Michigan Reading Corps Racial Equity Evaluation

Final Report and Recommendations

January 2, 2024

The content of this report does not necessarily reflect the views of the College of Education at Michigan State University or Michigan State University.

Michigan Reading Corps Racial Equity Evaluation

Authors
This study was conducted by the following researchers:

- Sara Witmer, PhD., Department of Counseling, Educational Psychology, and Special Education (CEPSE), Michigan State University
- Ruxin Shi, Center for Statistical Training and Consulting, Michigan State University
- Julia Bachmann, MA, Department of Counseling, Educational Psychology, and Special Education (CEPSE), Michigan State University

Acknowledgements
The evaluation plan, including original research questions and analysis framework applied for this study, were developed by staff from ServeMinnesota and Michigan Education Corps.

- Patrick Kaiser, ServeMinnesota
- Peter Nelson, Ph.D., ServeMinnesota
- Holly Windram, Ph.D., Michigan Education Corps

The data used for the analyses were compiled and provided via collaboration across staff from FastBridge, ServeMinnesota, and Michigan Education Corps.

For citations and reference to this publication, please use the following:

Witmer, S., Shi, R., & Bachmann, J. (2023). *Michigan Reading Corps Racial Equity Evaluation Report*, Michigan State University, East Lansing, MI.

Table of Contents

EXECUTIVE SUMMARY

Michigan Education Corps (MEC) contracted with Michigan State University (MSU) to complete a racial equity evaluation of the Michigan Reading Corps program (MRC). MEC and staff from ServeMinnesota developed the evaluation and analytic plan. The analyses were led by Dr. Sara Witmer with support provided by Ruxin Shi and Julia Bachmann.

Purpose of the Study

The goal of this study was to quantitatively compare several targeted experiences and outcomes of Black or African American and White students who participated in MRC during the 2022-2023 academic year. Prior work has identified positive effects of Reading Corps, and so the current study was intended to carefully analyze whether differences in targeted experiences and outcomes differ among Black or African American and White students. The findings are intended to help inform future work to address any identified racial disparities in experiences and outcomes among student participants in MRC. The individual student experience and outcome variables targeted for quantitative comparison were the following: proportion of intervention sessions attended, end-of-year program status, materials for family engagement sent home, intervention changes, spring reading score, and average weekly growth rate during intervention provision.

Study Design

For the six aforementioned variables of interest, separate multi-level regression models were applied to subsets of Black or African American and White MRC participants who were matched in terms of grade level and fall reading score using a nearest-neighbor matching process. Regression models accounted for school effects (Level 2) and student effects (Level 1), with a focus on the student-level effect of race/ethnicity when other relevant predictor variables were included as covariates in the prediction models. To address the fact that different reading measures were applied at different grade-levels and allow for aggregated analyses across grades, both reading level scores and reading growth scores were transformed into z-scores prior to analyses.

Key Findings

1) For only one of the targeted six variables was race a statistically significant predictor, when controlling for other relevant variables. More specifically, White students were more likely to be exited from the program by the end of the academic year than Black or African American students, when controlling for a variety of other relevant variables.
2) The relationship between race and many of the targeted program experience or outcome variables was found to vary by school site. In each of these situations, the relationship between race and the targeted variable was

lower (i.e., either weaker or in the direction of Black students experiencing more positive experiences or outcomes than White students) in schools with high baseline rates for the targeted experience or outcome variable.

3) Fall reading score and proportion of sessions attended consistently and positively predicted a variety of other variables within the regression models.

Recommendations

To more carefully analyze and seek to address any racial disparities (i.e., Black or African American vs. White) in MRC experiences and outcomes, the researchers offer the following recommendations.

Recommendations for Data Collection

- Consider encouraging and supporting participating schools to collect and make available de-identified student reading scores according to race for all students in the school, using the same measures used for the program. This could help ensure benchmark scores are available for all MRC participants to help facilitate analysis of all program participants, as well as offer an opportunity to better disentangle race-related effects that may correspond to general school programming and those that correspond to MRC programming.
- To better explore any differences in family engagement according to race, identification and use of a better measurement tool for family engagement would be helpful.
- Collecting information on which students experience low-income economic marginalization (LIEM) and incorporating that information in future analyses could help in disentangling the extent to which racial minoritization as opposed to low-income economic marginalization may play a role in any racial disparities identified.

Recommendations for Future Study

- Further consideration of the impact that the differential exit rates of students of different races has on their school experiences and reading growth may be a helpful direction for future work to ensure that the program is facilitating equity in reading and other school-related outcomes.
- To better understand racial disparities in experiences and outcomes that are specifically related to MRC (as opposed to potentially due to differences in more general school programming and timing of intervention provision), a more controlled study involving matching students within school and matching of intervention start time may be necessary. Understandably, identifying close student matches of different races within a single school may not be possible, but to the extent that they are, such analyses are encouraged.
- Consider identifying the schools where race was a strong predictor of experiences and outcomes as well as those where race was a weak predictor of experiences and outcomes. Further exploration of potential factors contributing to those differences may be helpful in developing ideas that may help to reduce racial disparities.
- To the extent that group sizes are large enough, separate analyses by grade level may be helpful to prevent the differences in measurement tools and associated score transformation process that were present in the current analyses from affecting the analyses and associated interpretations.

MICHIGAN READING CORPS RACIAL EQUITY EVALUATION

About the Michigan Reading Corps Program

Background

The Michigan Reading Corps program (MRC) is one of several AmeriCorps programs that is provided through the Michigan Education Corps (MEC). It involves application, within Michigan, of the Reading Corps program that has been developed and implemented in various places across the United States. MEC has been providing intervention to students in Michigan since its pilot year in 2012 and has served over 23,000 students to date. Separate reading, math, and early literacy programs are provided through MEC to facilitate math and reading skill development of elementary students within dozens of partner schools across Michigan.

MRC is based on the Reading Corps program that was first developed and implemented in Minnesota in 2003 and has since been implemented in several other states. MRC offers partner schools from across the state of Michigan tutoring services to accelerate the reading skill development of students in kindergarten through third grade who have been identified as at-risk for poor reading outcomes. AmeriCorps Members are selected, trained, and coached through MEC to provide evidence-based reading skill building activities with high fidelity to eligible students within participating schools. MRC is conceptualized by partner schools as a Tier 2 reading intervention and correspondingly provided to students who are identified as needing targeted intervention support, but not currently in need of intensive individualized interventions.

MRC is intended to promote the critical outcome of basic reading skill proficiency by the end of third grade. Third grade basic reading skill proficiency is an important predictor of long-term academic success, and yet many students, particularly those experiencing low-income economic marginalization, do not meet this important milestone (Hernandez, 2011). Students who do not have reading proficiency by the end of third grade lack the necessary literacy skills to access content in textbooks used in later schooling and correspondingly "read to learn." As such, they are less likely to graduate from high school and less likely to go on to experience employment success (Lesnick et al., 2010). Stanovich (2009) highlights the "Matthew Effects" of early reading skill development, in which a student's weak early literacy skills contributes to decreased literacy-related opportunities such that without appropriate and effective intervention, the gap in literacy skills among such a student and one with strong literacy skills will only increase over time. This occurs despite long-standing information on how to effectively intervene to address reading difficulties in the early elementary grade levels (Snow et al., 1998).

Concerted efforts to improve literacy are clearly necessary based on data indicating students across the nation, and particularly in Michigan, are not reading proficiently by the end of third grade. Data from the National Assessment of Educational Progress (NAEP) indicate that 37 percent of fourth-grade students performed below basic proficiency levels in reading in 2022, which was larger than the corresponding proportion of such students in 2019, with declines in average scores particularly evident among groups of students representing low-performing categories, and particularly low-performing Black students. (U.S. Department of Education, 2022). When disaggregating results by state, Michigan was identified as one of 30 states that showed a decrease in average fourth grade reading test score between 2019 and

2022, and was ranked 43rd overall (compared to 32nd in 2019) based on 2022 NAEP fourth grade reading score. Moreover, Michigan was ranked in the bottom five states with regard to fourth grade reading performance of Black students, with Michigan Black students' average score showing an 11 point decrease from 2019 to 2022, compared to the average fourth grade reading performance decline for White students which was just 4 points in the same time period. Relatedly, MRC seeks to increase the number of students in Michigan who are proficient by third grade through early identification and provision of effective intervention for students from kindergarten to grade; it is particularly interested in examining the extent to which it is specifically addressing the needs of low-performing Black students.

Programming Description

MRC correspondingly engages in a variety of activities to improve basic reading skill development among kindergarten through third graders, including the following:

- Communicating with schools across the state the opportunity to partner with MRC in the delivery of evidence-based tutoring services to promote student reading skill development and working with them to implement MRC.
- Recruiting, training, placing, and supporting AmeriCorps Members to provide 1:1 tutoring services comprised of evidence-based assessment and intervention procedures to kindergarten to third grade students who have not met grade-level benchmarks in reading within partnering elementary school sites.
- Providing a multi-level coaching model including site-based and external coaches who help to ensure AmeriCorps Members are engaging in the associated assessment and intervention procedures with high fidelity across the academic year.

MRC AmeriCorps Members are trained to criterion in specific evidence-based reading assessment and scripted intervention procedures, with systematic observations conducted to ensure they engage in the associated procedures with high fidelity in their provision of daily tutoring sessions to students. Elementary student participants are identified as needing intervention based in part on fall benchmark scores. The intervention strategy(s) used in a given intervention session is/are informed by data on the specific early literacy and basic reading skills that the student has and hasn't yet mastered. Progress is monitored, and decisions to make changes in the intervention strategy(s) provided to an individual student are based on ongoing data collection and analysis. Tutors provide students with practice reading passages and activities to take home for further practice with instructions to have their caregivers sign off on materials that were practiced at home and return to the school for recording purposes. AmeriCorps Members enter a variety of data for each student related to their assessment results, intervention session timing and strategies used, and family materials sent and returned daily.

MRC programming fully supports the General Education Leadership Network Essential Instructional Practices in Early Literacy: Grades K-3 and the Essential School-wide and Center-wide Practices in Literacy.

During the 2022-23 academic year, a total of 64 sites, 119 AmeriCorps Members, and 2114 kindergarten through third grade students were engaged in MRC. Of those students, 509 exited the program by the end of the academic year.

Theory of Change

The MRC theory of change is that providing trained AmeriCorps members and related support to help schools deliver proven academic interventions throughout the key developmental period from kindergarten to third grade will result in students who are on-track to be proficient in reading third grade.

Prior Effectiveness Evidence and Racial Equity Findings

MRC is a direct replication of Reading Corps, which has already been studied and supported via rigorous evaluations in other states, including Minnesota, Washington, and Wisconsin (Markovitz et al., 2014; Markovitz et al., 2018a; Markovitz et al., 2018b; Markovitz et al., 2022).

Based on a corresponding review of evidence, Reading Corps is listed through the Evidence for ESSA website as a program with strong evidence of impact, noting an average effect size estimate of 0.36.

Existing research on the program indicates that Reading Corps is similarly—if not more—beneficial for Black or African American students (Markovitz et al., 2018b). However, the nature of that finding is rooted in between group comparisons (e.g., the effects compared consist of treated v. untreated Black/African American student and treated v., untreated White students) that haven't considered potential differences in implementation activities. Apart from an evaluation study that explored whether race (i.e., Black vs. White) predicted months of growth and found that it was not a significant predictor (e.g., Tetta et al., 2019), existing work hasn't compared the implementation activities and growth rates for Black or African American students and White students who participate in Reading Corps in a particularly systematic manner.

About the Evaluation Study

Given existing evaluation work across time and geographic regions that has demonstrated statistically significant and practically meaningful effects of the Reading Corps program, the current evaluation work was intended to instead explore racial equity (or lack thereof) in program experiences and outcomes for participating students across an academic year. Such work was deemed necessary as important for informing efforts to enhance the program's capacity to address racial disparities in reading outcomes that are evident in the state of Michigan and beyond.

MRC currently is provided across over 50 elementary schools in Michigan who opt to participate. The program engages in recruitment and training of AmeriCorps Members to serve in the participating school sites. Fidelity checks to help ensure and offer documentation that the assessment tools and intervention strategies are implemented similarly and with strong fidelity by AmeriCorps members across all students who participate. Students are identified as eligible for MRC via grade-level Fastbridge benchmark assessment scores and receive 1:1 tutoring from a trained AmeriCorps tutor as a Tier 2 intervention. A specific evidence-based reading intervention strategy (or combination of strategies) is selected to target the specific reading skills the student needs to improve, which is then provided for approximately 20 minutes by the AmeriCorps tutor daily. Progress is monitored weekly during intervention by the AmeriCorps tutor using grade-level specific Fastbridge progress monitoring measures. Intervention change or exit decisions are made based on ongoing review of the associated progress monitoring data. Tutors receive supervision and guidance from internal

coaches at their school sites, who correspondingly receive supervision and guidance from external coaches to support intervention decision-making.

Although training and fidelity help to ensure assessment tools and intervention strategies are implemented in a consistent manner across all students, students' MRC-related experiences may vary considerably in other ways. An understanding of similarities and differences in these MRC experiences according to race was considered a critical first step toward understanding any corresponding race-related differences in outcomes that might exist. Six research questions that correspondingly explored race-related differences across six MRC-related experiences and outcomes were addressed in the current evaluation.

Study Design

The evaluation described here represents a quantitative comparison of the MRC-related experiences and outcomes of Black or African American students and White students who participated in the Michigan Reading Corps program during the 2022-2023 academic year. Following nearest neighbor matching procedures designed to create equivalent Black or African American and White groups based on grade-level and fall benchmark z-scores in reading, multi-level regression models were fit to the associated data to explore the relationship between race and various dependent variables reflecting students' program experiences and outcomes. A multi-level approach was considered necessary to help in accounting for variation that could be considered due to school-level effects (i.e., students nested within schools). Specific dependent variables for which racial differences were explored included the following:

- *Program experience variables*: (1) proportion of intervention sessions attended, (2) end-of-program-year status (i.e., exit, active, stop-stay), (3) intervention changes, (4) number of materials sent home
- *Program outcome variables*: (1) spring benchmark z-score in reading, (2) average weekly growth during intervention

Regression models included covariates, as deemed theoretically appropriate, to allow for meaningful analysis of differences in the dependent variables with a focus on student race as the predictor of interest. At times, multiple models were examined and the model showing best fit/most variance accounted for described to answer the corresponding research question.

Data Collection

To answer the research questions, the researchers obtained the following data files via ServeMinnesota which reflected data collected across the 2022-2023 academic year for all students receiving tutoring services through the Michigan Reading Corps program.

1) **Student demographics**. This dataset included a row for each participating student and included the following variables for each student: a school identification number, a tutor identification number, a student identification number, student grade, student gender, student race/ethnicity, student English learner status, student home language, and the student's latest status in the program.
2) **Assessments.** This dataset included a row for each assessment instance and included the following variables for each assessment instance: the student identification number associated with the assessment instance, the tutor identification number, the assessment date, whether the assessment involved benchmark or progress

monitoring data collection, the assessment tool administered, the correct score, and the error score (if applicable for the given assessment tool administered).

3) **Interventions.** This dataset included a row for each weekday (Monday through Friday) in which an intervention session might have been held, from the day of a student's first intervention session of the year to the day of their last intervention session of the year, and included the following variables for each intervention instance: the student identification number associated with the intervention instance, the school identification number, the tutor identification number, the intervention date, the amount of time the student attended the session, a reason if the student did not attend the session, and the specific intervention or combination of interventions used during the session. A row was included even for each weekday associated with a school break or holiday in which an intervention session was not provided to the student.
4) **Family engagement.** This dataset included a row for each participating student if data were available on family engagement and included the following variables for each student: the number of unique dates in which items were either sent home or returned from home, the total number of items sent home, and the total number of items returned from home.
5) **Fidelity.** This dataset included a row for each instance of a fidelity check, and included the following variables: the school identification number where the fidelity check occurred, the observer identification number, the observer role, the tutor identification number, the check date, the fidelity check type, the fidelity check tool, the linked intervention or assessment, the number of checklist items marked "yes", the number of checklist items marked "no", the maximum number of fidelity items for the given fidelity check and the percent fidelity for the given fidelity check instance.
6) **Average weekly growth.** This dataset included a row for each participating student and included the following variables for each student: the student identification number, the student grade level, and the average weekly growth for each of the tools used for progress monitoring the student during their experience of tutoring (e.g., average weekly growth during tutoring in Letter Sounds, average weekly growth during tutoring in Nonsense Words, average weekly growth during tutoring in Curriculum-Based Measurement-Reading).

Several variables within these files were directly used within the analyses; others were derived from the variables available within these files; all variables included in the analyses are named and described in Appendix A.

It is important to note that different reading assessment tools were used at different grade levels. To run analyses using aggregated data across grade levels as indicated for the evaluation plan, benchmark raw scores were transformed into grade-based z-scores using the raw score data available in the dataset for the given reading assessment tool/grade level/time period of administration (i.e., fall or spring). Similarly, average weekly growth scores during intervention were transformed into grade-based z-scores using the raw data available in the dataset for the given reading assessment tool/grade level. Z-scores were calculated using all available data in the dataset prior to selection of only Black or African American and White students for the purposes of the evaluation. For each grade level analyzed, benchmark and growth was based on different assessment tools from FastBridge as follows: kindergarten = Letter Sound Fluency (LSF), first grade = Nonsense Word Fluency (NWF), second grade = 2nd grade Curriculum-Based Measurement in Reading (CBM-R), third grade = 3rd grade CBM-R.

Student Selection

Following transformation of fall, spring, and growth scores to grade-level z-scores and creation of variables to facilitate analyses (further explained in Appendix A), students identified with any of the following as their race/ethnicity status were removed prior to analysis: American Indian or Alaska Native, Asian, Hispanic/Latino, Multi-Racial, or Unknown. This left only students with race/ethnicity statuses that were the focus of comparative analyses: Black or African American and White.

To facilitate meaningful comparison according to race to address each research question, a nearest neighbor matching process was used to match students of the two targeted race/ethnicities on grade and fall z-score. Only those students with valid values for grade and fall z-score and only those with values for the given dependent variable for a given research question were available to be included in the matching process for a given research question. The results of the matching processes applied are highlighted prior to summarizing the results of the analyses for each research question in the Analysis of Data section below.

As background, of the 2114 total cases, 809 were identified as Black or African American, and 712 were identified as White. All had valid grade levels (k-3) identified in the dataset. Of the Black or African American students, 381 (47% of all Black or African American) had valid fall z-scores. Of the White students, 417 (59% of all White) had valid fall z-scores.

Analysis of Data (including Matching) and Results

General Approach for Each Research Question

Each of the six targeted research questions and six associated dependent variables were addressed in separate analyses described below. For each analysis, matching procedures were first used to identify Black or African American and White student samples with non-missing values for the given dependent variable that were similar in terms of grade-level and fall z-score to permit appropriate comparisons. Next, a multi-variate regression model including the school site at Level 2, and other student-level variables at Level 1, including the predictor variable of interest (i.e., race_eth), the targeted dependent variable, and various covariate variables was identified and described. In several cases, multiple models were initially identified for a particular dependent variable that varied in terms of the covariates included; of those identified, the best fitting model is described below.

Research Question 1: To what extent do students who identify as Black or African American, and White students differ in the proportion of total tutoring sessions attended?

All cases included values for the variable “proportion of sessions attended”. Using a 1:1 nearest neighbor matching process of Black or African American students to White students, matching for grade and fall benchmark z-score yielded a final sample of 316 Black or African America students and 316 White students from 51 school sites. An additional 166 students were unmatched (65 Black or African American students and 101 White students) and were therefore excluded from the analysis, in addition to those mentioned earlier who did not have valid fall z-scores.

The results of the matching procedure (displayed in Table 1 below) produced a sample of Black or African American students and White students that were similar according to grade and fall z-score.

Before matching, the racial groups were different on these characteristics as follows:

- 6% of Black or African American students were in Kindergarten, compared with 9% of White students
- 22% of Black or African American students were in first grade, compared with 27% of White students
- 32% of Black or African American students were in second grade, compared with 36% of White students
- 40% of Black or African American students were in third grade, compared with 27% of White students
- Black students had an average fall z-score of -.17, compared with .09 for White students.

After matching, the racial groups were much more similar as follows:

- 6% of Black or African American students were in Kindergarten, compared with 7% of White students
- 26% of Black or African American students were in firsts grade, compared with 27% of White students
- 35% of Black or African American students were in second grade, compared with 30% of White students
- 33% of Black or African American students were in third grade, compared with 36% of White students
- Black students had an average fall z-score of -.08, compared with .00 for White students.

Table 1. Results of Matching for Student Selection for Research Question 1

Matching variable	Pre-matching		Post-matching	
	Black or African American	White	Black or African American	White
Kindergarten	6%	9%	6%	7%
First Grade	22%	27%	26%	27%
Second Grade	32%	36%	35%	30%
Third Grade	40%	27%	33%	36%
Fall Z-Score	-.17	.09	-.08	.00

Appendix B shows the distributions of the targeted variable “proportion of sessions attended” (i.e., prop_att) for the matched samples; it shows slightly higher values for White as opposed to Black or African American students, but further analysis was considered necessary to understand whether these differences were significant within the context of controlling for other variables via the multi-level regression model.

Research Question 1 was then answered using a generalized linear mixed model from the Gaussian family, given that the data, upon removal of three outliers (1 from the White group and 2 from the Black or African American group), was found to have a non-significant Shapiro-Wilk test result and therefore a normal distribution could be assumed. The model was fit using restricted (residual) maximum likelihood (REML) estimation method with t-tests conducted using the Satterthwaite method; all were conducted within the *R* statistical programming language.

The model took the form as described below.

For the student-level:

$$prop\_att_{ij} = \beta_{0j} + \beta_{1j} \cdot race\_eth_{ij} + \beta_{2j} \cdot Z\_fall_{ij} + \beta_{3j} \cdot grade1_{ij} + \beta_{4j} \cdot grade2_{ij} + \beta_{5j} \cdot grade3_{ij} + e_{ij}$$

For the school-level:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

where:
- γ_{00} = 0.610 and is the overall average of coefficient β_{0j}
- u_{0j} is drawn from a random variable with a normal distribution, having a mean of 0 and a variance of 0.004. Its value varies between different schools.
- γ_{10} = −0.001 and is the overall average of coefficient β_{1j} (NOTE: Black or African American was coded '0' and White was coded '1')
- u_{1j} is drawn from a random variable with a normal distribution, having a mean of 0 and a variance of 0.0006. Its value varies between different schools.
- Furthermore, the correlation between u_{0j} and u_{1j} is -0.46.
- β_{2j} is the estimate for fall z-score, which is 0.012
- β_{3j} is the estimate for first grade, which is 0.008
- β_{4j} is the estimate for second grade, which is -0.005
- β_{5j} is the estimate for third grade, which is -0.016

The coefficients in the regression equation represent the change in the proportion of sessions attended for a one-unit change in the corresponding student-level predictor variables. Additionally, the model includes random effects that account for the unobserved heterogeneity across different school sites identified by the site identification number, which are crucial in capturing variability and improving the model's accuracy and predictive power.

The model's random effects suggested some variability in prop_att across different school sites (N = 51), especially considering variance in intercepts and race_eth slopes. More specifically, the variance for the intercept was .004 and the variance for the race_eth effect within site was .001. This indicates that while the intercept (0.610) contributes significantly to the model, the variation among groups within this component is relatively low. In other words, this intercept demonstrates consistency across different groups. The correlation between intercept and race_eth effect was -.46. Overall, this means that when a school site had a high baseline value for prop_att, the effect related to race_eth was lower, and when the site had a lower baseline value for prop_att, the effect related to race_eth was higher. The strength of this inverse relationship was measured at -.46.

The model's fixed effects are highlighted in Table 2. Cells containing a "*" indicate that a variable was a statistically-significant predictor of prop_att. **As the table shows, the race_eth variable was not a statistically significant predictor of prop_att.** However, the random effects suggested that the influence of race_eth on prop_att varied considerably by school site. In those sites where there is a particularly high overall value for prop_att, Black or African Amercan students tended to have a slightly higher proportion of sessions attended than White students, whereas in those sites where there is a particularly low overall value for prop_att, White students tended to have a slightly higher prop_att than Black or African American students.

In terms of fixed effects examined apart from race_eth, the fall z-score was significantly positively related with proportion of sessions attended, suggesting that those who were higher achieving at the beginning of the year tended to have a higher prop_att. Grade level was not significantly associated with proportion of sessions attended.

Table 2. Fixed effects regression summary for multilevel model associated with Research Question 1 *(proportion of sessions attended as dependent variable)*

Variable	Coefficient (standard error)
(*intercept*)	.610 (.02)***
race_eth	-.001 (.01)
fall z-score	.012 (.00)**
first grade	.008 (.02)
second grade	-.005 (.02)
third grade	-.016 (.02)

Note. "*" denotes statistical significance ($p < .05$). "**" denotes statistical significance ($p < .01$). "***" denotes statistical significance ($p < .001$).

For the sake of interpretation, it is important to note that variance in prop_att could be influenced by many factors. School breaks and holidays, as well as individual student and tutor absences from school would lower a student's value for prop_att. Relatedly, the time of year at which intervention was initiated and ended for a student may have substantially influenced their proportion of sessions attended. Further analysis that considers and incorporates intervention timing may allow for a more direct and sensitive comparison of program attendance according to race_eth.

Research Question 2: To what extent do students differ in the rate at which they exit intervention as a function of race?

Due to missing data for the dependent variable of "bin_exit" (i.e., a binary variable created from the datafile based on latest_status, in which 0 = either "Active" or "Stop-Stay" status at end of the year, 1 = exit status at the end of the year, and students who moved out were coded as "missing"), rematching was needed to address this research question. Using a 1:1 nearest neighbor matching process of Black or African American students to White students, matching for grade and fall z-score yielded a final sample of 300 Black or African America students and 300 White students. An additional 171 students were unmatched (66 Black or African American students and 105 White students) and were therefore excluded from the analysis, in addition to those mentioned earlier who did not have valid fall z-scores or data for bin_exit.

The results of the matching procedure (displayed in Table 3 below) produced a sample of Black or African American students and White students that were similar according to grade and fall z-score.

Before matching, the racial groups were different on these characteristics as follows:

- 5% of Black or African American students were in kindergarten, compared with 9% of White students
- 22% of Black or African American students were in first grade, compared with 27% of White students
- 32% of Black or African American students were in second grade, compared with 37% of White students
- 41% of Black or African American students were in third grade, compared with 27% of White students
- Black students had an average fall z-score of -.18, compared with .11 for White students.

After matching, the racial groups were much more similar as follows:

- 6% of Black or African American students were in Kindergarten, compared with 7% of White students
- 25% of Black or African American students were in first grade, compared with 26% of White students
- 35% of Black or African American students were in second grade, compared with 31% of White students

- 34% of Black or African American students were in third grade, compared with 36% of White students
- Black students had an average fall z-score of -.07, compared with .00 for White students.

Table 3. Results of Matching for Student Selection for Research Question 2

Matching variable	Pre-matching		Post-matching	
	Black or African American	White	Black or African American	White
Kindergarten	5%	9%	6%	7%
First Grade	22%	27%	25%	26%
Second Grade	32%	37%	35%	31%
Third Grade	41%	27%	34%	36%
Fall Z-Score	-.18	.11	-.07	.00

Appendix B shows the distributions of the targeted variable "bin_exit" for the matched samples; it shows that 75% of the White students (within the matched sample) had exited whereas 66% of the Black students (within the matched sample) had exited. However, further analysis was considered necessary to understand whether these differences were significant within the context of controlling for other variables via the multi-level regression model.

Research Question 2 was answered using a multi-level logistic regression model, given the binary nature of the dependent variable (i.e., exit vs. no exit) and completed within the *R* statistical programming language. The model took the form as described below.

For the student-level:

$$\text{logit}(\text{bin\_exit}_{ij}) = \beta_{0j} + \beta_{1j}\cdot\text{race\_eth}_{ij} + \beta_{2j}\cdot\text{Z\_fall}_{ij} + \beta_{3j}\cdot\text{grade1}_{ij} + \beta_{4j}\cdot\text{grade2}_{ij} + \beta_{5j}\cdot\text{grade3}_{ij} + \beta_{6j}\cdot\text{att\_ses}_{ij} + \beta_{7j}\cdot\text{prop\_att}_{ij} + e_{ij}$$

For the school-level:

$\beta_{0j} = \gamma_{00} + u_{0j}$

$\beta_{1j} = \gamma_{10} + u_{1j}$

where:

- γ_{00} = -1.235 and is the overall average of coefficient β_{0j}
- u_{0j} is drawn from a random variable with a normal distribution, having a mean of 0 and a variance of 0.540. Its value varies between different schools.
- γ_{10} = 0.688 and is the overall average of coefficient β_{1j} (NOTE: Black or African American was coded '0' and White was coded '1')
- u_{1j} is drawn from a random variable with a normal distribution, having a mean of 0 and a variance of 0.071. Its value varies between different schools.
- Furthermore, the correlation between u_{0j} and u_{1j} is -0.32.
- β_{2j} is the estimate for fall z-score, which is 1.163
- β_{3j} is the estimate for first grade, which is -1.723

- β_{4j} is the estimate for second grade, which is -2.555
- β_{5j} is the estimate for third grade, which is -1.809
- β_{6j} is the estimate for att_ses (i.e., total number of sessions attended across the year), which is -0.025
- β_{7j} is the estimate for prop_att, which is 5.593

The coefficients in the logistic regression equation represent the change in the log-odds of exiting the program for a one-unit change in the corresponding student-level predictor variables, all other variables being equal. Additionally, the model includes random effects that account for the unobserved heterogeneity across different school sites identified by the site identification number, which are crucial in capturing variability and improving the model's accuracy and predictive power.

The model's random effects suggested variability in bin_exit across different school sites (N = 51), especially considering variance in intercepts and race_eth slopes. More specifically, the variance for the intercept was .540 and the variance for the race_eth effect within site was .071. Their correlation was -.32. Overall, this means that when a school site had a high baseline likelihood value for bin_exit, the effect related to race_eth was lower, and when the site had a low baseline likelihood value for bin_exit, the effect related to race_eth was higher. The strength of this inverse relationship was -.32.

The model's fixed effects are highlighted in Table 4. Cells containing a "*" indicate that a variable was a statistically-significant predictor of the log-odds of exit. **As the table shows, the race_eth variable was a statistically significant predictor; white students were significantly more likely to have exited the program at the end of the year compared to Black students**. However, it is important to remember that the random effects suggested that the influence of race_eth on bin_exit varied considerably by school site. In those sites where there is a particularly high overall value for bin_exit, the effect of race was weaker than that when there was a low overall value for bin_exit.

In terms of fixed effects examined apart from race_eth, the fall z-score was significantly positively related with bin_exit, suggesting that those who were higher achieving at the beginning of the year tended to be more likely to exit. Those students in grades 1, 2, and 3 were significantly less likely to exit than those in kindergarten. The number of sessions attended was significantly negatively associated with bin_exit, while the proportion of sessions attended was significantly positively associated with bin_exit.

Table 4. Fixed effects regression summary for multilevel model associated with Research Question 2 *(bin_exit as dependent variable)*

Variable	Coefficient (standard error)
(*intercept*)	-1.235 (.98)
race_eth	0.689(.30)*
fall z-score	1.163 (.16)***
first grade	-1.722(.52)***
second grade	-2.554 (.53)***
third grade	-1.809 (.53)***
att_ses	-0.025(.00)***
prop_att	5.593 (1.42)***

Note. "*" denotes statistical significance ($p < .05$). "**" denotes statistical significance ($p < .01$). "***" denotes statistical significance ($p < .001$)

Research Question 3: To what extent do students who identify as Black or African American and White differ in regard to the number of intervention change decisions made?

All cases included values for the variable "number of intervention changes made" (i.e., inter_change). The same matchings were therefore used as those described for Research Question 1 (see Research Question 1 above for those results).

Appendix B shows the distributions of the targeted variable "number of intervention changes made" for the matched samples; although there are some minor differences evident, further analysis was considered necessary to understand whether these differences were significant within the context of controlling for other variables via the multi-level regression model.

Research Question 3 was answered using a Poisson regression model, given sample data exhibited a mean (1.25) that was approximately equal to the variance (1.03). This proximity suggests that the data could reasonably comply with the equidispersion assumption—where the mean equals the variance—a fundamental prerequisite for the validity of Poisson regression models in analyzing data that are presumed to follow a Poisson distribution. Model estimation was completed within the *R* statistical programming language. The model took the form as described below.

For the student-level:

$$log(E(inter\_change)) = \beta_{0j} + \beta_{1j} \cdot race\_eth_{ij} + \beta_{2j} \cdot Z\_fall_{ij} + \beta_{3j} \cdot grade1_{ij} + \beta_{4j} \cdot grade2_{ij} + \beta_{5j} \cdot grade3_{ij} + \beta_{6j} \cdot att\_ses_{ij} + e_{ij}$$

For the school-level:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

where:

- γ_{00} = -.607 and is the overall average of coefficient β_{0j}
- u_{0j} is drawn from a random variable with a normal distribution, having a mean of 0 and a variance of 0.058. Its value varies between different schools.
- γ_{10} = 0.108 and is the overall average of coefficient β_{1j} (NOTE: Black or African American was coded '0' and White was coded '1')
- u_{1j} is drawn from a random variable with a normal distribution, having a mean of 0 and a variance < 0.001. Furthermore, the correlation between u_{0j} and u_{1j} is 1.00.
- β_{2j} is the estimate for fall z-score, which is -0.096
- β_{3j} is the estimate for first grade, which is -0.070
- β_{4j} is the estimate for second grade, which is -0.494
- β_{5j} is the estimate for third grade, which is -0.499
- β_{6j} is the estimate for att_ses, which is 0.013

The coefficients in the regression equation signify the change in the logarithm of the expected count of inter_change due to a one-unit modification in the corresponding predictor variable, with other variables held constant.

The model's random effects suggested variability in inter_change across different school sites (N = 51), with variance for the intercept under site_id at .058 with a standard deviation of .240. Additionally, the correlation between the intercept and race_eth within site_ID was perfect (1.00), suggesting potential multicollinearity or that the variance is coming solely from the difference between sites rather than the interaction between site and race.

The model's fixed effects are highlighted in Table 5. Cells containing a "*" indicate that a variable was a statistically-significant predictor of the logarithm of the expected count of inter_change, with other variables held constant. **As the table shows, the race_eth variable was not a statistically significant predictor of number of intervention changes.**

In terms of fixed effects separate from race/ethnicity, the fall z-score was significantly negatively related with inter_change, suggesting that those who were lower achieving at the beginning of the year tended to have a higher number of intervention changes. Both grade 2 and grade 3 students had significantly lower inter_change than kindergarten students, and the number of sessions attended was significantly positively related to inter_change.

Table 5. Fixed effects regression summary for multilevel model associated with Research Question 3 (*Intervention change count as dependent variable*)

Variable	Coefficient (standard error)
(*intercept*)	-.607 (.18)***
race_eth	.108 (.09)
fall z-score	-.096 (.04)*
first grade	-.070 (.15)
second grade	-.494 (.16)**
third grade	-.499 (.17)**
att_ses	.013 (.00)***

Note. "*" denotes statistical significance ($p < .05$). "**" denotes statistical significance ($p < .01$). "***" denotes statistical significance ($p < .001$).

Research Question 4: To what extent do students who identify as Black or African American and White differ in regard to the number of materials for family engagement sent home?

Due to missing data for the dependent variable of "total number of program materials sent home with the student for practice during the year" (i.e., fam_sent), rematching was needed prior to further analysis. Using a 1:1 nearest neighbor matching process of Black or African American students to White students, matching for grade and fall z-score yielded a final sample of 199 Black or African American students and 199 White students. An additional 128 students were unmatched (26 Black or African American students and 102 White students) and were therefore excluded from the analysis, in addition to those mentioned earlier who did not have valid fall z-scores or data for fam_sent.

The results of the matching procedure (displayed in Table 6 below) produced a sample of Black or African American students and White students that were similar according to grade and fall z-score.

Before matching, the racial groups were different on these characteristics as follows:

- 7% of Black or African American students were in Kindergarten, compared with 11% of White students
- 25% of Black or African American students were in first grade, compared with 27% of White students
- 34% of Black or African American students were in second grade, compared with 39% of White students

- 34% of Black or African American students were in third grade, compared with 24% of White students
- Black students had an average fall z-score of .09, compared with .16 for White students.

After matching, the racial groups were much more similar as follows:

- 6% of Black or African American students were in Kindergarten, compared with 6% of White students
- 28% of Black or African American students were in first grade, compared with 31% of White students
- 38% of Black or African American students were in second grade, compared with 35% of White students
- 28% of Black or African American students were in third grade, compared with 28% of White students
- Black students had an average fall z-score of .08, compared with .13 for White students.

Table 6. Results of Matching for Student Selection for Research Question 4

Matching variable	Pre-matching		Post-matching	
	Black or African American	White	Black or African American	White
Kindergarten	7%	11%	6%	6%
First Grade	25%	27%	28%	31%
Second Grade	34%	39%	38%	35%
Third Grade	34%	24%	28%	28%
Fall Z-Score	.09	.16	.08	.13

Appendix B shows the distributions of the targeted variable fam_sent for the matched samples; although there are some minor differences evident, further analysis was considered necessary to understand whether these differences were significant within the context of controlling for other variables via the multi-level regression model.

Research Question 4 was answered using negative binomial link function as part of the multilevel regression model, given that the sample mean was much lower than the observed variance, suggesting overdispersion. The negative binomial distribution is particularly applicable to such situations due to its flexibility in handling extra-Poisson variation. The regression work was completed within the *R* statistical programming language. The model took the form as described below.

For the student-level:

$$log(E(fam\_sent)) = \beta_{0j} + \beta_{1j} \cdot race\_eth_{ij} + \beta_{2j} \cdot Z\_fall_{ij} + \beta_{3j} \cdot grade1_{ij} + \beta_{4j} \cdot grade2_{ij} + \beta_{5j} \cdot grade3_{ij} + \beta_{6j} \cdot att\_ses_{ij} + e_{ij}$$

For the school-level:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

where:

- γ_{00} = 1.117 and is the overall average of coefficient β_{0j}
- u_{0j} is drawn from a random variable with a normal distribution, having a mean of 0 and a variance of 0.649. Its value varies between different schools.

- γ_{10} = 0.124 and is the overall average of coefficient β_{1j} (NOTE: Black or African American was coded '0' and White was coded '1')
- u_{1j} is drawn from a random variable with a normal distribution, having a mean of 0 and a variance of .016. Furthermore, the correlation between u_{0j} and u_{1j} is 1.00.
- β_{2j} is the estimate for fall z-score, which is 0.116
- β_{3j} is the estimate for first grade, which is 0.383
- β_{4j} is the estimate for second grade, which is 0.418
- β_{5j} is the estimate for third grade, which is 0.397
- β_{6j} is the estimate for att_ses, which is 0.013

The coefficients in the regression equation signify the change in the logarithm of the expected count of fam_sent due to a one-unit modification in the corresponding predictor variable, with other variables held constant.

The model's random effects suggested variability in fam_sent across different school sites (N = 40), with variance for the intercept under site_id at .65 with a standard deviation of .81. Additionally, the correlation between the intercept and race_eth within site_ID was perfect (1.00), suggesting potential multicollinearity or that the variance is coming solely from the difference between sites rather than the interaction between site and race.

The model's fixed effects are highlighted in Table 7. Cells containing a "*" indicate that a variable was a statistically-significant predictor of the logarithm of the expected count of materials sent home. **As the table shows, the race_eth variable was not a statistically significant predictor of the number of materials sent home.**

In terms of fixed effects examined apart from race/ethnicity, the fall benchmark z-score was significantly positively related with the count of materials sent home, suggesting that those who were higher achieving at the beginning of the year tended to have a higher proportion of materials sent home. Being in first, second, or third grade (as opposed to kindergarten) was associated with higher counts of materials sent home. Finally, the number of sessions attended was associated with higher counts of materials sent home.

Table 7. Fixed effects regression summary for multilevel model associated with Research Question 4 *(family materials sent home as dependent variable)*

Variable	Coefficient (standard error)
(*intercept*)	1.117 (.24)***
race_eth	.124 (.12)
fall z-score	.116 (.05)*
first grade	.383 (.19)*
second grade	.418 (.19)*
third grade	.397 (.20)*
att_ses	.013 (.00)***

Note. "*" denotes statistical significance ($p < .05$). "**" denotes statistical significance ($p < .01$). "***" denotes statistical significance ($p < .001$).

Research Question 5: To what extent do students who identify as Black or African American and White differ in regard to spring benchmark scores, when holding constant implementation variables associated with the program?

Due to missing data for the dependent variable of "spring z-score", rematching was needed prior to further analysis. One outlier was identified (White student) and removed prior to matching. Using a 1:1 nearest neighbor matching process of Black or African American students to White students, matching for grade and fall z-score yielded a final sample of 277 Black or African American students and 277 White students. An additional 162 students were unmatched (51 Black or African American students and 111 White students) and were therefore excluded from the analysis, in addition to those mentioned earlier who did not have valid fall z-scores or data for spring z-score.

The results of the matching procedure (displayed in Table 8 below) produced a sample of African American students and White students that were similar according to grade and fall z-score.

Before matching, the racial groups were different on these characteristics as follows:

- 6% of Black or African American students were in kindergarten, compared with 9% of White students
- 21% of Black or African American students were in first grade, compared with 27% of White students
- 34% of Black or African American students were in second grade, compared with 37% of White students
- 39% of Black or African American students were in third grade, compared with 27% of White students
- Black students had an average fall z-score of -.15, compared with .10 for White students.

After matching, the racial groups were much more similar as follows:

- 6% of Black or African American students were in kindergarten, compared with 7% of White students
- 23% of Black or African American students were in first grade, compared with 27% of White students
- 40% of Black or African American students were in second grade, compared with 31% of White students
- 31% of Black or African American students were in third grade, compared with 35% of White students
- Black students had an average fall z-score of -.08, compared with .00 for White students.

Table 8. Results of Matching for Student Selection for Research Question 5

Matching variable	Pre-matching		Post-matching	
	Black or African American	White	Black or African American	White
Kindergarten	6%	9%	6%	7%
First Grade	21%	27%	23%	27%
Second Grade	34%	37%	40%	31%
Third Grade	39%	27%	31%	35%
Fall Z-Score	-.15	.10	-.08	.00

Appendix B shows the distributions of the targeted variable spring z-score for the matched samples; although scores appear slightly higher for white students, further analysis was considered necessary to understand whether these differences were significant within the context of controlling for other variables via the multi-level regression model.

Research Question 5 was answered using a linear mixed-effects model, given that the dependent variable (spring z-score) followed a normal distribution, completed within the *R* statistical programming language. The model took the form as described below.

For the student-level:

spring z-score= $\beta_{0j} + \beta_{1j} \cdot race\_eth_{ij} + \beta_{2j} \cdot Z\_fall_{ij} + \beta_{3j} \cdot grade1_{ij} + \beta_{4j} \cdot grade2_{ij} + \beta_{5j} \cdot grade3_{ij} + \beta_{6j} \cdot att\_ses_{ij} + \beta_{7j} \cdot prop\_att_{ij} + \beta_{8j} \cdot fidelity_{ij} + \beta_{9j} \cdot fam\_sent_{ij} + e_{ij}$

For the school-level:

$\beta_{0j} = \gamma_{00} + u_{0j}$

$\beta_{1j} = \gamma_{10} + u_{1j}$

where:

- γ_{00} = -1.444 and is the overall average of coefficient β_{0j}
- u_{0j} is drawn from a random variable with a normal distribution, having a mean of 0 and a variance of 0.075. Its value varies between different schools.
- γ_{10} = 0.119 and is the overall average of coefficient β_{1j} (NOTE: Black or African American was coded '0' and White was coded '1')
- u_{1j} is drawn from a random variable with a normal distribution, having a mean of 0 and a variance of 0.150. Its value varies between different schools.
- Furthermore, the correlation between u_{0j} and u_{1j} is -0.49.
- β_{2j} is the estimate for fall z-score, which is 0.482
- β_{3j} is the estimate for first grade, which is -0.389
- β_{4j} is the estimate for second grade, which is -0.230
- β_{5j} is the estimate for third grade, which is -0.181
- β_{7j} is the estimate for att_ses, which is -0.004
- β_{8j} is the estimate for prop_att, which is 0.798
- β_{9j} is the estimate for fidelity, which is 1.708
- β_{10j} is the estimate for fam_sent, which is -0.001

The coefficients in the linear mixed-effects model represent the average change in the z_spring score for a one-unit change in the respective predictors, considering all other predictors to be constant. Additionally, the model includes random effects that account for the unobserved heterogeneity across different school sites identified by the site identification number, which are crucial in capturing variability and improving the model's accuracy and predictive power.

The model's random effects suggested variability across different school sites (N = 37), especially considering variance in intercepts and race_eth slopes. More specifically, the variance for the intercept was .075 and the variance for the race_eth effect within site was .150. Their correlation was -.49.

The model's fixed effects are highlighted in Table 9. Cells containing a "*" indicate that a variable was a statistically-significant predictor of spring z-score. **As the table shows, the race_eth variable was not a statistically significant predictor**; however, it is important to remember that the random effects suggested that the influence of race_eth on spring z-score varied considerably by school site. When a school site had a high baseline value for spring z-score, the effect related to race_eth was lower (i.e., Black or African American students tended to have either more similar z-scores or higher spring z-scores than White students), and when the site had a lower baseline value for spring z-score, the effect related to race_eth was higher (i.e., White students tended to have higher spring z-scores than Black students).

In terms of fixed effects examined apart from race/ethnicity, the fall z-score was significantly positively related with spring z-score, suggesting that those who were higher achieving at the beginning of the year tended to be higher achieving at the end of the year. Compared to kindergarteners, first graders tended to have significantly lower spring z-scores; this may reflect how kindergarteners are the individuals most likely to respond to intervention and achieve higher spring scores. The number of sessions attended (i.e., att_ses) displayed a significant negative association with spring z-scores, and the proportion of sessions attended (i.e. prop_att) displayed a significant positive relationship with spring z-scores. The negative association between number of sessions attended and spring z-scores may reflect the notion that those who were not making progress tended to be more likely to continue in the program for a longer period of time. The marginally significant positive effect for proportion of sessions attended aligns with the notion that consistency in program attendance and engagement is associated with higher spring scores. Fidelity was not a significant predictor; however, this may be due to the fact that fidelity in carrying out the intervention strategies within sessions was measured to be consistently high.

Table 9. Fixed effects regression summary for multilevel model associated with Research Question 5 *(spring z-score as dependent variable)*

Variable	Coefficient (standard error)
(*intercept*)	-1.444 (1.80)
race_eth	.119 (.11)
fall z-score	.482 (.02)***
first grade	-.389 (.14)**
second grade	-.230 (.14)
third grade	-.181 (14)
att_ses	-.004 (.00)**
prop_att	.798 (.46)^
fidelity	1.708 (1.83)
fam_sent	-.001 (.001)

Note. "^" denotes marginal significance ($p < .1$). "*" denotes statistical significance ($p < .05$). "**" denotes statistical significance ($p < .01$). "***" denotes statistical significance ($p < .001$).

Research Question 6: To what extent do students who identify as Black or African American and White differ in regard to their growth scores during intervention?

Due to missing data for the dependent variable of growth z-score, rematching was needed prior to further analysis. Using a 1:1 nearest neighbor matching process of Black or African American students to White students, matching for grade and fall benchmark z-score yielded a final sample of 289 Black or African America students and 289 White

students. An additional 160 students were unmatched (58 Black or African American students and 102 White students) and were therefore excluded from the analysis, in addition to those mentioned earlier who did not have valid fall z-scores or data for growth z-score.

The results of the matching procedure (displayed in Table 10 below) produced a sample of Black or African American students and White students that were similar according to grade and fall z-score.

Before matching, the racial groups were different on these characteristics as follows:

- 6% of Black or African American students were in kindergarten, compared with 9% of White students
- 20% of Black or African American students were in first grade, compared with 27% of White students
- 33% of Black or African American students were in second grade, compared with 36% of White students
- 41% of Black or African American students were in third grade, compared with 28% of White students
- Black students had an average fall z-score of -.23, compared with .07 for White students.

After matching, the racial groups were much more similar as follows:

- 5% of Black or African American students were in Kindergarten, compared with 6% of White students
- 23% of Black or African American students were in first grade, compared with 24% of White students
- 39% of Black or African American students were in second grade, compared with 33% of White students
- 33% of Black or African American students were in third grade, compared with 37% of White students
- Black students had an average fall z-score of -.14, compared with -.03 for White students.

Table 10. Results of Matching for Student Selection for Research Question 6

Matching variable	Pre-matching		Post-matching	
	Black or African American	White	Black or African American	White
Kindergarten	6%	9%	5%	6%
First Grade	20%	27%	23%	24%
Second Grade	33%	36%	39%	33%
Third Grade	41%	28%	33%	37%
Fall Z-Score	-.23	.07	-.14	-.03

Appendix B shows the distributions of the targeted variable “growth z-score” for the matched samples; although scores appear slightly higher for Black or African American students, further analysis was considered necessary to understand whether these differences were significant within the context of controlling for other variables via the multi-level regression model.

Research Question 6 was answered using a linear mixed-effects model, given that the dependent variable (z-score for growth during program participation) followed a normal distribution. It was completed within the *R* statistical programming package.The model took the form as described below.

For the student-level:

$$\text{growth z-score} = \beta_{0j} + \beta_{1j}\cdot\text{race\_eth}_{ij} + \beta_{2j}\cdot\text{Z\_fall}_{ij} + \beta_{3j}\cdot\text{grade1}_{ij} + \beta_{4j}\cdot\text{grade2}_{ij} + \beta_{5j}\cdot\text{grade3}_{ij} + \beta_{6j}\cdot\text{att\_ses}_{ij} + \beta_{7j}\cdot\text{prop\_att}_{ij} + e_{ij}$$

For the school-level:

$\beta_{0j} = \gamma_{00} + u_{0j}$

$\beta_{1j} = \gamma_{10} + u_{1j}$

where:

- γ_{00} = -0.401 and is the overall average of coefficient β_{0j}
- u_{0j} is drawn from a random variable with a normal distribution, having a mean of 0 and a variance of 0.093. Its value varies between different schools.
- γ_{10} = -.020 and is the overall average of coefficient β_{1j} (NOTE: Black or African American was coded '0' and White was coded '1')
- u_{1j} is drawn from a random variable with a normal distribution, having a mean of 0 and a variance of 0.010. Its value varies between different schools.
- Furthermore, the correlation between u_{0j} and u_{1j} is -0.91.
- β_{2j} is the estimate for fall z-score, which is 0.084
- β_{3j} is the estimate for first grade, which is .338
- β_{4j} is the estimate for second grade, which is 0.273
- β_{5j} is the estimate for third grade, which is 0.248
- β_{7j} is the estimate for att_ses, which is -0.008
- β_{8j} is the estimate for prop_att, which is 1.028

The coefficients in the linear mixed-effects model represent the average change in the growth z-score for a one-unit change in the respective predictors, considering all other predictors to be constant. Additionally, the model includes random effects that account for the unobserved heterogeneity across different school sites identified by the site identification number, which are crucial in capturing variability and improving the model's accuracy and predictive power.

The model's random effects suggested variability across different school sites (N = 51), especially considering variance in intercepts and race_eth slopes. More specifically, the variance for the intercept was 0.093 and the variance for the race_eth effect within site was 0.010. Their correlation was -.91. Overall, this means that when a school site had a high baseline value for growth z-score, the effect related to race_eth was lower, and when the site had a lower baseline value for growth z-score, the effect related to race_eth was higher. The strength of this inverse relationship was measured at -.91.

The model's fixed effects are highlighted in Table 11. Cells containing a "*" indicate that a variable was a statistically-significant predictor of z_growth score. **As the table shows, the race_eth variable was not a statistically significant predictor**; however, it is important to remember that the random effects suggested that the influence of race_eth on growth z-score varied considerably by school site. More specifically, when a school site had a high baseline value for growth z-score, the effect related to race_eth was lower than the overall fixed effect identified (i.e., Black students tended to have higher growth z-scores than White students), and when the site had a lower baseline value for growth z-score, the effect related to race_eth was particularly weak (i.e., students had similar growth rates regardless of race).

In terms of fixed effects examined apart from race/ethnicity, the fall benchmark z-score was significantly positively related with growth z-score, suggesting that those who were higher achieving at the beginning of the year tended to experience higher growth rates during intervention. Compared to kindergarteners, first graders, second graders, and third graders tended to have significantly higher growth z-scores. The number of sessions attended (i.e., att_ses) displayed a significant negative association with growth z-scores, which aligns with the notion that students who were making less growth were more likely to continue in the program over time rather than exit. The proportion of sessions attended (i.e., prop_att) displayed a significant positive relationship with z-growth score, suggesting that the rate of receiving intervention sessions was significant related to growth per week during intervention.

Table 11. Fixed effects regression summary for multilevel model associated with Research Question 6 *(growth z-score as dependent variable)*

Variable	Coefficient (standard error)
(*intercept*)	-0.401 (0.22)
race_eth	-0.02 (0.07)
fall z-score	0.084 (0.03)**
first grade	0.338 (0.12)**
second grade	0.273 (0.12)*
third grade	0.348 (0.12)*
att_ses	-0.008 (0.00)***
prop_att	1.028 (0.316)**

Note. "*" denotes statistical significance ($p < .05$). "**" denotes statistical significance ($p < .01$). "***" denotes statistical significance ($p < .001$).

Limitations

Although the data available for this analysis were substantial and involved use of similar measures across school sites (which is an advantage compared to other similar analyses conducted at an earlier point in time, it is important to acknowledge important limitations of the data available for analysis and corresponding analyses. These are described below. Related suggestions to consider for future investigations of racial equity in programming and outcomes for MRC are indicated in the recommendations section.

Data-related challenges

Several limitations related to data availability are important to note. First, the availability of fall and spring benchmark data for students was in many cases related to whether the student participated in the program near to the time of the given benchmark data collection. Less than half of the student pool from whom the matched students were selected had both fall and spring benchmark scores available in the dataset. Because of the desire to match students on fall benchmark scores, the analyses reflect results for only a subset of cases and is likely less representative of students who participated only in the spring. With regard to matching, it is also important to note that students could not be matched within school site, which would have been ideal to help in controlling for differences that might be due to existing school programming. Moreover, just a little over half of the students had values entered for materials for family engagement sent home; results were limited to those who had such values included. As a result of missing data, as well as situations in which appropriate matches were not available, it is important to note that the results for each regression analysis

represent only subsets of students who participated in the program and may not comprehensively represent the program.

Although similar measurement tools were used across schools, different measurement tools were used across grade levels. To facilitate group sizes for meaningful analysis and offer a helpful summary of equity in program activities and outcomes across grade levels, data were transformed so that analyses were run in a manner that aggregated results across grade-levels. It is important to note that measuring reading in different ways across different grade levels may have played a role in our ability to detect differences present according to race. Earlier evaluation studies have run analyses separately by grade level (e.g., Markowicz et al., 2019), which may be an important consideration for future work.

With respect to Research Question 1, it is important to highlight (as was indicated earlier) that the variable "proportion of sessions attended" may have been substantially influenced by the timing with which the student received intervention due to holiday breaks; conducting comparisons of students who receive intervention at the same time of year and for the same amount of time may be helpful toward engaging in a more sensitive analysis of differences across race.

With respect to Research Question 4 about family engagement, it is important to note that the materials for "family materials sent home" may have represented very different activities and may not necessarily be a particularly valid quantitative indicator of less and greater efforts to engage families in reading activities at home.

Interpretation challenges

Within Michigan, it is important to note that schools vary considerably with respect to the core reading instructional programming provided to students. Because all participating students included in the analysis were receiving both their core school programming and Michigan Reading Corps, it is impossible to disentangle the extent to which growth and outcomes are attributable to the existing school programming vs. attributable to Michigan Reading Corps. Failure to identify significant differences based on race suggests that altogether, the school programming and Michigan Reading Corps programming that students experienced corresponded to relatively similar outcomes by race. However, without control groups of Black and White students who did not participate it is not possible to know the specific contributions that Michigan Reading Corps made to achieving those similar outcomes. Moreover, results for several of the analyses highlighted variation in findings across school sites, which highlights the potential role that school programming and school-related differences may play in facilitating or diminishing equity in student outcomes. In the future, it may be helpful to collect data on school programming and existing racial differences in outcomes among non-participating students to better understand the contribution that MRC offers for improving equity in outcomes for students in different races.

It is important to acknowledge that this series of analyses did not involve consideration or control for socio-economic status, which is often recommended in attempts to disentangle what differences may be due to factors such as low-income economic marginalization vs. racial minoritization. The analyses did include extensive consideration and removal of effects of a variety of variables that may be different across students of different races (e.g., initial z-score) and which may correspond to the outcome variables of interest. Although this was the desired approach for the analysis as

indicated in the evaluation plan, it is important to note that the small race-related effects identified may be due to the associated decisions to match based on fall z-score and include a variety of covariates in each analysis.

Conclusions and Implications

Conclusions

Several key findings from the study are important to highlight:

1) Of the six targeted program experiences and outcomes examined, for only one (i.e., exit status at the end of the year) was race a statistically significant predictor when controlling for other relevant variables. More specifically, White students were more likely to be exited by the end of the academic year than Black or African American students, controlling for a variety of other relevant variables.
2) For several of the targeted program experiences and outcomes examined, the relationship between race and the given program experience or outcome was found to vary by school site. In each of these situations, the relationship between race and the targeted variable was lower (i.e., either weaker or in the direction of Black students experiencing more positive experiences or outcomes than White students) in schools with high baseline rates for the targeted experience or outcome variable.
3) Fall benchmark score was consistently positively related with student reading growth and outcomes, as was proportion of program sessions attended.

Although further investigation of decision-making around exit may be warranted from the perspective of seeking to better understand differences in how MRC is experienced by students of different races, it is important to note that it does not appear that a correspondingly different reading growth rate or spring score was evident based on race. Prior work has suggested that the program is effective for both groups, and so it is possible that the differences in exit rates as identified in the current study may correspond to helping to reduce race-related disparities in student reading outcomes. However, further investigation could help to determine if such occurs within the current contexts where Reading Corps is being applied.

The identification of differences across schools in the nature of the relationship between race and the targeted program experiences and outcomes highlights a challenge in understanding the potential contribution of MRC - as opposed to the contribution of other school or community-related experiences - to racial disparities (or lack thereof) in MRC experiences and overall outcomes. Racial disparities in outcomes (or lack thereof) may correspond to a variety of other school-related factors apart from those that are directly a part of MRC. Relatedly, in the current study, there was no control group utilized to help understand racial discrepancies that might be accounted for by school programming to be able to more carefully investigate the unique contribution of MRC to either increasing or reducing racial discrepancies in outcomes, and so it is critical to interpret the findings of the current evaluation and its implications for improving the program very cautiously. Moreover, a considerable amount of missing data for both predictor and outcome variables of interest was evident, making it such that the results reported represent only unique subgroups of program participants.

The trends identified, however, do align with a general notion that where involvement in MRC experiences is high and where student growth and spring scores are high, racial disparities are limited, and in some cases racial disparities are the direction of showing better experiences and outcomes among those from the racially-minoritized sub-group (i.e.,

Black or African American students). Moreover, the consistent positive predictive relationship between proportion of sessions attended and reading growth and outcomes aligns with a notion that MRC receipt is associated with growth. At the same time, the positive relationship identified between fall z-score and outcomes also suggests that MRC may not result in particularly accelerated gains for those participating students who start the year particularly behind compared to those participating students who start the year less behind. Although potentially serving as an effective tool for those who receive the program, positive relationship between fall benchmark score and growth suggests that among those who receive Reading Corps, the program does not currently result in greater advantage among those MRC participants who appear to be in most need of additional support.

Implications

Accurate measurement of the effects of the program and the extent to which they may or may not facilitate great racial equity in reading outcomes requires a more systematic approach than that used in the current study. Collecting and analyzing data for control groups of Black or African American and White students who are eligible but do not receive the MRC program at the corresponding school sites would be needed to better understand the specific program-related contributions to racial equity in student outcomes. The racial disparities (and lack thereof) identified within the current analysis must correspondingly be considered carefully, as they simply represent the outcome of both school- and program-related program experiences and not necessarily those specific to the Reading Corps program.

Given the significant difference identified in exit rates by race, examination of decision-making related to exit may be worthwhile toward a goal of better understanding any existing racial disparities in program experiences, given that it was the one area in which overall program racial differences were identified. It is important to recognize that differences in exit that retain Black students in the program longer than White students may be a mechanism by which existing racial disparities in reading outcomes (i.e., reading outcomes in general tend to be higher for White students than Black students) are diminished, given existing support that the program does promote positive outcomes.

Given that the variation in racial disparities often varied by school site, it may be worthwhile to explore variables that may explain that variation; such work might help to identify what could be done to reduce particularly harmful racial disparities in those sites where Black or African American students tend to have fewer program-related experiences and lower outcomes than White students. Exploration of the dynamics at play in both those sites where racial equity appears to be present (i.e., Black students are gaining either the same or more than White students) and comparing that to those sites where problematic disparities exist may shed light on efforts that might be helpful toward a goal of promoting greater equity at all sites.

The proportion of program sessions attended was the single most consistently positive predictor of growth and spring z-score; relatedly, attendance may be the single most important vehicle for improving reading growth and outcomes for all students eligible to participate.

More consistent reading achievement data collection (e.g., fall and spring benchmarks for all program participants, reporting for all students of family engagement materials sent home) would allow for the results to be more representative of all students who participated in the program. In addition, although the parameters of this evaluation work were intended to allow for an understanding of aggregated program effects (combined k-3 analyses), it may be

helpful to examine results separately by grade, particularly given that different measures are used at different grade levels. Moreover, controlling for the time of events with regard to individual student intervention start and ending may allow for a more direct comparison of experiences according to race. Finally, information on socio-economic status could help explore potential contributions of low income economic marginalization versus racial minoritization in future evaluations and analyses.

Recommendations

To better understand racial disparities in experiences and outcomes, and more specifically MRC's contributions to racial equity in reading outcomes, the following recommendations are offered:

Recommendations for Data Collection

- Participating schools could be encouraged and supported to engage all students (even non-program participants) in the reading benchmarking tools used by MRC. If the participating schools collect the same benchmarking data for all students in the school at all three time points, it is anticipated that this will help (a) prevent the high levels of missing benchmark data in the data sets used in the current analysis and allow the results to correspondingly reflect those of all MRC participants, and (b) facilitate opportunities to more carefully disentangle contributions of school programming as opposed to MRC programming to the reading growth and outcomes of program participants.
- If engaging families in MRC-related reading activities is of particular interest, it is recommended that in addition to ensuring collection of such data for all participating students, a better measurement tool for examining this be identified, as the current tool (i.e., number of materials sent home) did not account for the nature of the materials provided that may have differed in expectations and scope depending on the intervention used within sessions. It also did not offer information on the family's use of those materials.
- Collecting information on which students experience low-income economic marginalization (LIEM) and incorporating that information in future analyses could help in disentangling the extent to which racial minoritization as opposed to low-income economic marginalization may play a role in any racial disparities identified.

Recommendations for Future Study

- Further consideration of the impact that the differential exit rates of students of different races has on their school experiences and reading growth, along with the decision-making processes and procedures that contribute to those differential exit rates may be a helpful direction for future work in an effort to ensure that the program is facilitating equity in reading and other school-related outcomes.
- To more comprehensively understand the contributions of MRC, as opposed to other school programming, to racial disparities in outcomes (or lack thereof), systematic manipulation of who receives the MRC tutoring services and when they receive could be used to allow for more controlled comparisons of differences in experiences and outcomes by race. Understandably, this may be difficult to achieve due to a desire to provide the intervention quickly to all those who qualify. Matching students within evaluation-related analyses based on site-location (when possible) and start time of intervention provision may also allow more helpful comparisons.
- Consider identification and analysis of site-based predictors that may contribute to racial disparities in the sites where strong relationships between race and experiences and outcomes were identified. An understanding of

such differences may help in identifying unique supports that may be needed in certain sites to promote greater racial equity in student experiences and outcomes.

- To the extent that group sizes are large enough, separate analyses by grade level may be helpful in order to prevent the differences in measurement tools and associated need to transform scores that was present in the current analyses from affecting the findings.

References

Hernandez, D. J. (2011). Double jeopardy: How third-grade reading skills and poverty influence high school graduation. *Annie E. Casey Foundation*.

Markovitz, C., Hernandez, M., Hedberg, E., & Silberglitt, B. (2014). *Outcome Evaluation of the Minnesota Reading Corps K3 Program*. NORC at the University of Chicago: Chicago, IL.

Markovitz, C., Hernandez, M., Hedberg, E., & Neishi, K. (2018a). *Impact Evaluation of the Wisconsin Reading Corps Program*. NORC at the University of Chicago: Chicago, IL.

Markovitz, C., Hernandez, M., Hedberg, E., & Whitmore, H. (2022). Evaluating the effectiveness of a volunteer one-on-one tutoring model for early elementary reading intervention: A randomized controlled trial replication study. *American Educational Research Journal, 59*(4), 788–819.

Markovitz, C., Hernandez, M., Hedberg, E., Whitmore, H., Satorius, J. (2018b). *Impact Evaluation of the Minnesota Reading Corps K-3 Program (2017-18*). NORC at the University of Chicago: Chicago, IL.

Snow, C., Burns, S., Griffin, P. (1998). *Preventing reading difficulties in young children*. National Academic Press. https://nap.nationalacademies.org/catalog/6023/preventing-reading-difficulties-in-young-children

Stanovich, K. E. (2009). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy." *Journal of Education*, 189, 23-55.

Tetta, P., Baker, J., & Medhanie, A. (2019). *Washington Reading Corps Impact Evaluation Final Report and Recommendations*. Washington Employment Security Department: Olympia, WA.

U.S. Department of Education (2022). *NAEP Report Card: 2022 NAEP Reading Assessment: Highlighted results at grades 4 and 8 for the nation, states, and districts*. https://www.nationsreportcard.gov/highlights/reading/2022/

Appendix A. Data Files, Original Variables, and Variables Created for Analysis

Data File	Variables Used For Description or Analysis (*directly entered for the targeted regression analyses)
Student_Demographics *(each row represents a student)*	<u>Direct from datafile</u> • site_ID*: a numeric label for the school the student attended • student_ID: a numeric label for the student (facilitates linking with data from other data files) • grade_id: student's grade with the following possible values - k = kindergarten, 1 = first grade, 2 = second grade, 3 = third grade • gender: student's gender with the following possible values: male, female, unknown • race/ethnicity: student's race with the following possible values: American Indian or Alaska Native, Asian, Black or African American, Hispanic/Latino, Multi-Racial, Unknown, White • latest_status: student's Reading Corps program status at the end of the program year with the following possible values: Active, Exit, Move, Stop-Stay
	<u>Created from datafile</u> • first grade*: dummy variable created based on grade_id with following possible values – 0 = not first grade, 1 = first grade • second grade*: dummy variable created based on grade_id with following possible values – 0 = not second grade, 1 = second grade • third grade*: dummy variable created based on grade_id with following possible values – 0 = not third grade, 1 = third grade • race_eth*: student's race created based on race/ethnicity with the following possible values – 0 = Black or African American, 1 = White • bin_exit*: student's program status at the end of the program year created from latest_status with the following possible values – 0 = Active or Stop-Stay, 1 = Exit, Missing = Move
Assessments *(each row represents an assessment instance for a student)*	<u>Direct from datafile</u> • student_ID: a numeric label for the student associated with the given assessment session (facilitates linking with data from other data files) • assessment_date: date of assessment tool administration session • bpa_id: type of assessment tool administered with the following possible values – benchmark, progress monitoring • assessment_tool_abbrev: specific grade-level assessment tool administered with the following possible values – CBM-R, LS, NW • correct score: the total raw score for the given assessment tool administration
	<u>Created from datafile</u> • fall z-score*: standardized score (mean = 0, standard deviation = 1) created using data for all cases (including all race/ethnicities in original dataset) in a

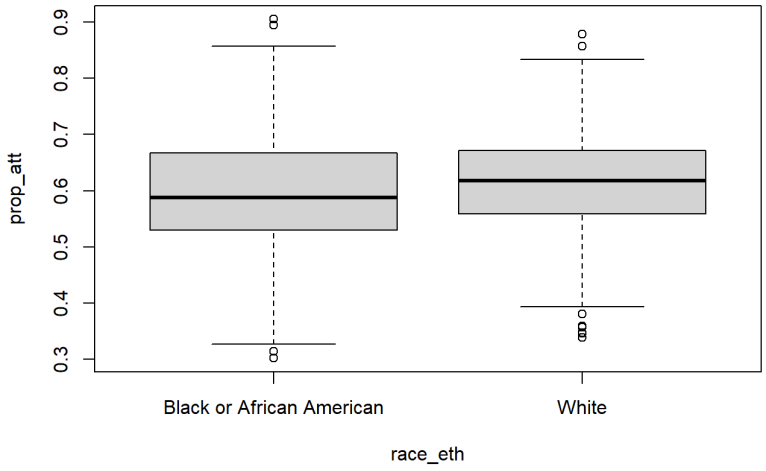
	given grade level that had benchmark score for the given grade-level assessment tool as listed below with an assessment date between 9/1/2022 and 9/30/2022; scores were transformed separately by grade level using the following tools- LS for kindergarten, NW for first grade, and CBM-R for second grade, and CBM-R for third grade • spring z-score*: standardized score (mean = 0, standard deviation = 1) created using data for all cases (including all race/ethnicities in original dataset) in a given grade level that had benchmark score for the given grade-level assessment tool as listed below with an assessment date between 4/24/2023 to 5/26/2023; scores were transformed separately by grade level using the following tools- LS for kindergarten, NW for first grade, and CBM-R for second grade, and CBM-R for third grade
Interventions *(each row represents a student's weekday experience during intervention programming, from the first weekday of their intervention programming to their final weekday of intervention programming, including days when school was not in session)*	<u>Direct from datafile</u> • student_ID: a numeric label for the student associated with the given weekday during intervention provision (facilitates linking with data from other data files) • tutor_id: a numeric label for the tutor associated with the given student's intervention programming • intervention_date: date associated with the given weekday • attendance: a number indicating the amount of time (in minutes) spent on the given weekday in an intervention session with values ranging from 15 to 20 OR marked "NULL" if student did not receive intervention on the given weekday • interventions: the specific intervention strategies or combination of intervention strategies provided (or intended to be provided) on the given weekday
	<u>Created from datafile</u> • att_ses*: student-level variable created using the attendance data for a given student – the value is the total number of intervention sessions the given student attended, and was calculated based on the total number of intervention sessions recorded for the given student_ID that included a number as opposed to being marked "NULL". • prop_att*: student-level variable created using the attendance data; the number of intervention sessions the given student attended divided by the number of weekdays from the first weekday of intervention provision to the last weekday of intervention provision (i.e., a proportion of sessions attended). • inter_change*: student-level variable created by ordering intervention sessions chronologically for the given student, and then identifying the number of times the intervention changed over time; a change was counted if either the strategy completing changed or if a new strategy was added or deleted compared to the prior strategy used

Family Engagement (*each row represents a student*)	<u>Direct from datafile</u> • student_ID: a numeric label for the student (facilitates linking with data from other data files) • fam_sent*: the total number of program materials sent home with the student for practice during the year
Fidelity (*each row represents a fidelity observation session*)	<u>Direct from datafile</u> • tutor_ID: a numeric label for the tutor associated with the fidelity check • fidelity_check_type: indicates whether the fidelity check related to an assessment or an intervention • pc_fidelity: the proportion of items on the fidelity checklist for the given assessment or intervention strategy that was identified as present during the session by the observer.
	<u>Created from datafile</u> • fidelity*: a student-level variable that represented an average pc_fidelity score for the student's tutor based on all intervention fidelity observation sessions held for the student's tutor
Avg_Weekly_Growth (*each row represents a student*)	<u>Direct from datafile</u> • student_ID: a numeric label for the student (facilitates linking with data from other data files) • grade_ID: student's grade with the following possible values - k = kindergarten, 1 = first grade, 2 = second grade, 3 = third grade • avg_ls: student's average weekly growth for letter sounds on grade-level probes during receipt of intervention programming • avg_nw: student's average weekly growth for nonsense word fluency on grade-level probes during receipt of intervention programming • avg_cbmr: student's average weekly growth for CBM-R on grade-level probes during receipt of intervention programming
	<u>Created from datafile</u> • growth z-score*: standardized score (mean = 0, standard deviation = 1) created using data for all cases (including all race/ethnicities in original dataset) in a given grade level that had average weekly growth score for the given grade-level assessment tool; scores were transformed separately by grade level using the following tools- LS for kindergarten, NW for first grade, CBM-R for second grade, and CBM-R for third grade.

Appendix B. Descriptive Statistics for the Six Dependent Variables Among Matched Samples

1. Proportion of Sessions Attended

	Black or African American	White
Minimum	0.0714	0.0000
1st Quartile	0.5274	0.5581
Median	0.5864	0.6176
Mean	0.5931	0.6088
3rd Quartile	0.6667	0.6719
Maximum	0.9048	0.8780

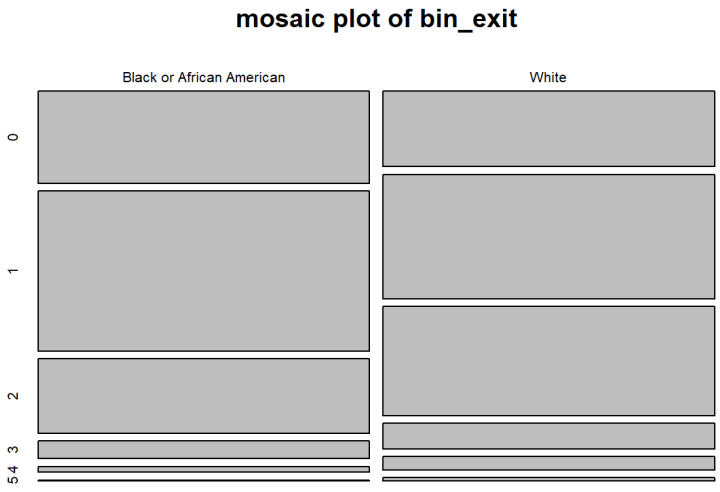


2. Exit Status at End of Year

	Black or African American	White
Not exit (i.e, active or stop-stay status at end of year)	226 (75%)	201 (67%)
Exit	74 (25%)	99 (33%)

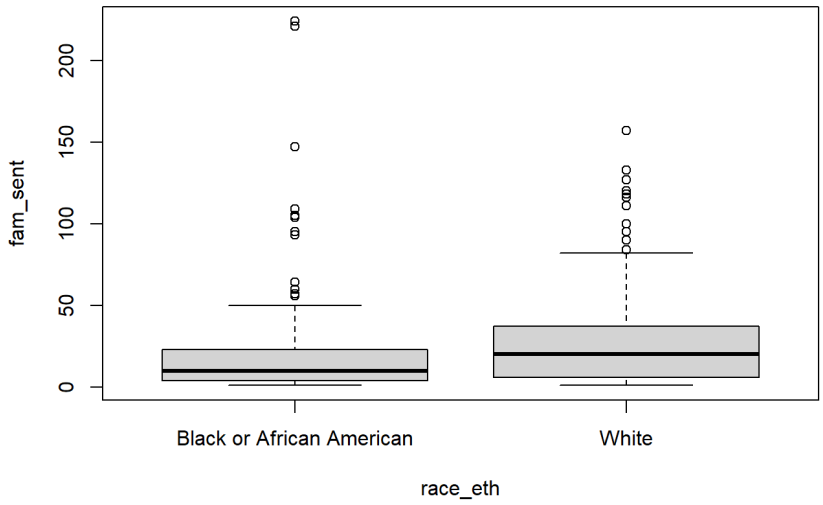
3. Number of Intervention Changes

Number of Intervention Changes	Black or African American	White
0	83	68
1	144	112
2	67	98
3	16	23
4	5	12
5	1	3

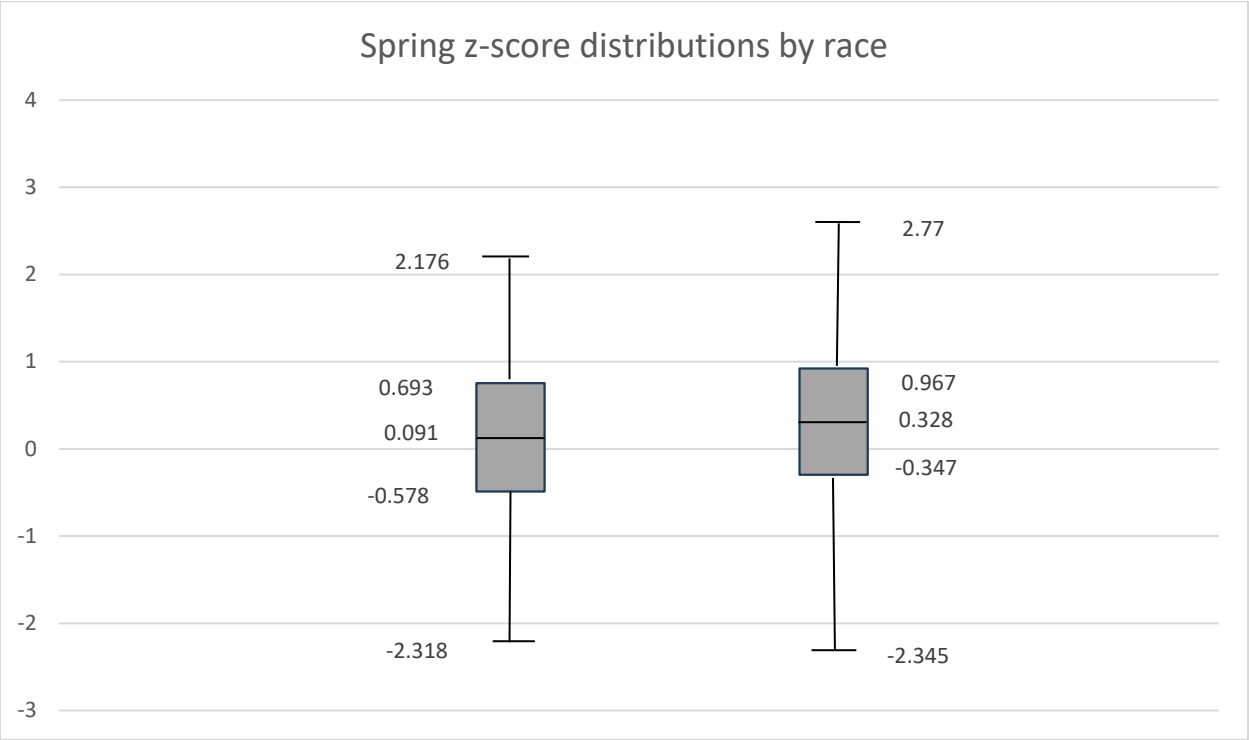


4. Materials for Family Engagement Sent Home

	Black or African American	White
Min.	1.0	1.0
1st Qu.	4.0	6.0
Median	10.0	20.0
Mean	19.0	26.9
3rd Qu.	23.0	37.0
Max.	224.0	157.0



5. Spring Z-score



6. Growth Z-score*

*The negative median growth z-scores should not be interpreted as negative growth; negative z-scores in this case indicate that the samples of students who were included in the matched groups happened to be students who had lower growth scores than the entire group of MRC participants that formed the normative group used for raw score to z-score transformations.