EFFICACY

# Impact of Amplify English Language Arts 6–8

## Amplify ELA Grade 6 students show significantly better reading outcomes in Seminole County, Florida.

by Amplify Education, Inc.

Amplify.

Amplify.

# Table of contents


# ESSA Abstract

Amplify ELA is a 6–8 English Language Arts curriculum designed specifically for the middle grade student, based on extensive research into learning, cognition, and how students develop literacy skills. In the 2018–2019 school year, we explored the effect of Amplify ELA on reading outcomes for students in Grade 6 in Seminole County, FL. Students in grade 6 who used Amplify ELA showed significantly better reading outcomes on the ELA Florida State Exam (FSA) than those who did not use Amplify for their ELA curriculum.

- The study compared 246 6th grade Amplify ELA students to 483 non-Amplify ELA students in a propensity-score weighted sample, controlling for Grade 5 FSA scores and demographic variables including race, gender, free-and-reduced lunch status, English learner status, special education status, and gifted status.
- The average treatment effect (ATE) of Amplify ELA was positive and statistically significant ($p < .001$), with a moderate effect size (ATE = 5.91, $d = .08$).
- Amplify ELA students outperformed non-Amplify students by an average of 6.72 points on the ELA Florida State Exam.
- Approximately 64% of the Amplify students scored at or above Proficient, compared to 48% of the non-Amplify ELA students.

# Introduction

The purpose of this study is to investigate the effect of Amplify's MS-ELA curriculum on end-of-year standardized assessment scores using the potential outcomes framework. Both the average treatment effect on the treated (ATT) and the average treatment effect (ATE) in the population were estimated. The distinction between these two types of effects is as follows:

> The ATT is the average effect that would be seen if everyone in the treated group received the treatment compared with if no one in the treated group received the treatment. In contrast, the ATE is the average effect that would be seen if all individuals (treated and comparison) received the treatment compared with if none of these individuals (treated and comparison) received the treatment.
> (Harder, Stuart, & Anthony, 2010, p. 240)

This study estimated both the ATT and the ATE of the MS-ELA curriculum, as one or the other may be more relevant for a particular school district's RFP. It would be more appropriate to report the ATE if the district is seeking a program for universal adoption across their schools; otherwise, the ATT would be more applicable (Steiner & Cook, 2013).

Approximately 64% of the Amplify students scored at or above Proficient, compared to 48% of the non-Amplify ELA students.

**Non-Amplify ELA students**

48%
Proficient

**Amplify ELA students**

64%
Proficient

# Method

## Sample

This study used SY1819 grade 6 student-level data1 from a convenience sample of four middle schools in Florida's Seminole County Public Schools (SCPS) district. Two of the schools adopted Amplify's MS-ELA curriculum (hereafter referred to as "MS-ELA") in SY1718. For this study, SCPS selected two comparison schools such that each treatment condition was composed of a magnet and a non-magnet school.

Within the SY1819 grade 6 cohort, about 69% of students who had self-selected into advanced ELA courses at the treatment schools were exposed to MS-ELA. Therefore, only students in the comparison schools who had also self-selected into advanced ELA courses were included in the study. None of the students had used an Amplify product the previous year. The final study sample consisted of $N = 246$ MS-ELA students from the treatment schools and $N = 483$ non-MS-ELA students from both treatment and comparison schools, all with complete data.

## Treatment variable

Students were designated as MS-ELA students if they had handed in at least one MS-ELA Learning Object at any time during SY1819. The treatment variable equals 1 for MS-ELA students and 0 for non-MS-ELA students.

## Outcome variable

The outcome variable is grade 6 ELA scale scores on the 2019 Florida Standards Assessment (FSA). The score scale ranges from 259 to 391 and is divided into five performance levels, with the cut score for the proficient level (level 4) set at 339[2].

1 The original dataset included additional cohorts of grade 6 and grade 7 students who were exposed to MS-ELA during the first year of implementation (SY1718) at the treatment schools. Outcome analyses were not conducted on these cohorts due to the problem of quasi-complete separation during propensity score estimation.

2 http://www.fldoe.org/core/fileparse.php/5663/urlt/UnderFSARpt19.pdf

## Covariates

### Grade 5 FSA scores

Students' prior year FSA scores were included as a covariate to adjust for baseline differences in ability between the two groups. Only students with a reported grade level of 5 for the previous year were retained in the study. The grade 5 score scale ranges from 257 to 385, with the proficient cut score set at 336.

### Student demographics

Student demographic variables are listed in Table 5 and include gender, race/ethnicity, free and reduced lunch (FRL) status, English language learner (ELL) status, special education (SPED) status, and gifted status. All variables were dichotomized and dummy coded, and the variable names indicate the non-reference groups (e.g., *male* = 1).

## Propensity score analysis

In the absence of random assignment to treatment condition, propensity score analysis was conducted to balance the two groups on observed baseline characteristics and, thereby, reduce bias in the estimate of the treatment effect. All analyses were performed in R.

### Estimation

Propensity scores were estimated using a logistic regression model with grade 5 FSA scores and student demographics as covariates. The initial model included school fixed effects to account for the clustering of students within schools; however, this model produced propensity scores with an inadequate area of common support. On the other hand, a model without school fixed effects resulted in improved overlap in the distribution of propensity scores between the two groups (see Appendix A). According to Leite et al., (2015) and Li, Zaslavsky, and Landrum (2013), if clustering is accounted for in the outcome model, it can be ignored in the propensity score model; therefore, the final propensity score model is as follows:

$$logit(T_i = 1) = \beta_0 + \beta_1(gr5.FSA_i) + \beta_2(male_i) + \beta_3(white_i) + \beta_4(FRL_i) + \beta_5(ELL_i) + \beta_6(SPED_i) + \beta_7(gifted_i)$$

Figure 1

where the outcome is the natural log of the odds of $T_i = 1$, or exposure to MS-ELA, conditional on the covariates. The propensity score is the conditional probability of $T_i = 1$.

### Weighting

The propensity scores estimated from Equation (fig. 1) were converted to weights via weighting by the odds and inverse probability of treatment weighting. The former leads to estimates of the ATT while the latter estimates the ATE.

With weighting by the odds, students in the MS-ELA group each receive a weight equal to 1. Students in the non-MS-ELA group each receive a weight equal to

$$w_i = \widehat{ps}_i / 1 - \widehat{ps}_i$$

Figure 2

where $w_i$ is the weight for student $i$ and $\widehat{ps}_i$ is the student's estimated propensity score.

The formulas for calculating inverse probability of treatment weights (IPTW) for MS-ELA and non-MS-ELA students, respectively, are

$$IPTW_{i1} = 1/\widehat{ps}_{i1} \qquad IPTW_{i0} = 1/1 - \widehat{ps}_{i0}$$

Figure 3a, 3b

The IPTWs were then normalized by dividing them by the mean of the weights so that they sum to the total sample size (Leite, 2017).

## Covariate balance

The R package *twang* was used to calculate the standardized difference in weighted and unweighted means between the treatment groups on each of the covariates. The criterion for covariate balance used in this study is a standardized difference less than a magnitude of 0.10 (Austin, 2011). Although additional balance diagnostics are often recommended (Stuart & Rubin, 2008), this study relied only on standardized mean differences to assess baseline equivalence, as per What Works Clearinghouse guidelines for quasi-experimental designs[3].

3 https://ies.ed.gov/ncee/wwc/Multimedia/23

## Outcome model

The effect of MS-ELA on grade 6 FSA scores was estimated using a propensity score weighted regression model with school fixed effects to account for the clustering of students within four schools. Fixed effects regression has been shown to perform better than multilevel modeling when the number of clusters is small (McNeish & Stapleton, 2016; McNeish & Wentzel, 2017). The inclusion of school fixed effects in the model also controls for school-level differences, such as school type (magnet or non-magnet).

To further reduce bias in estimates of the treatment effect, the covariates used in the propensity score model were included in the outcome model, along with their interactions with the treatment variable (Leite, 2017; Schafer & Kang, 2008; Shadish & Steiner, 2010). The initial outcome model included all treatment-by-covariate interactions except for the interactions with ELL and SPED because there were only one ELL and three SPED students in the MS-ELA group. The final outcome model, with non-significant interactions removed, is:

$$Y_{ij} = \sum_{k=1}^{7} \beta_k (X_{ki_j}^{c}) + \beta_8 (T_{ij}) + \beta_9 (T_{ij} \times FRL_{i_j}^{c}) + \sum_{j=1}^{4} \delta_j g_j$$

Figure 4

where $Y_{ij}$ is the grade 6 FSA score for student $i$ in school $j$, $\sum_{k=1}^{7}\beta_k(X_{kij}^{c})$ represents the linear combination of the covariates and their main effects, $\beta_8$ is the estimate of the ATT (with weighting by the odds) or ATE (with IPTW), $T_{ij}$ equals 1 for MS-ELA students and 0 for non-MS-ELA students, and $\beta_9$ is the differential effect of MS-ELA on FRL students. Each covariate was centered (as indicated by the superscript $c$) on the mean of the MS-ELA group for estimating the ATT, or on the grand mean for estimating the ATE (Leite, 2017). Finally, the intercept was suppressed to allow for four (instead of three) dummy-coded variables for school membership ($g_j$) and four school-specific intercepts ($\delta_j$) (Clarke, Crawford, Steele, & Vignoles, 2015).

## Effect size

Standardized differences in means (*d*) and their associated 95% confidence intervals (CI) were calculated to assess the practical significance of the ATT and ATE estimated from Equation (fig. 4). Kraft's (2019) guideline for interpreting values of *d* are:

- $d < 0.05$: small effect
- $0.05 \leq d < 0.20$: medium effect
- $d \geq 0.20$: large effect

Point estimates of *d* were obtained by dividing the effect estimate by the pooled standard deviation of the weighted grade 6 FSA scores. The formula for the pooled standard deviation is

$$s_{pooled} = \sqrt{[(n_1 - 1)(s_1^2) + (n_0 - 1)(s_0^2)]/(n_1 + n_0 - 2)}$$

Figure 5

where $n_1$ and $n_0$ are the sample sizes for the MS-ELA and non-MS-ELA groups respectively, and $s_0^2$ and $s_1^2$ are the variances of the propensity score weighted outcome scores (i.e., observed grade 6 FSA score × propensity score weight, see Appendix B). Interval estimates of *d* were calculated using the following formula for standard error (SE) (Konstantopoulos, 2008):

$$SE = \sqrt{[(n_1 + n_0)/(n_1 n_0)] + [d^2/2(n_1 + n_0)]}$$

Figure 6

# Results

## Descriptives

Tables 1 and 2 summarize student performance on the grade 6 ELA FSA. MS-ELA students outperformed non-MS-ELA students by an average of 6.72 points. Furthermore, approximately 64% of the MS-ELA students scored at or above the proficient level, compared to 48% of non-MS-ELA students.

**Table 1. Descriptive Statistics of SY1819 Grade 6 ELA FSA Scores**

| | Mean | SD | Min | Max |
|---|---|---|---|---|
| MS-ELA | 344.49 | 14.50 | 305 | 386 |
| Non-MS-ELA | 337.77 | 16.94 | 283 | 391 |

**Table 2. SY1819 Grade 6 Distribution of Performance Levels on the ELA FSA**

| | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 |
|---|---|---|---|---|---|
| MS-ELA | 0.81% | 5.69% | 29.67% | 44.31% | 19.51% |
| Non-MS-ELA | 4.35% | 21.95% | 25.67% | 34.78% | 13.25% |

Level 1 = *inadequate*; level 2 = *below satisfactory*; level 3 = *satisfactory*; level 4 = *proficient*; level 5 = *mastery*.

Tables 3 and 4 show that, at baseline, MS-ELA students also outperformed non-MS-ELA students on the prior year FSA by an average of 5.06 points, with approximately 68% scoring at or above the proficient level in grade 5, compared to 53% of the non-MS-ELA students.

**Table 3. Descriptive Statistics of SY1718 Grade 5 ELA FSA Scores**

| | Mean | SD | Min | Max |
|---|---|---|---|---|
| MS-ELA | 341.93 | 13.41 | 310 | 385 |
| Non-MS-ELA | 336.87 | 15.24 | 295 | 385 |

**Table 4. SY1718 Grade 5 Distribution of Performance Levels on the ELA FSA**

| | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 |
|---|---|---|---|---|---|
| MS-ELA | 0.00% | 4.88% | 26.83% | 45.93% | 22.36% |
| Non-MS-ELA | 1.24% | 10.97% | 34.99% | 35.40% | 17.39% |

Level 1 = *inadequate*; level 2 = *below satisfactory*; level 3 = *satisfactory*; level 4 = *proficient*; level 5 = *mastery*.

Table 5 shows the frequencies of student demographics in the MS-ELA and non-MS-ELA groups.

**Table 5. Frequencies of Student Demographics (N = 729)**

| | MS-ELA (N = 246) | | Non-MS-ELA (N = 483) | |
|---|---|---|---|---|
| | N | % | N | % |
| Male | 116 | 47% | 248 | 51% |
| White | 150 | 61% | 262 | 54% |
| FRL | 103 | 42% | 224 | 46% |
| ELL | 1 | 0.4% | 7 | 1.4% |
| SPED | 3 | 1.2% | 7 | 1.4% |
| Gifted | 57 | 23% | 101 | 21% |

## Covariate balance

Prior to propensity score weighting, the average difference of 5.06 points on the grade 5 FSA corresponds to a difference of 0.34 standard deviations. The standardized mean differences for white students and ELL students also failed to meet the 0.10 criterion for covariate balance. However, covariate balance was achieved with use of weighting by the odds and IPTW, as shown in Table 6. There were no issues with extreme weights, as weights ranged from 0.83 to 3.14 for the MS-ELA group and from 0.08 to 1.43 for the non-MS-ELA group.

**Table 6. Standardized Mean Differences on Measured Covariates Before and After Weighting**

| | Before PS weighting | Weighting by the odds | IPTW |
|---|---|---|---|
| Grade 5 FSA | 0.34 | -0.04 | 0.04 |
| Male | -0.08 | -0.01 | 0.01 |
| White | 0.14 | -0.01 | 0.01 |
| FRL | -0.09 | 0.01 | -0.01 |
| ELL | -0.10 | 0.01 | -0.04 |
| SPED | -0.02 | 0.00 | 0.00 |
| Gifted | 0.05 | -0.03 | 0.02 |

## Effect of MS-ELA on ELA FSA scores

Results from the fixed effects regression analyses with weighting by the odds and with IPTW are both presented in Table 7. Recall that the estimate of the average treatment effect on the treated (ATT) is obtained with weighting by the odds while the average treatment effect (ATE) is estimated using IPTW. The estimated ATT and ATE were both positive and statistically significant ($p < .001$). The ATT estimate of 6.35 indicates that advanced ELA students who are exposed to MS-ELA are expected to perform better on the FSA than if they had not been exposed. This estimate corresponds to a moderate effect of about 0.10 standard deviations (95% CI [-0.05, 0.25]). The ATE estimate of 5.91 indicates that students randomly drawn from a population of advanced ELA students are expected to perform better with MS-ELA than without. This effect is also moderate ($d = 0.08$, 95% CI [-0.07, 0.24]). Note that these estimates reflect the effect of MS-ELA averaged across student demographic subgroups.

The significant interaction between MS-ELA and FRL status indicates that the effect of MS-ELA was significantly smaller for FRL students than for non-FRL students by 3.70 points for the ATT and 3.50 points for the ATE ($p < 0.05$). For the ATT, non-FRL students who are exposed to MS-ELA are expected to perform better than they otherwise would by 7.90 points ($d = 0.12$, [-0.03, 0.28], while the effect for FRL students was 4.20 points ($d = 0.07$, [-0.09, 0.22][4]. For the ATE, the effect for non-FRL students was 7.48 points ($d = 0.11$, [-0.05, 0.26]) and the effect for FRL students was 3.98 points ($d = 0.06$, [-0.10, 0.21])[5].

4 Centering on the mean of the MS-ELA group for the ATT changed the original dummy coding of the FRL variable to .581 for FRL students and -.419 for non-FRL students.

5 Centering on the grand mean for the ATE changed the original dummy coding of the FRL variable to .551 for FRL students and -.449 for non-FRL students.

The point estimates of d provided above all fell within the range for a medium effect size (i.e., $0.05 \leq d < 0.20$). The interval estimates, however, all contain zero, indicating that the *true* effect of MS-ELA for advanced ELA students may, in fact, be null.

**Table 7. Estimates from Propensity Score Weighted Fixed Effects Regression**

| | Weighting by the odds | | | IPTW | | |
|---|---|---|---|---|---|---|
| | Estimate | SE | *p*-value | Estimate | SE | *p*-value |
| MS-ELA | **6.35**[a] | 1.44 | <0.001 | **5.91**[b] | 1.36 | <0.001 |
| Grade 5 FSA | **0.66** | 0.03 | <0.001 | **0.66** | 0.03 | <0.001 |
| Male | **-2.39** | 0.83 | .004 | **-2.64** | 0.81 | .001 |
| White | 1.59 | 0.87 | .07 | **1.82** | 0.85 | .03 |
| FRL | -0.59 | 1.22 | .63 | -0.97 | 1.19 | .42 |
| Gifted | **6.09** | 1.03 | <0.001 | 6.32 | 1.03 | <.001 |
| ELL | -9.11 | 6.47 | .16 | -5.34 | 4.22 | .21 |
| SPED | -6.92 | 3.71 | .06 | **-7.62** | 3.49 | .03 |
| MS-ELA × FRL | **-3.70** | 1.66 | .03 | **-3.50** | 1.63 | .03 |

a Estimate of the ATT.

b Estimate of the ATE.

Significant effects at alpha = 0.05 are bolded. School fixed effects are not reported here.

# Discussion

This study investigated the effect of Amplify's MS-ELA curriculum on end-of-year standardized assessment scores with a highly select sample of grade 6 students who were enrolled in advanced ELA courses during the 2018–2019 school year. While this sample is not representative of the general population of students or of Amplify's target population, the results may be valuable, nonetheless, for internal purposes.

First, the finding of a significant but *negative* interaction between MS-ELA and FRL indicates that perhaps the program may be less effective for advanced ELA students who are economically disadvantaged. Second, while the overall effect of MS-ELA, averaged across FRL and non-FRL students, is positive and moderate, future impact studies should investigate whether these results hold across different samples.

# Appendix A

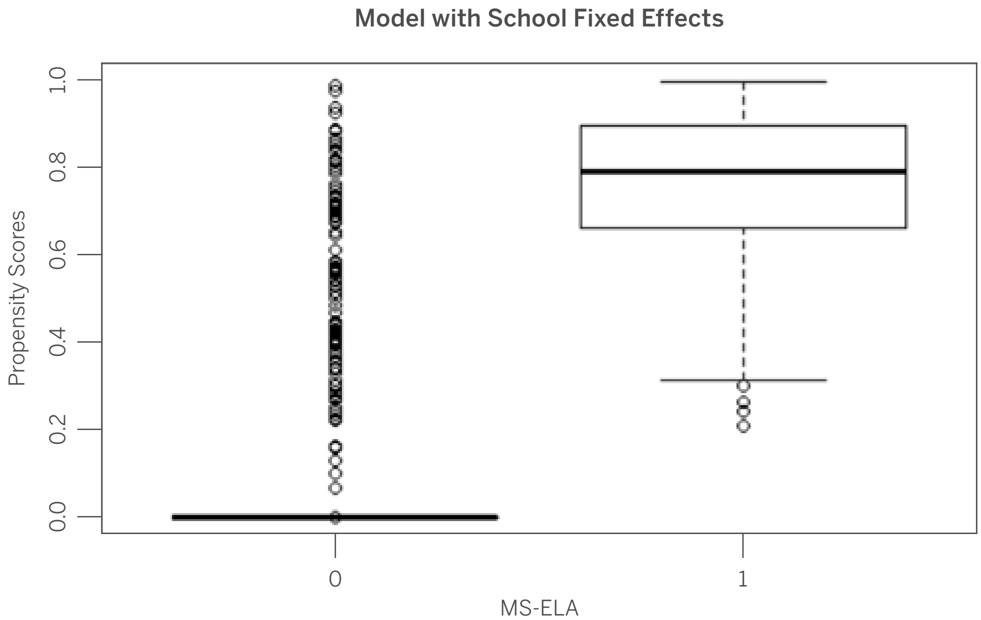


Figure A1. Distributions of propensity scores for the MS-ELA and non-MS-ELA groups from the logistic regression model that accounts for the clustering of students within schools.

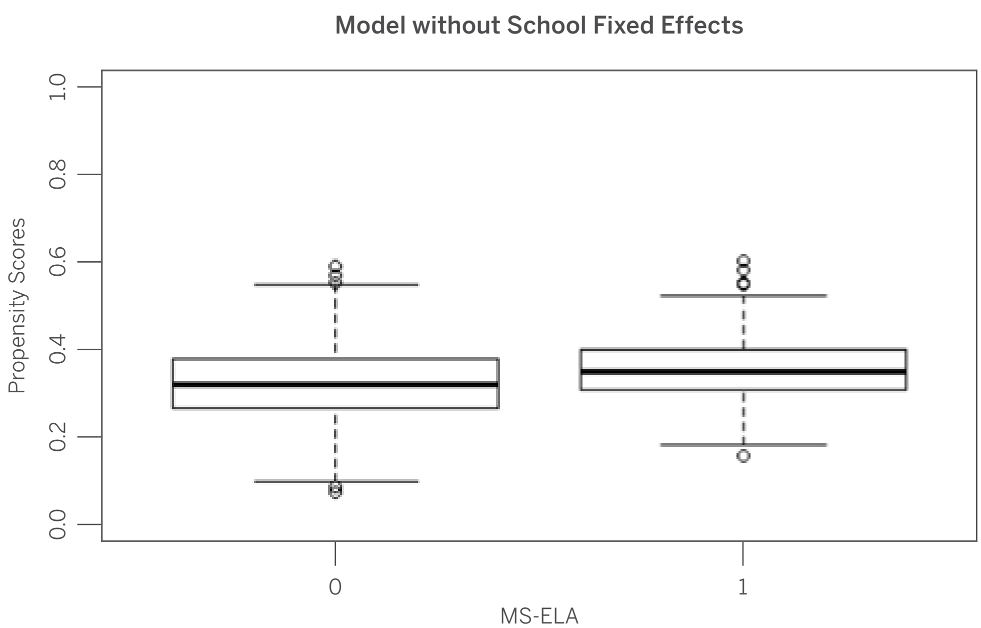


Figure A2. Distributions of propensity scores for the MS-ELA and non-MS-ELA groups from the logistic regression model that ignores the clustering of students within schools.

# Appendix B

**Table B1. Means and Standard Deviations of the Weighted Grade 6 ELA FSA Scores**

| | Weighting by the odds | | IPTW | |
|---|---|---|---|---|
| | **Mean** | **SD** | **Mean** | **SD** |
| MS-ELA | 344.49 | 14.50 | 504.45 | 104.19 |
| Non-MS-ELA | 175.25 | 78.03 | 257.49 | 45.46 |

# References


Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3), 399–424.

Borenstein, M., Hedges, L., & Rothstein, H. (2007). *Introduction to meta-analysis*. Retrieved from www.Meta-Analysis.com

Clarke, P., Crawford, C., Steele, F., & Vignoles, A. (2015). Revisiting fixed- and random-effects models: Some considerations for policy-relevant education research. *Education Economics*, 23(3), 259–277.

Harder, V. S., Stuart, E. A., & Anthony, J. C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological Methods*, 15(3), 234–249.

Konstantopoulous, S. (2008). An introduction to meta-analysis. In J. Osborne (Ed.), *Best practices in quantitative methods* (pp. 177–194). Thousand Oaks, CA: Sage Publications.

Kraft, M. A. (2019). *Interpreting effect sizes of education interventions* (EdWorkingPaper No. 19-10). Retrieved from Annenberg Institute at Brown University: https://www.edworkingpapers.com/ai19-10

Leite, W. (2017). *Practical propensity score methods using R*. Thousand Oaks, CA: Sage Publications.

Leite, W. L., Jimenez, F., Kaya, Y., Stapleton, L. M., MacInnes, J. W., & Sandbach, R. (2015). An evaluation of weighting methods based on propensity scores to reduce selection bias in multilevel observational studies. *Multivariate Behavioral Research*, 50, 265–284.

Li, F., Zaslavsky, A. M., & Landrum, M. B. (2013). Propensity score weighting with multilevel data. *Statistics in Medicine*, 32(19), 3373–3387.

McNeish, D., & Stapleton, L. (2016). Modeling clustered data with very few clusters. *Multivariate Behavioral Research*, 51(4), 495–518.

McNeish, D., & Wentzel, K. R. (2017). Accommodating small sample sizes in three-level models when the third level is incidental. *Multivariate Behavioral Research*, 52(2), 200–215.

Schafer, J. L., & Kang, J. (2008). Average causal effects from nonrandomized studies: A practical guide and simulated example. *Psychological Methods*, 13(4), 279–313.

Shadish, W. R., & Steiner, P. M. (2010). A primer on propensity score analysis. *Newborn and Infant Nursing Reviews*, 10(1), 19–26.

Steiner, P. M., & Cook, D. (2013). Matching and propensity scores. In T. D. Little (Ed.), *Oxford handbook of quantitative methods, volume 1: Foundations* (pp. 237–259). New York, NY: Oxford University Press.

Stuart, E. A., & Rubin, D. B. (2008). Best practices in quasi-experimental designs: Matching methods for causal inference. In J. Osborne (Ed.), *Best practices in quantitative methods* (pp. 155–176). Thousand Oaks, CA: Sage Publications.

For more insights and observations from Amplify's experts, visit **amplify.com/ela**.

Amplify.