizational conditions necessary to fully integrate student supports into a school's existing model for addressing nonacademic barriers to learning.

Implementation of the BARR model within a school involves restructuring the ninth grade into blocks of three- to four-person teacher teams, which engage in collaborative assessment, problem solving, and planning in weekly block meetings. In each school, these meetings and other components of the BARR model are supported by a designated BARR coordinator whose time commitment ranges from 50% to 100% depending on the size of the school. To support school staff and leaders newly implementing the model, the BARR program developers provide in-situation coaching, phone-based support, quarterly site-to-site mentoring visits, and technology-enabled learning opportunities. They also maintain an ongoing learning community of BARR schools, which come together in semiannual in-person program meetings to share their implementation experiences and challenges.

The BARR model is built around eight specific strategies that address a range of different aspects of teacher effectiveness, teacher practice, and collaboration. These strategies include the following:

Strategy 1: Focus on the Whole Student. The BARR model explicitly focuses on the whole student, not just on a student's performance in a particular subject or his/her specific academic or nonacademic challenges. Thus, teachers and administrators are instructed to identify each student's assets and leverage them in addressing challenges and barriers. Working across multiple core courses makes it easier to identify these assets and to address challenges that manifest themselves differently in different settings.

Strategy 2: Provide Professional Development for Teachers, Counselors, Administrators. Teachers and school administrators receive practical and hands-on training and coaching to improve their communication about students' progress, assets, and barriers as well as their ability to identify and implement necessary interventions to help keep students on track. Much of this professional development is focused on improving communication among teachers, school administrators, support staff, students, and parents.

Strategy 3: Use BARR's I-Time Curriculum to Foster a Climate of Learning. The BARR model includes a weekly I-Time lesson, which is taught by one of the core teachers and explicitly addresses students' social-emotional development and related issues. The I-Time activities specifically aim to improve student-student and student-teacher communication and support mutual understanding and collaboration.

Strategy 4: Create Cohorts of Students. The course schedule is restructured such that distinct groups of students share the same group of teachers for their core subjects. This structure is intended to increase feelings of community and belonging among students and enables their teachers to compare and improve students' academic progress across the different subjects.

Strategy 5: Hold Regular Meetings of the Cohort Teacher Teams. A key feature of the BARR model is weekly block meetings during which the academic progress, assets, and challenges of each student are discussed by their core teachers working together with the BARR coordinator. During these meetings, the team agrees upon any interventions that individual students may need. The implementation and effectiveness of these interventions are discussed and monitored in subsequent meetings.

Strategy 6: Conduct Risk Review Meetings. Students who persistently fail or exhibit major attendance or behavioral problems are referred to risk review meetings, which include counselors, school administrators, and other support staff. These meetings also result in the identification of specific interventions whose success is monitored on an ongoing basis.

Strategy 7: Engage Families in Student Learning. The BARR model includes extensive, ongoing interaction with parents to ensure their continued engagement in their child's education. The BARR program believes that actively involved parents are key to supporting student success, especially when students encounter academic or nonacademic challenges.

Strategy 8: Engage Administrators. The BARR program requires serious and ongoing commitment from school leadership (time, attention, staff resources) and directly involves school administrators in the day-to-day implementation of the model. The model aims to enhance administrators' ability to make decisions, support their teaching staff, and take an active role in their students' academic and nonacademic success.

Together, these strategies are designed to improve the ninth-grade experience for students (e.g., feeling more connected to school, cultivating better relationships with teachers, receiving coordinated support) and teachers (e.g., developing better relationships with colleagues, working collaboratively). If the BARR model works, these improvements in students' short-term outcomes and experiences should then translate into improved midterm outcomes for students (e.g., earning more course credits toward graduation, attaining better test scores, being more engaged in learning) and eventually result in long-term benefits for students (e.g., increased graduation rates, higher college acceptance rates, acceptance to better colleges). This evaluation is designed to capture both the short- and midterm impacts of the program. Future evaluations also will include analyses of the long-term effects of BARR.

Methods

Based on the first two of three study cohorts, this report provides preliminary answers to the following three confirmatory questions about the impact of the BARR model on student outcomes:

1. What is the impact of BARR on ninth-grade students' educational attainment as measured by the percentage of credits completed in three core subjects (e.g., English, mathematics, science)?
2. What is the impact of BARR on ninth-grade students' mean level of mathematics achievement as measured by the NWEA MAP assessment?
3. What is the impact of BARR on ninth-grade students' mean level of reading achievement as measured by the NWEA MAP assessment?

In addition to the three confirmatory outcomes listed here, the evaluation includes impact analyses for a range of exploratory outcomes, measured at the student and teacher levels. These outcomes, some of which are addressed in this report, include student-reported measures of engagement, belonging, and school experiences and teacher-reported measures of student behavior, teacher collaboration, and teacher use of data for instruction. The confirmatory academic outcome measures will drive the evaluation's summary assessment of BARR's effectiveness, and the various exploratory analyses will provide supplemental and contextual evidence.

Sample and Randomization

The six study schools featured in this report include two rural schools in Maine, three suburban schools in California, and one suburban school in Minnesota. Table 1 shows that there was considerable variation in the demographic background characteristics of the students across these six study schools. The three California schools were considerably larger and included much greater percentages of students of color, English language learners (ELLs), and low-income students. Students in the two schools in rural Maine were predominantly white, and only about half of them qualified for free or reduced-price lunch (FRPL). The study school in Minnesota had the lowest rate of low-income students, with fewer than one in five FRPL-eligible.

Table 1. Characteristics of Schools in Cohorts 1 and 2

School	State	Locale	Number of Grade 9 Students	Percent Students of Color	Percentage of English Language Learners	Percentage of Special Education Students	Percentage of Students Eligible for Free or Reduced-Price Lunch
Cohort 1							
School A[a]	Maine	Rural	214	7.0%	< 1%	17.7%[b]	39.2%
School B[c]	California	Suburban	648	94.0%	15.4%	11.2%[d]	89.0%
School C[c]	California	Suburban	453	72.4%	8.3%	16.6%	80.7%
Cohort 2							
School D[a]	Maine	Rural	167	3.5%	0.3%	12.5%[b]	53.4%
School E[e]	California	Suburban	583	93.1%	14.3%	10.7%	80.6%
School F[f]	Minnesota	Suburban	183	25.0%	2.6%	9.6%	19.7%

Sources: [a] Maine School Accountability Report Card (2014–15); [b] Maine Department of Education Data Warehouse (2014–15); [c] California Department of Education Educational Demographics Unit database (2014–15); [d] California School Accountability Report Card (2013–14); [e] Ed-Data Education Data Partnership (2014–15); [f] Minnesota Report Card (2015–16)

In the summer before the school year started, each of these schools provided AIR with a list of incoming ninth-grade students, which we then randomized into two groups. One group was assigned to receive BARR supports, and one was assigned as the "business-as-usual" control group. As detailed earlier, the BARR group of students took their ninth-grade core courses with BARR teachers (who themselves were not randomly assigned) and were included in the block meetings and risk review meetings if necessary. BARR students also participated in I-Time lessons designed to strengthen the relationship between the students and their teachers. Control-group students were eligible to participate in all other school activities and received all usual supports at the school.

After randomization, we determined that the two research groups in the study sample were equivalent in gender, ethnicity, ELL status, special education (SPED) status, and FRPL status.[2]

The total combined sample for Cohorts 1 and 2 included 2,172 ninth-grade students: 981 BARR students and 1,191 control students (Table 2). There are more control students than BARR students in the study sample because capacity constraints in one large high school (School E) were addressed by applying a 1:2 BARR/control random-assignment ratio in that school.

Table 2. Randomized Student Sample for Cohorts 1 and 2

Assigned Sample	BARR	Control	Total
Cohort 1			
School A	104	104	208
School B	321	321	642
School C	180	179	359
Cohort 2			
School D	82	81	163
School E	211	422	633
School F	83	84	167
Total	981	1,191	2,172

Source: AIR sample from school-provided administrative data.

Data Sources

At the beginning of the school year, school staff administered two standardized tests to all ninth-grade students: the NWEA MAP assessment in mathematics and reading. These assessment data served as a baseline measure of academic skill for students in the study and were also used as the basis for calculating projected achievement growth scores.

At the end of the school year, school staff again administered the two NWEA MAP assessments to all ninth-grade students, and these data provide two of the three measures of academic achievement for the confirmatory impact analyses. In addition to these standardized assessments, we collected data on students' core credits earned, which constituted the third confirmatory outcome measure.

We administered student surveys in the spring semester, which were used to assess the effects of BARR on student experiences (e.g., engagement). BARR and control group students received and completed the same surveys. Table 3 summarizes the follow-up data collection efforts, showing the number of BARR and control students in the originally randomized sample from which AIR successfully collected outcome data. The table also shows what percentage of the original sample these students represented.

[2] The difference between the BARR group and control group for each baseline characteristic was less than 0.25 standard deviations, thus meeting the What Works Clearinghouse (WWC) baseline equivalence guidelines.

Overall attrition and differential attrition (between the BARR and control group) were relatively low for the combined Cohort 1 and 2 sample. Overall attrition ranged from 29.5% for the student surveys to 15.8% for the core credit data. (It was 24.9% for the confirmatory NWEA outcomes.) See Appendix A for the sample sizes and attrition rates for each student outcome measure.

Table 3. Student Data Sources by Group (Cohorts 1 and 2)

	BARR *N* = 981			Control *N* = 1,191			Total *N* = 2,172	
Data Source	***n***	**%**		***n***	**%**		***n***	**%**
NWEA Reading scores	740	75.4%		891	74.8%		1,631	75.1%
NWEA Mathematics scores	745	75.9%		886	74.4%		1,631	75.1%
Core credits data	824	84.0%		1,004	84.3%		1,828	84.2%
Student surveys	730	74.4%		801	67.3%		1,531	70.5%

Source: AIR calculations from NWEA scores, school-provided administrative data, and AIR-administered surveys.

To capture the experiences of teachers serving students in the BARR and control groups, we administered a survey at the end of the spring semester to any core-subject teachers who taught ninth-grade students during the school year. Twenty-nine BARR teachers and 37 control teachers completed the teacher surveys, for a total response rate of 69.9%. Note that teachers were not randomized to the BARR or control group. Therefore, observed differences in outcomes between teachers in these two groups do not constitute estimates of program effects.

Analytic Approach

We estimated program effects on all outcomes presented in this report by using an ordinary least squares (OLS) model to compare outcomes for students assigned to the BARR group with outcomes for students assigned to the control group. All impact models included student-level background characteristics (e.g., race, gender, FRPL status, ELL status), a test of prior student achievement (i.e., students' fall NWEA MAP scores), an indicator of a student's assignment to BARR, and a set of dummy variables to control for school effects and variation in the random-assignment ratio across schools. Of note, we imputed missing baseline variables and pretest scores using the dummy variable imputation methods recommended in Puma et al. (2009).

We estimated the impact of BARR on students' academic achievement using scale scores generated by the NWEA from the MAP assessments administered by each school at the end of the spring semester. Each student's RIT (Rasch Unit) score on these assessments represents performance across a series of subtests within the content areas of mathematics and reading. For the purposes of the regression analysis, we standardized the original RIT scores for each student to a mean of 0 and a standard deviation of 1.

In addition to the impacts on the NWEA scale scores, we estimated impacts on the likelihood that students would meet the "projected growth" target connecting the NWEA pretest in the fall and the NWEA posttest in the spring. This is a measure of the expected growth that a typical NWEA test taker would demonstrate during the year. For a student to meet this projected growth target, the change in his or her RIT scores must equal or exceed the NWEA calculated growth projection. This calculation is based on the average growth observed in the latest NWEA norming study for students who had the same starting RIT score.[3] Hence, this is a 0/1 indicator, which is 1 if students meet or exceed their growth target, and 0 otherwise.

We also assessed the program's impact on core credits earned and on passing all core courses during the school year. Although these are important measures of student progress in school, it should be noted that these outcomes can be affected by variability in how teachers assign grades, which may not always reflect a change in student academic achievement.

We examined BARR's impact on credits earned in four core subjects: ELA, mathematics, science, and social studies. Each of the study schools implemented BARR in at least three out of four of these subjects, and it was up to the school which subjects they chose. For the purpose of this analysis, we assumed that students could earn one credit for each quarter or semester in which they completed a core course. We then estimated what percentage of the credits each student earned compared to the credits they attempted, and then examined the differences between the two groups overall and for each core subject area.

As described earlier, we administered a survey to ninth-grade students in each of the BARR and control classrooms toward the end of the school year. These surveys measured six constructs of student experience related to academic achievement and social and emotional outcomes. The findings are presented in the form of scale scores, which were derived mathematically as a summary of responses to multiple survey questions within each construct (see Table 13). The survey scale scores for each student were standardized to a mean of 50 and a standard deviation of 10 for the regression model.

The effect sizes shown in the results tables are the differences between the BARR and control groups that takes into consideration the standard deviation of the outcome measure. Lipsey et al. (2012) explain that effect size is useful because the "standardized form (i.e., representing effects in standard deviation units) allows comparison of the magnitude of effects on different outcome variables and across different studies." Therefore, researchers can compare the effect size from the impact of the BARR model to the impact of similar programs on high school populations, with a possible estimate of ES = .10 or higher as being noteworthy.

Results

As stated earlier, the impact of the BARR model on student achievement was estimated using four outcome measures:

- NWEA reading and mathematics scores
- Projected growth met on NWEA assessments
- Percentage of core course credits earned
- Passing all core courses in ninth grade

[3] Source: https://legacysupport.nwea.org/sites/www.nwea.org/files/resources/AnnotatedReports-MAP_0.pdf, pp. 5–6. Cohort 1 followed 2011 norms, and Cohort 2 followed 2015 norms.

The analyses for each of the student academic outcomes are presented below for the combined sample of Cohort 1 and Cohort 2 students. Discussion about any notable differences by cohort will be included in the final impact report for this study.

NWEA Scores

Table 4 shows that students who were assigned to BARR scored higher, on average, than students in the control group on both the NWEA reading and mathematics assessments. The differences between the groups' achievement scores were modest but statistically significant for both measures. Expressed in terms of effect sizes, the evaluation found positive impacts for both of the measures (ES = .08).

Table 4. Impacts on Standardized Reading and Mathematics Scale Scores (Cohorts 1 and 2)

Student Outcome Measure	*N*	BARR	Control	Difference	Effect Size
Reading NWEA RIT score	1,631	222.81	221.69	1.13*	0.08
Mathematics NWEA RIT score	1,631	231.21	229.74	1.47**	0.08

Source: AIR calculations from school-administered NWEA assessments.
Notes: * = statistically significant at the $p < .05$ level; ** = statistically significant at the $p < .01$ level.

In Table 5, NWEA reading impacts are organized by key demographic subgroups. The table shows positive impact estimates for all key subgroups. However, these impacts were considerably stronger for male students, students of color, and low-income students. Because of the smaller effect sizes and sample sizes, only the impact estimates for male students (ES = .14) and low-income students (ES = .10) were statistically significant.

Table 5. Impacts on Reading NWEA RIT Scores (Subgroups)

Student Subgroup	*N*	BARR	Control	Difference	Effect Size
Female	825	223.88	223.38	0.50	0.04
Male	806	221.88	219.79	2.10**	0.14
Students of color	1,084	220.18	218.92	1.26	0.09
White	547	228.06	227.20	0.86	0.06
Free and reduced-price lunch	1,175	220.50	219.03	1.47*	0.10
Non-free and reduced-price lunch	456	228.88	228.43	0.45	0.03

Source: AIR calculations from school-administered NWEA assessments.
Notes: * = statistically significant at the $p < .05$ level; ** = statistically significant at the $p < .01$ level.

Of particular note, Table 6 reveals a somewhat different pattern of subgroup impacts for NWEA mathematics scores. Although male students similarly benefited more than female students, NWEA mathematics impacts were stronger for white students (ES = 0.13) than for students of color (ES = 0.06) and were also somewhat stronger for students who were not FRPL-eligible (ES = .11). It is possible that these subgroup impact patterns are related to how the different subgroups are distributed across schools. We will further explore this phenomenon in future evaluation reports that include all 11 study schools.

Table 6. Impacts on Mathematics NWEA RIT Scores (Subgroups)

Student Subgroup	*N*	BARR	Control	Difference	Effect Size
Female	828	230.44	229.91	0.53	0.03
Male	803	232.02	229.51	2.51**	0.13
Students of color	1,085	226.54	225.48	1.07	0.06
White	546	240.41	238.13	2.27**	0.13
Free and reduced-price lunch	1,177	227.59	226.21	1.38*	0.08
Non-free and reduced-price lunch	454	240.72	238.78	1.94*	0.11

Source: AIR calculations from school-administered NWEA assessments.
Notes: * = statistically significant at the $p < .05$ level; ** = statistically significant at the $p < .01$ level.

Table 7 shows how these impacts on NWEA reading and mathematics test scores translate into impacts on a projected growth measure associated with these assessments. For a student to meet projected growth, the change in his or her RIT scores from fall to spring needs to equal or exceed NWEA's fall-to-spring growth projection, which is based on the average growth observed in the latest NWEA norming study for students who had the same starting RIT score.[4]

Based on this information, students were assigned a 0/1 indicator for this outcome (i.e., 1 = met or exceeded projected growth, 0 = did not meet projected growth). For Cohorts 1 and 2, the percentage of BARR students who met or exceeded their projected growth in reading and math was greater than their control group counterparts, and these differences were statistically significant, (ES = 0.13 and ES = 0.15, respectively).

Table 7. Impacts on Projected Growth Met (Cohorts 1 and 2)

Student Outcome Measure	*N*	BARR	Control	Difference	Effect Size
Reading projected growth met (%)	1,332	73.3	67.3	6.0*	0.13
Mathematics projected growth met (%)	1,239	78.6	71.7	6.9**	0.15

Source: AIR calculations from school-administered NWEA assessments.
Notes: * = statistically significant at the $p < .05$ level; ** = statistically significant at the $p < .01$ level.

Table 8 shows additional subgroup analyses of the NWEA projected growth measure in reading. As expected, the pattern of subgroup impacts closely mirrors that found for the average NWEA reading scores, with male students and FRPL-eligible students benefiting from BARR the most. Male students were almost 12 percentage points more likely to meet the NWEA projected growth target, and FRPL-eligible students were nearly 8 percentage points more likely to do so.

Table 9 shows that BARR significantly increased the proportion of students who met their projected growth target in mathematics among male students, white students, and students not eligible for FRPL. Again, these estimates mirror those found for the NWEA RIT scores. All other point estimates for the subgroup impacts on this outcome were positive as well, although not statistically significant.

---

[4] Source: https://legacysupport.nwea.org/sites/www.nwea.org/files/resources/AnnotatedReports-MAP_0.pdf, pp. 5–6. Cohort 1 followed 2011 norms, and Cohort 2 followed 2015 norms.

Table 8. Impacts on Reading Projected Growth Met (Subgroups)

Student Subgroup	*N*	BARR	Control	Difference	Effect Size
Female	677	71.2	71.8	-0.6	-0.01
Male	655	79.0	67.4	11.6***	0.25
Students of color	804	80.2	74.7	5.5	0.13
White	528	33.9	28.5	5.5	0.12
Free and reduced-price lunch	902	75.8	68.2	7.6*	0.16
Non-free and reduced-price lunch	421	76.8	74.2	2.6	0.06

Source: AIR calculations from school-administered NWEA assessments.
Notes: * = statistically significant at the p < .05 level; *** = statistically significant at the p < .001 level.

Table 9. Impacts on Mathematics Projected Growth Met (Subgroups)

Student Subgroup	*N*	BARR	Control	Difference	Effect Size
Female	628	79.7	75.3	4.4	0.10
Male	611	82.4	73.6	8.8**	0.20
Students of color	715	79.3	77.0	2.2	0.05
White	524	65.7	49.0	16.7***	0.33
Free and reduced-price lunch	816	74.1	69.4	4.7	0.10
Non-free and reduced-price lunch	413	86.6	77.3	9.4**	0.22

Source: AIR calculations from school-administered NWEA assessments.
Notes: ** = statistically significant at the p < .01 level, *** = statistically significant at the p < .001 level.

Core Credits Earned

Table 10 shows that BARR significantly increased the number of core credits students earned.[5] These outcome variables were calculated as a percentage of possible credits earned out of credits attempted to account for differences across schools in the number of core credits available to earn. Note that for the individual subject areas, statistically significant, positive differences were observed for both ELA and science courses but not for mathematics or social studies courses.

Table 10. Impacts on Students' Core Credits Earned (Cohorts 1 and 2)

Student Outcome Measure	*N*	BARR	Control	Difference	Effect Size
Total credits earned (%)	1,828	84.3	79.0	5.3***	0.05
ELA credits earned (%)	1,822	87.4	76.9	10.5***	0.11
Mathematics credits earned (%)	1,823	79.1	81.3	-2.2	-0.02
Science credits earned (%)	1,661	85.3	77.1	8.2***	0.08
Social studies credits earned (%)	300	98.2	97.8	0.4	0.0

Source: AIR calculations from school-provided administrative data.

[5] During data collection, there was some concern that course passing rates in the control group in one of the six study schools may have been inflated. Because it was impossible to determine the precise extent of this data problem, we did not adjust the impact results or remove the school from the impact analyses. As a result, all impact estimates in this section may be attenuated (biased toward zero). Detailed results of sensitivity analyses will be included in an appendix of the final impact report for this study.

Notes: *** = statistically significant at the $p < .001$ level.

Subgroups. Additional analyses revealed statistically significant impacts on total credits earned for several subgroups of students. As was the case with the NWEA impacts, Table 11 shows that these impacts were statistically significant for male students (ES = .09), students of color (ES = .07), and FRPL-eligible students (ES = .07).

Table 11. Impacts on Students' Core Credits Earned (Subgroups)

Student Subgroup	*N*	BARR	Control	Difference	Effect Size
Female	908	87.4	85.4	2.0	0.02
Male	920	81.3	72.4	8.9***	0.09
Students of color	1,249	81.9	75.1	6.8***	0.07
White	579	89.5	87.8	1.7	0.02
Free and reduced-price lunch	1,346	82.2	75.4	6.8***	0.07
Non-free and reduced-price lunch	482	90.7	88.3	2.4	0.02

Source: AIR calculations from school-provided administrative data.
Notes: *** = statistically significant at the p < .001 level.

Passing All Core Courses

With respect to core credits earned, we also investigated whether BARR had an impact on students passing all of their core courses during the fall and spring semesters without any course failures (i.e., they had no F's on their report cards). For the regression model, this outcome was transformed into log odds coefficients that estimated the probability that a student would meet this criterion (i.e., 1 = passed all core courses and 0 = did not pass all core courses).[6] Table 12 shows that BARR students were significantly more likely to pass all their courses than their control group counterparts (ES = .25).

Table 12. Impacts on Students Passing All Core Courses (Cohorts 1 and 2; Full Sample and Subgroups)

Student Outcome Measure	*N*	BARR	Control	Difference	Effect Size
Passing all core courses	1,828	70.3	57.7	12.6***	0.25
Student subgroup					
Female	908	85.2	79.5	5.7*	0.14
Male	920	68.4	52.0	16.3***	0.33
Students of color	1,249	64.4	47.2	17.2***	0.34
White	579	71.0	69.8	1.1	0.02
Free and reduced-price lunch	1,346	69.5	53.4	16.0***	0.32
Non-free and reduced-price lunch	482	37.8	36.8	1.0	0.02

Source: AIR calculations from school-provided administrative data.

[6] For two schools, students who failed a core course during the first semester had the opportunity to make up that credit during the second semester. Total core credits earned during the school year includes these retakes. However, students who failed a course at any time during the school year are marked as not passing all core courses.

Note: ** = statistically significant at the $p \leq .01$ level.

Subgroups. Additional analyses revealed statistically significant impacts on passing all core courses for most subgroups of students. Reflecting the earlier findings about the percentage of core credits earned, these impacts were strongest for male students (ES = .33), students of color (ES = .34), and FRPL-eligible students (ES = .32). There was also a smaller but statistically significant impact on female students (ES = .14).

Student Social and Emotional Outcomes

To assess impact of the BARR model on students' experiences in ninth grade, we collected data through surveys that measured the following six constructs:

- Expectations and rigor
- Student engagement
- Supportive relationships
- Social and emotional learning
- Sense of belonging
- Grit

Impact analyses for these findings are presented in Table 13. All but one of the impact estimates shown in this table favored the BARR group, with impacts on three constructs reaching statistical significance. Students in the BARR group indicated that they felt more supported by their teachers (ES = 0.34) and felt that their teachers had higher expectations of them (ES = 0.23). BARR students also reported significantly greater engagement (ES = 0.12).

Table 13. Impacts on Student Experiences From Student Survey (Cohorts 1 and 2)

Student-Reported Outcome	*N*	BARR	Control	Difference	Effect Size
Expectations and rigor	1,524	51.21	48.89	2.32***	0.23
Student engagement	1,510	50.64	49.42	1.22*	0.12
Supportive relationships	1,507	51.75	48.39	3.36***	0.34
Social and emotional learning	1,495	50.01	49.99	0.01	0.00
Sense of belonging	1,501	50.23	49.79	0.44	0.04
Grit	1,497	49.81	50.17	-0.36	-0.04

Source: AIR calculations from AIR-administered student surveys.
Notes: * = statistically significant at the $p < .05$ level; *** = statistically significant at the $p < .001$ level.

Subgroups. Additional analyses revealed that for several subgroups of students, impact estimates also were statistically significant for the same three constructs (supportive relationships, expectations and rigor, and student engagement). Of note, these impacts were the largest for students of color: supportive relationships (ES = .43), expectations and rigor (ES = .31), and student engagement (ES = .16). Appendix B provides the subgroup mean scale scores and effect sizes for each student survey construct.

Differences Between Teacher Groups

An important objective of the BARR model is to change how teachers view and interact with their students and with each other. By creating structures and activities to bring teachers together and to deepen teachers' relationships with their students, BARR aims to enhance teacher effectiveness and student engagement.

To examine differences in the experiences of the BARR teachers and the control teachers, we collected data through surveys that measured the following eight constructs:

- View of students' actual behavior, commitment, and attitudes (what teachers observe students doing)
- Perception of student behavior, commitment, and attitudes (what teachers think students would do)
- View of the school's supports provided to teachers
- Interaction with parents
- Teacher self-efficacy
- View of student accountability (e.g., completing work on time)
- Collaboration with and view of colleagues
- Use of data

We administered an online teacher survey at the end of the school year to systematically assess whether teachers in the BARR and control groups reported different experiences and opinions during the study year. Table 14 presents the results from a statistical comparison between the responses of the BARR and control teachers on a range of teacher survey measures, in the form of scale scores. Note that the survey scale scores for each teacher were standardized to a mean of 50 and a standard deviation of 10 for the independent samples *t*-test comparisons.

Table 14. Differences in Teacher Experiences From Teacher Survey (Cohorts 1 and 2)

Teacher-Reported Outcome	*N*	BARR	Control	Difference	Effect Size
View of students' actual behavior	65	52.7	47.8	4.9*	0.05
Perception of student behavior	64	51.9	48.4	3.5	0.04
View of the school's supports	65	52.5	48.0	4.5	0.05
Interaction with parents	65	51.6	48.7	2.9	0.03
Teacher self-efficacy	65	53.0	47.6	5.4*	0.05
View of student accountability	64	49.3	50.6	-1.3	-0.01
Collaboration with and view of colleagues	64	56.3	44.8	11.5***	0.11
Use of data	63	55.5	45.3	10.3***	0.10

Source: AIR calculations from AIR-administered teacher surveys.
Note: * = statistically significant at the $p < .05$ level; *** = statistically significant at the $p < .001$ level.

The table shows statistically significant differences between BARR and control group teachers for four of the eight survey measures. BARR teachers had more positive views about their colleagues and collaborating with them (ES = .11), and they reported greater levels of data use to inform their instruction (ES = .10). BARR teachers also were somewhat more positive about their students' behavior, commitment, and attitudes in the classroom (ES = .05), and they and reported greater self-efficacy (ES = .05) than their control group counterparts.

Note that the effect sizes for the teacher surveys are calculated in a manner similar to those presented in the earlier student-level impact tables. However, these group differences are not impact estimates because teachers were not randomly assigned to BARR or the control group. They could be partially or fully explained by pre-existing differences among the teachers in the two groups and how the teachers were selected for each group before the study began.

Summary

For schools in Cohorts 1 and 2 (the 2014–15 and 2015–16 school years), BARR had modest but statistically significant impacts on student achievement as measured through independent achievement tests. The program significantly reduced ninth-grade course failure rates. Students also reported better relationships with their teachers and greater levels of engagement in the classroom.

Teachers in the BARR model reported more collaboration with their peers and more data use in their instruction compared to teachers in the control group. Data continue to be collected at schools in Cohort 3 (the 2016–17 school year), and the impacts on the full student sample will be available at the conclusion of the validation study.

References

Puma, M. J., Olsen, R. B., Bell, S. H., & Price, C. (2009). *What to do when data are missing in group randomized controlled trials* (NCEE 2009-0049). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/pdf/20090049.pdf

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms* (NCSER 2013-3000). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education. Retrieved from https://ies.ed.gov/ncser/pubs/20133000/pdf/20133000.pdf

Appendix A. Attrition Rates for Outcome Measures

Tables A-1 through A-4 list the overall attrition (i.e., the rate of attrition for the entire sample) and differential attrition (i.e., the difference in the rates of attrition for the BARR and control groups) for the student outcome measures for the combined Cohorts 1 and 2 sample. Overall attrition and differential attrition were low and within acceptable conservative WWC standards for the three confirmatory academic outcomes. The differential attrition for the student survey outcome measure was over the conservative boundary but below the liberal boundary for the WWC standards.

Table A-1. Attrition for NWEA Reading Scores (Cohorts 1 and 2)

Data Source	BARR	Control		Total
Assigned sample	981	1,191		2,172
Analytic sample	740	891		1,631
Overall attrition rate	24.6%	25.2%		24.9%
Differential rate				0.6%

Table A-2. Attrition for NWEA Mathematics Scores (Cohorts 1 and 2)

Data Source	BARR	Control		Total
Assigned sample	981	1,191		2,172
Analytic sample	745	886		1,631
Overall attrition rate	24.1%	25.6%		24.9%
Differential rate				1.6%

Table A-3. Attrition for Core Credits Earned (Cohorts 1 and 2)

Data Source	BARR	Control		Total
Assigned sample	981	1,191		2,172
Analytic sample	824	1,004		1,828
Overall attrition rate	16.0%	15.7%		15.8%
Differential rate				0.3%

Table A-4. Attrition for Student Surveys (Cohorts 1 and 2)

Data Source	BARR	Control		Total
Assigned sample	981	1,191		2,172
Analytic sample	730	801		1,531
Overall attrition rate	25.6%	32.7%		29.5%
Differential rate				7.1%

Appendix B. Student Social and Emotional Outcomes by Subgroup

To collect a measure of student experience to compare across groups, a student survey was administered with six constructs. The findings from the student surveys for the combined sample of Cohorts 1 and 2 are presented in Tables B-1 through B-6, disaggregated into six student subgroups: females, males, students of color, white students, FRPL-eligible students, and students not eligible for FRPL.

Table B-1. Impacts on Student Experiences From Student Survey (Females)

Student-Reported Outcome	*N*	BARR	Control	Difference	Effect Size
Expectations and rigor	771	50.8	48.9	1.9 **	0.20
Student engagement	766	51.0	49.6	1.4 *	0.13
Supportive relationships	766	51.8	48.5	3.2 ***	0.32
Social and emotional learning	762	49.9	49.2	0.7	0.07
Sense of belonging	765	49.6	48.4	1.2	0.12
Grit	762	50.2	50.6	-0.4	-0.04

Source: AIR calculations from AIR-administered student surveys.
Notes: * = statistically significant at the $p < .05$ level; ** = statistically significant at the $p < .01$ level, *** = statistically significant at the $p < .001$ level.

Table B-2. Impacts on Student Experiences From Student Survey (Males)

Student-Reported Outcome	*N*	BARR	Control	Difference	Effect Size
Expectations and rigor	753	51.6	48.9	2.7 ***	0.28
Student engagement	744	50.3	49.2	1.1	0.11
Supportive relationships	741	51.8	48.2	3.6 ***	0.36
Social and emotional learning	733	50.0	51.0	-1.0	-0.10
Sense of belonging	736	50.8	51.4	-0.6	-0.06
Grit	735	49.3	49.8	-0.5	-0.05

Source: AIR calculations from AIR-administered student surveys.
Notes: *** = statistically significant at the $p < .001$ level.

Table B-3. Impacts on Student Experiences From Student Survey (Students of Color)

Student-Reported Outcome	*N*	BARR	Control	Difference	Effect Size
Expectations and rigor	1,010	51.4	48.3	3.1 ***	0.31
Student engagement	1,001	50.9	49.2	1.6 **	0.16
Supportive relationships	1,002	51.6	47.4	4.2 ***	0.43
Social and emotional learning	991	50.8	50.6	0.2	0.02
Sense of belonging	998	50.3	50.1	0.2	0.02
Grit	995	49.7	50.1	-0.4	-0.04

Source: AIR calculations from AIR-administered student surveys.
Notes: ** = statistically significant at the $p < .01$ level, *** = statistically significant at the $p < .001$ level.

Table B-4. Impacts on Student Experiences From Student Survey (White Students)

Student-Reported Outcome	*N*	BARR	Control	Difference	Effect Size
Expectations and rigor	514	51.1	50.0	1.0	0.10
Student engagement	509	50.3	49.8	0.5	0.05
Supportive relationships	505	52.2	50.4	1.7	0.17
Social and emotional learning	504	48.3	48.9	-0.5	-0.05
Sense of belonging	503	50.2	49.0	1.1	0.11
Grit	502	50.2	50.2	0.0	0.00

Source: AIR calculations from AIR-administered student surveys.

Table B-5. Impacts on Student Experiences From Student Survey (FRPL-Eligible Students)

Student-Reported Outcome	*N*	BARR	Control	Difference	Effect Size
Expectations and rigor	1,102	50.8	48.2	2.6 ***	0.26
Student engagement	1,092	50.3	49.1	1.2 *	0.12
Supportive relationships	1,092	51.0	47.5	3.5 ***	0.35
Social and emotional learning	1,080	50.3	50.1	0.2	0.02
Sense of belonging	1,088	49.6	49.6	0.0	0.00
Grit	1,087	49.4	49.6	-0.2	-0.02

Source: AIR calculations from AIR-administered student surveys.
Notes: * = statistically significant at the $p < .05$ level; *** = statistically significant at the $p < .001$ level.

Table B-6. Impacts on Student Experiences From Student Survey (Non-FRPL-Eligible Students)

Student-Reported Outcome	*N*	BARR	Control	Difference	Effect Size
Expectations and rigor	421	52.2	50.7	1.5	0.15
Student engagement	418	51.5	50.3	1.1	0.12
Supportive relationships	415	53.4	50.9	2.5*	0.25
Social and emotional learning	415	49.0	49.8	-0.8	-0.08
Sense of belonging	413	51.6	50.6	1.0	0.10
Grit	410	51.1	51.6	-0.5	-0.05

Source: AIR calculations from AIR-administered student surveys.
Notes: * = statistically significant at the $p < .05$ level.

ABOUT AMERICAN INSTITUTES FOR RESEARCH

Established in 1946, with headquarters in Washington, D.C., American Institutes for Research (AIR) is an independent, nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance both domestically and internationally. As one of the largest behavioral and social science research organizations in the world, AIR is committed to empowering communities and institutions with innovative solutions to the most critical challenges in education, health, workforce, and international development.

AIR®
AMERICAN INSTITUTES FOR RESEARCH®

1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
202.403.5000

www.air.org

Making Research Relevant

LOCATIONS

Domestic

Washington, D.C.
Atlanta, GA
Austin, TX
Baltimore, MD
Cayce, SC
Chapel Hill, NC
Chicago, IL
Columbus, OH
Frederick, MD
Honolulu, HI
Indianapolis, IN
Metairie, LA
Naperville, IL
New York, NY
Rockville, MD
Sacramento, CA
San Mateo, CA
Waltham, MA

International

Egypt
Honduras
Ivory Coast
Kyrgyzstan
Liberia
Tajikistan
Zambia

8022_01/17