1-28-2021

Effects of an Intensive English Program on Students' Math and English Scores

Camila N. Morales
*University of Texas at Dallas*

Tim Sass
*Georgia State University*

Follow this and additional works at: https://scholarworks.gsu.edu/gpl_reports

Part of the Education Policy Commons, Policy Design, Analysis, and Evaluation Commons, and the Public Policy Commons

**Recommended Citation**

This Report is brought to you for free and open access by the Georgia Policy Labs at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Georgia Policy Labs Reports by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.
Effects of an Intensive English Program on Students’ Math and English Scores

January 2021

Camila N. Morales  Tim R. Sass
University of Texas at Dallas  Georgia State University

DISCLAIMER: The results in this policy brief are preliminary and have not undergone the peer review process. All opinions or views expressed herein are those of the authors and do not necessarily represent the opinions or views of any school district partner.
**HIGHLIGHTS**

- We investigate the academic impacts of eligibility for and participation in an intensive English program aimed at English Learners (ELs) with very low English proficiency.
- Our results indicate that ELs who are eligible for an intensive English program have lower English Language Arts (ELA) test scores one year after program eligibility, relative to ELs who receive traditional English as a Second Language (ESL) support. The effect size is equivalent to a widening of the EL/non-EL achievement gap in reading by roughly 20 percent.
- The negative effects of program eligibility are concentrated among older students. Specifically, the impact on students who are first screened in grades 5 to 7 is over five times greater, relative to students screened in grades 3 to 4.
- In subsample analyses, we find that refugee ELs who are eligible for the intensive English program have higher ELA and math test scores one year after program eligibility, which is in contrast to our finding of negative effects overall.

**MOTIVATION AND BACKGROUND**

English Learners (ELs) represent 10 percent of all public-school students in the United States and are the fastest-growing student population in the country (Hussar et al., 2020). In Georgia, as of school year (SY) 2016-17, ELs represented roughly 6.5 percent of public-school students—ranking among the 10 states with the fastest-growing EL enrollment (Batalova & McHugh, 2010; Snyder et al., 2019).

According to the 2019 National Assessment of Educational Progress (NAEP), the fourth-grade reading achievement gap between non-ELs and ELs (33 points) was higher than the White-Black student achievement gap (26 points) and the high-low-income achievement gap (28 points), making ELs among the lowest-performing sub-group of students.

Given their growing prevalence and lag in achievement, understanding the efficacy of policies designed to enhance the educational outcomes of ELs is a priority for public schools around the country.

As part of the school enrollment process, parents are asked to report whether their child speaks a language other than English at home. Responses are then used to identify students who may benefit from additional language support, where final EL classification is largely determined by a screening test. Broadly, EL education policies center on initial EL classification, type of English as a Second Language (ESL) instruction, and recategorization out of ESL support services.

While there is some evidence on the impact of these policies, rigorous studies present mixed results. Some researchers document gains in performance for students who are classified as ELs in early elementary grades, relative to students who score just above the maximum score for EL classification (Pope, 2016; Shin, 2016). Students who are no longer classified as ELs are known as former ELs. Schools commonly monitor the progress of these students for up to three years.

---

1 See nationsreportcard.gov/reading/nation/groups/?grade=4.
2 ESL instruction is also referred to as English for Speakers of Other Languages (ESOL) instruction.
2018), while others find negative effects of being classified as an EL on short- and long-term outcomes (Umansky, 2016; Johnson, 2019). These studies also tend to focus on students who are close to a given threshold of EL classification (i.e., students with relatively high levels of English proficiency). Thus, the findings may not be applicable to students with more limited proficiency in English. Additionally, there is little rigorous evidence on the relative efficacy of different types of ESL instruction, specifically programs aimed at recently-arrived ELs, such as refugees.

For this project, we partnered with a school district in the metro-Atlanta area to study the impact of an intensive English program on the academic achievement of ELs with very low initial English proficiency. As is common of most ESL programs geared toward newcomers (Short & Boyson, 2012), this intensive English program is a short-term intervention designed as a specialized environment for ELs. Students who participate in the program attend classes in the district’s specialized EL instruction center or one of eight satellite schools for up to one year. The primary goal is to introduce students to basic English skills—academic and social—before they transfer to their neighborhood school and begin receiving a combination of traditional ESL instruction and English-only classes. The program is also designed to immerse students in a new culture and education system.

In this project, we explore the impact of eligibility for and participation in an intensive English program relative to receiving traditional ESL instruction. We also estimate whether eligibility for the intensive English program has different impacts across refugee and non-refugee ELs, thereby conducting the first large-scale study on the impact of ESL instruction on refugee ELs.

**RESEARCH QUESTIONS**

1) Relative to traditional ESL instruction, what is the impact of eligibility for an intensive English program on the English Language Arts (ELA) and math test scores of ELs with very low English proficiency?

2) Are there differences in the effect of program eligibility by grade?

3) Are there differences in the effect of program eligibility across refugee and non-refugee ELs?

**DATA AND METHODOLOGY**

We utilize individual-level data from one of the five largest school districts in Georgia. Our sample consists of students in grades 1 through 8 who, at the time of first enrollment in the

---

4 Exceptions include Valentino and Reardon (2015), Steele et al. (2017), and Bibler (2018) who provide evidence on the differential impact of bilingual education.

5 Newcomer students are recently arrived immigrants or refugees. They commonly have little to no knowledge of English and have had interrupted formal education.

6 In this context, traditional ESL instruction refers to a combination of “push-in” or “pull-out” services. “Push-in” refers to a model where ELs remain in their core academic class (e.g., reading) where they receive instruction from their content area teacher and a co-teacher who specializes in ESL instruction. “Pull-out” refers to a model where ELs are taken out of the core academic class and receive ESL instruction for a portion of the school day.

7 For the purposes of this brief, we focus only on the academic effects as measured by ELA and math test scores. It is likely that participation in the program can also impact students’ non-cognitive and social-emotional skills.

8 The district serves a county where a large proportion of refugees who are resettled in the state reside. As of 2017, roughly 4 percent of the students in the district self-identified as refugees.
district, report speaking a language other than English at home and are screened for EL classification. Our sample includes data for SY 2007-08 through SY 2017-18.

Our sample consists of 4,479 students of which 2,545 (57 percent) correspond to those eligible to enroll in the intensive English program. These students come from diverse backgrounds and are likely to live in low-income households. Specifically, the ethnic/racial composition is 40 percent Asian, 30 percent Hispanic, 23 percent Black, and 14 percent White. Among students eligible for the program, 86 percent are eligible for Free or Reduced-Price Lunch (FRL) and 52 percent are self-reported refugees.

Students are screened for EL classification using the WIDA Screener, an English proficiency assessment that measures four language domains, and receive a score for each domain and an overall composite proficiency level ranging from 1.0 to 6.0 (in increments of 0.1). The composite score and grade in which a student was screened are used to determine initial EL classification and type of ESL instruction. Students who are screened in grades 3 to 8 and score between 1.0 and 1.9 are classified as ELs and are eligible to enroll in an intensive English program for up to one year. Alternatively, students who are screened in grades 1 or 2 and score between 1.0 and 4.9 and students screened in grades 3 to 8 who score between 2.0 and 4.9 are classified as ELs and receive traditional ESL instruction in their home school. Lastly, students who score 5.0 or above on the WIDA screener, regardless of grade level, are not classified as ELs. See Table 1 for a summary of these classification criteria.

<table>
<thead>
<tr>
<th>Grade of EL Screening</th>
<th>Initial WIDA Score</th>
<th>EL and ESL Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>1.0-4.9</td>
<td>EL and Traditional ESL Instruction</td>
</tr>
<tr>
<td>3-8</td>
<td>1.0-1.9</td>
<td>EL and Intensive English Program</td>
</tr>
<tr>
<td>3-8</td>
<td>2.0-4.9</td>
<td>EL and Traditional ESL Instruction</td>
</tr>
<tr>
<td>1-8</td>
<td>5.0-6.0</td>
<td>Not EL</td>
</tr>
</tbody>
</table>

Notes. Students in grades 9-12 are subject to the same screening and ESL classification criteria as those screened in grades 3-8. Students in kindergarten are screened using a different test instrument and criteria.

The program eligibility criteria outlined above allow us to estimate the impact of program eligibility and participation using two analytical strategies: “difference in differences” (DiD) and a “fuzzy regression discontinuity” (RD).

In the DiD approach, we estimate the effect of program eligibility by comparing the ELA and math test scores of students who met both criteria for Intensive English program eligibility (having an initial WIDA score of 1.9 or less and

---

9 We further limit our sample to students who are first screened in grades 1 to 7 in order to observe student outcomes up to grade 8. In addition, we omit students who are never classified as ELs after initial screening.

10 “Hispanic” is defined as ethnicity, separate from race. Other racial groups not listed include Indigenous/Native American and Multiracial.

11 Our unique data allow us to directly identify refugees apart from other foreign-born students by using self-reported information collected at the time of school registration. The sample district does not track students’ immigration status. Rather, the data on refugee identification is entirely voluntary and self-reported, and it is mainly used to offer targeted services that can benefit school integration among refugee students.

12 The four language domains are listening, speaking, reading, and writing.

13 These students are omitted from our sample to limit the control group to ELs who receive traditional ESL instruction.
enrollment in grades 3-8) with students who only met the WIDA criterion (i.e., had a WIDA score of 1.9 or less but were in grades 1 or 2 when they were initially screened). Because there is considerable diversity in English skills within both the fully-eligible and partly-eligible groups, we control for each student’s initial WIDA score in the analysis. For instance, we compare the ELA test scores of two students with a WIDA score of 1.0 where one student is first screened in grade 2 (and thus ineligible for the program) and another is screened in grade 3 (and thus eligible for the program). While both students have the same level of initial English proficiency, each faces a different type of ESL instruction based solely on their grade of EL screening.

In the RD approach, we utilize the fact that, among students in grades 3-8, program eligibility is based on having an initial WIDA score of 1.9 or less. We use this to compare the achievement outcomes of students who fall within a relatively small window of the eligibility cutoff. This approach enables us to estimate the impact of program participation among a subset of students with similar initial English proficiency. An advantage of the RD method is that, unlike the DiD approach, we can compare students of similar ages, thus avoiding any spurious correlation between language proficiency and age of arrival. However, one disadvantage of this method is that there are not large numbers of students with scores right at the eligibility cutoff. Thus, the smallest window we can employ compares outcomes for students with initial WIDA scores of 1.5-1.9 to those with initial scores in the range of 2.0-2.3. As in the DiD approach, however, we statistically control for variation in initial WIDA scores within these bands.

RESULTS

RESEARCH QUESTION #1

Results from the DiD specification provide evidence that EL students who are eligible for the intensive English program have lower ELA achievement one year after program eligibility, relative to EL students who receive traditional ESL services in their neighborhood school. We find no evidence that program eligibility impacts math test scores on average. Specifically, ELs who are eligible for the intensive English program score 0.17 standard deviations lower in ELA than do EL students who do not qualify due to their grade level or initial screening score (see Figure 1). The effect size can be interpreted as a widening of the EL/non-EL achievement gap in reading by roughly 20 percent. This is a large and significant result, higher than most negative estimates of the

14 Our preferred specification also includes controls for demographic characteristics, school fixed effects, and grade fixed effects. We cluster all errors at the school level to account for within-school correlation in test scores.

15 We estimate that each 0.1 reduction in the initial WIDA score lowers the later ELA score by 0.04 standard deviations and the later math score by 0.03 standard deviations.

16 In the fuzzy regression discontinuity design, we limit our sample to students screened in grades 3 to 7 between SY 2014-15 and SY 2016-17.

17 In other words, we compare students who score at proficiency level 1 (or “Entering”) and are over halfway toward achieving proficiency level 2 with students who score just above proficiency level 2 (or “Emerging”).

18 Our findings are robust to modifications in our preferred specification that include year fixed effects, school-by-year fixed effects, controls for the month in which students are screened for EL classification, and variations in the variable used to cluster the standard errors.

19 See the related academic paper for a full discussion of the math results, including the estimates from the regression discontinuity analysis.

20 The EL/non-EL achievement gap in reading is computed using the difference in NAEP test scores among fourth graders. Alternatively, the program eligibility effect can be interpreted as the widening of the EL/non-EL reading achievement gap in eighth grade by 15 percent.
impact of EL classification in kindergarten on ELA scores (Umansky, 2016).

As one may suspect, a possible explanation for the findings is that we might simply be estimating the “effect” of being older (i.e., first screened for EL classification in grades 3 to 7) and having low initial English proficiency.\(^{21}\) In other words, our estimates might reflect a difference in achievement among early and late EL arrivals, irrespective of exposure to the intensive English program.\(^{22}\) Moreover, by comparing students across grades 1-2 and 3-8, our results are likely confounded by age and grade-level developmental and instructional factors (differences in curricula, classroom instructional models, etc.) that could explain why younger students showed more progress in math and ELA over time.\(^{23}\)

---

\(^{21}\) Another reason that may explain our negative results is student mobility. Due to the structure of the intensive English program, students may exit the program and transfer to their neighborhood school mid-year, which can have a negative impact on achievement (Schwartz et al., 2016). We check whether our results are robust by limiting our sample to students who are screened in June to August and are less likely to move. Results from this exercise are qualitatively the same as our baseline estimates, indicating that it is unlikely that mobility is driving the negative findings.

\(^{22}\) Previous literature shows significant differences in English proficiency by children’s age at arrival to the United States (Bleakley & Chin, 2004).

\(^{23}\) Due to data limitations, we are unable to control for these factors.
Considering the limitations of the DiD approach, we turn to the RD analysis, which enables us to compare achievement outcomes across students who are screened in the same grade but whose initial WIDA scores fall within a relatively narrow range around the small window of the program eligibility cutoff. While we are unable to draw robust conclusions from this approach, most estimates align with the previous DiD results. We find that students who participate in the program, as a result of their initial WIDA score, have lower ELA scores one year after program enrollment. Notably, we find no evidence of positive program enrollment effects.

What may explain these negative impacts? Two potentially unintended consequences of participating in the intensive English program are a delay in access to general education resources and core-content classes. We indirectly examine this hypothesis by comparing the math effects resulting from program eligibility relative to participation. Simply put, if math is a core content class to which ELs in the intensive English program have delayed access, then we expect the impact on math achievement to be greater among those who participate compared to those who are simply eligible. In fact, we find that Intensive English program participation is associated with a reduction in math test scores of up to 0.5 standard deviations among participants. In addition, prior studies find that EL classification can lead to differential exposure to school resources in elementary and middle school (Umansky, 2018). Thus, there is suggestive evidence from our analysis and prior literature that the negative results may be driven by delayed access to general education resources and content.

**RESEARCH QUESTION #2**

We further investigate whether there are differences in the effect of program eligibility across grade levels. Results from our DiD specification suggest that the negative findings are concentrated among older students. Specifically, we estimate large negative impacts of program eligibility among students who are first screened for EL classification in grades 5 to 7. Results from the RD approach also align with these findings. We find smaller and statistically insignificant impacts among students who are screened in grades 3 and 4 (see Figure 2).

Overall, our results align with previous research showing that EL classification is negatively associated with student outcomes in middle school (Umansky, 2018).

---

24 This alleviates concerns that arise from comparing students of different ages in the DiD approach.
25 Our preferred fuzzy RD specification limits the sample from SY 2014-15 to SY 2017-18. Enrollment compliance was highest at 81 percent on average during this time.
26 While we limit our sample to students whose WIDA scores fall within 0.4 units of the program eligibility cutoff, it is likely that these small numerical differences reflect larger differences in English language ability. See the WIDA Interpretive Score Guide and Can Do Descriptors for a detailed explanation on the WIDA English proficiency levels. We account for differences in proficiency levels by controlling for individual WIDA scores; however, due to data limitations, qualitative differences are accounted for in our model.
27 RD results are sensitive to the choice of bandwidth and level of clustering and thus must be viewed with some caution.
28 Setren (2019) finds that non-targeted education interventions can have large positive effects among ELs, especially those with low initial English proficiency.
29 As an exception, we find positive program effects among students screened for EL classification in grade 7; however, these results are not robust to the level of clustering.
Figure 2. Estimated Effect of Eligibility for an Intensive English Program Relative to Traditional ESL Instruction, by Grade of EL Screening

Notes. The height of each bar indicates the estimated short-term effect of program eligibility in standard deviation units, relative to the achievement of students who receive standard ESL support. The vertical lines indicate the 95 percent confidence intervals. Each bar shows the estimated effect of eligibility by grade of EL screening. These results are obtained from regression models that include controls for demographic characteristics, time-invariant school characteristics, and test grade. Test scores are measured one year after program eligibility.

RESEARCH QUESTION #3

In contrast to the results from the full sample, we find large positive impacts of program eligibility on achievement among the self-reported refugee subsample. Specifically, results from the DiD specification indicate that refugee students who are eligible for the intensive English program have higher ELA test scores by 0.41 standard deviations, relative to refugee students who can only receive traditional ESL instruction in their neighborhood school.\(^{30}\) This effect size is equivalent to reducing the EL/non-EL reading achievement gap by 43 percent. We also find positive and large effects on math test scores. Further, we find large and positive effects on ELA achievement across all grade levels (see Figure 3).

We note that our RD results for the refugee subsample do not support our DiD findings. Rather, with the RD method, we estimate negative but statistically insignificant effects.

Figure 3. Estimated Effect of Eligibility for an Intensive English Program Relative to Traditional ESL Instruction, by Refugee Status and Grade of EL Screening

Notes. The height of each bar indicates the estimated short-term effect of program eligibility in standard deviation units, relative to the achievement of students who receive standard ESL support. The vertical lines indicate the 95 percent confidence intervals. Each bar shows the estimated effect of eligibility by grade of EL screening and refugee status. These results are obtained from regression models that include controls for demographic characteristics, time-invariant school characteristics, and test grade. Test scores are measured one year after program eligibility.

---

\(^{30}\) Our DiD results from the refugee subsample are robust to limiting our sample to students in grades 2 and 3 only, students whose WIDA score falls within one unit of the eligibility cutoff, and students who enroll in June to August and are less likely to experience a move mid-year.
POLICY CONSIDERATIONS

We find that refugee students who are eligible for an intensive English program have higher test scores in ELA and math one year after program eligibility, relative to refugee ELs who receive traditional ESL support in their neighborhood school. These findings are robust to several variations of our DiD analytical approach.

Among non-refugee ELs, students who are in grades 5 and above and eligible for the intensive English program experience lower ELA achievement than do non-refugee ELs in the same grades who are not eligible for the program and receive traditional EL support. While we are unable to uncover the mechanisms behind this differential impact, these results suggest that the overall benefits of the intensive English program could be enhanced by expanding the program to grades below grade 3. Of course, any potential modification to existing policies would need to consider the costs of the intensive English program relative to the expense of providing traditional EL support. Other non-academic factors, such as socialization and access to non-EL peers, may be a consideration as well.

In addition, for students who are screened in grades 5 and above, policymakers may want to consider altering the eligibility criteria for the intensive English program in order to place greater emphasis on serving students with very low initial English proficiency (of which a large proportion are refugees). However, it is likely that some form of specialized English instruction will still be needed for non-refugee immigrant ELs in this older age group.

While our work has provided important new insights into the delivery of ESL services, additional research into longer-run impacts on test scores and outcomes beyond test scores is warranted.
REFERENCES


ABOUT THE AUTHORS

Camila N. Morales is an assistant professor of economics at the University of Texas at Dallas. She earned her Ph.D. in economics from Georgia State University and was a graduate research assistant with the Georgia Policy Labs. Her research interests lie at the intersection of education economics, labor economics, and immigration policy. Her current work focuses on the educational outcomes of refugee and immigrant students, second language learners, and their peers. Prior to her Ph.D., she earned a B.S. in economics and a minor in mathematics from Georgia State University.

Tim R. Sass is a Distinguished University Professor in the department of economics at Georgia State University and the W.J. Usery Chair of the American Workplace in the Andrew Young School of Policy Studies. He is also the faculty director of the Metro Atlanta Policy Lab for Education (MAPLE). Specific areas of interest include teacher labor supply, the measurement of teacher quality, and school choice. His work has been published in numerous academic journals and has been supported by grants from the U.S. Department of Education, the Gates Foundation, the Smith Richardson Foundation, Arnold Ventures, and the Spencer Foundation. He has acted as a consultant to school systems in New York City, Washington, D.C., Charlotte, NC, the state of Florida, and the state of New York. He is also a senior researcher at the Center for Analysis of Longitudinal Data in Education Research (CALDER).

ABOUT GEORGIA POLICY LABS

The Georgia Policy Labs (GPL) is a collaboration between Georgia State University and a variety of government agencies to promote evidence-based policy development and implementation. Housed in the Andrew Young School of Policy Studies, GPL works to create an environment where policymakers have the information and tools available to improve the effectiveness of existing government policies and programs, try out new ideas for addressing pressing issues, and decide what new initiatives to scale. The goal is to help government entities more effectively use scarce resources and make a positive difference in people’s lives. GPL has three components: The Metro Atlanta Policy Lab for Education works to improve K-12 educational outcomes; the Career & Technical Education Policy Exchange focuses on high-school-based career and technical education in multiple U.S. states; and the Child & Family Policy Lab examines how Georgia’s state agencies support the whole child and the whole family. In addition to conducting evidence-based policy research, GPL serves as a teaching and learning resource for state officials and policymakers, students, and other constituents. See more at gpl.gsu.edu.