Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 24-1085

Leveling the Playing Field: Default Policy and its Effects on English Learner Reclassification

Caroline Bartlett
Michigan State University

Joseph R. Cimpian
New York University

Madeline Mavrogordato
Michigan State University

Reclassification, the process by which English learner (EL) students exit EL classification, often determines ELs' access to mainstream academic coursework. While existing research finds that many students who demonstrate English proficiency do not reclassify, few studies evaluate policies that effectively reclassify eligible students. This study examines the impact of shifting reclassification responsibility from school districts to the state in Michigan. Using a difference-in-regression discontinuities design, we find that state-level responsibility increases reclassification rates by 35 percentage points compared to district-level responsibility. The effects are larger for Spanish speakers, indicating that state procedures may reduce linguistic bias. Our findings contribute to the literature on default policies in K-12 education and provide evidence on policies that promote equity in EL education.

VERSION: November 2024

Suggested citation: Bartlett, Caroline, Joseph R. Cimpian, and Madeline Mavrogordato. (2024). Leveling the Playing Field: Default Policy and its Effects on English Learner Reclassification. (EdWorkingPaper: 24 -1085). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/z4s2-c474

Leveling the Playing Field: Default Policy and its Effects on English Learner Reclassification

Caroline Bartlett, Michigan State University

Joseph Cimpian, New York University

Madeline Mavrogordato, Michigan State University

Abstract: Reclassification, the process by which English learner (EL) students exit EL classification, often determines ELs' access to mainstream academic coursework. While existing research finds that many students who demonstrate English proficiency do not reclassify, few studies evaluate policies that effectively reclassify eligible students. This study examines the impact of shifting reclassification responsibility from school districts to the state in Michigan. Using a difference-in-regression discontinuities design, we find that state-level responsibility increases reclassification rates by 35 percentage points compared to district-level responsibility. The effects are larger for Spanish speakers, indicating that state procedures may reduce linguistic bias. Our findings contribute to the literature on default policies in K-12 education and provide evidence on policies that promote equity in EL education.

Equal access to educational opportunities and the equitable application of policy are fundamental principles in modern education policy, as exemplified by the Every Student Succeeds Act of 2015 (ESSA). Implemented in the 2016-17 school year, ESSA represents a shift towards standardizing student benchmarks and away from local discretion. Understanding the effects of policy decisions aimed at increasing standardization and uniformity in processes is critical for assessing if this trend leads to more equitable decision-making. On the one hand, discretionary policies can allow for flexibility and nuanced decision-making regarding the education of individual students. On the other hand, they can also allow for differential treatment of similarly situated students, which may reflect biases against particular groups of students.

The trend toward education policy standardization is particularly true for multilingual students classified as English learners (ELs), who make up nearly ten percent of the US K-12 student population and attend schools in almost every district (National Center for Education Statistics, 2023). While classified as ELs, students are legally entitled to linguistic support services and additional resources to help them meet academic content standards as they develop English proficiency (ESSA, 2015). EL classification also often results in students being educated in different settings or receiving different educational services than non-ELs. Once ELs demonstrate English proficiency[1], they qualify to exit EL services through a process called "reclassification." Reclassification is significant because it ends schools' legal obligation to provide linguistic support services and routes students into mainstream academic coursework that is often unavailable to students classified as ELs. If timed appropriately, reclassification should result in a smooth transition to mainstream academic content and ensure students receive access to a developmentally appropriate educational setting (Robinson, 2011).

Once an EL student qualifies for reclassification, their state or school district must administratively formalize their exit from EL status using either a manual or automatic reclassification process. Many states require districts to identify and manually reclassify students who have met reclassification criteria. Manual procedures often result in many eligible students remaining classified as ELs (e.g., Cimpian et al., 2017; Estrada & Wang, 2018; Mavrogordato & White, 2017). Further, reclassification outcomes under manual procedures are uneven, with eligible Spanish speakers being notably less likely to reclassify than students speaking other home languages (Mavrogordato & White, 2017; Umansky et al., 2020). This can lead to restricted opportunities to learn in academically rigorous settings and, in many cases, to students being denied access to mainstream core coursework (Estrada & Wang, 2018). A handful of states implement automatic reclassification procedures, in which students who meet reclassification criteria are automatically reclassified via state administrative data systems, thus removing the manual determination by districts. However, the ability of automatic procedures to (1) increase reclassification rates of eligible students and (2) alleviate linguistic biases in reclassification decision-making processes remains unexplored.

A state's choice to implement a manual or automatic reclassification procedure has meaningful consequences because it sets the default approach to reclassification. In general, automatic processes assume that the qualifying student will reclassify unless an extenuating circumstance indicates the student is not ready. Manual procedures, on the other hand, require districts to decide to reclassify eligible students actively. In other words, automatic reclassification procedures generally require districts to "opt out" of reclassifying students, while manual procedures require districts to "opt in." Default policies may increase uptake of a desired outcome (Jachimowicz et al., 2019), making them potentially powerful tools in states new to serving EL students, especially in contexts where educators have less experience identifying whether students will benefit from reclassification or in contexts where capacity is limited.

In light of ESSA's (2015) call for standardization of reclassification, identifying processes that effectively reclassify eligible students is vital for policymakers to facilitate equitable reclassification outcomes for ELs. To better understand the efficacy of different reclassification procedures for reclassifying eligible students, we present estimates of the effect of shifting statewide reclassification policy in Michigan. Michigan is a new immigrant diaspora state and, like most other states, serves a rapidly growing EL population. In the 2019-20 school year, Michigan shifted from a manual (school district responsibility) to an automatic (state responsibility) reclassification process. We use a difference-in-regression discontinuities design to estimate (1) how shifting the default reclassification procedure in Michigan impacts eligible EL students' likelihood of reclassifying and (2) how the magnitude of impact varies across subgroups of ELs.

We find that shifting from manual to automatic reclassification procedures in Michigan results in significant and meaningful default effects on reclassification rates for eligible students. In Michigan, eligible students are over 35 percentage points more likely to formally reclassify under automatic reclassification than manual reclassification. We also find preliminary evidence of larger effect sizes for ELs reporting Spanish as a home language, suggesting that automatic procedures have the potential to "level the playing field" for groups of eligible students who may experience bias under manual reclassification procedures. Given ESSA's (2015) call to standardize reclassification policy, state education agencies may look towards procedures like automatic reclassification to facilitate standardization of the reclassification process and equitable outcomes for students.

Our analysis contributes to the literature on default policies and EL education policy. First, we present what we believe to be the first evaluation of a default policy in K-12 education. The increasing prevalence of administrative data sets and the goal of closing education opportunity gaps have resulted in a rise in default policies in education, particularly around enrolling students in advanced coursework. For example, legislators in Texas recently adopted a policy that requires 5th graders who performed in the top 40 percent on a standardized math assessment to automatically be enrolled in advanced math for sixth grade (Richman, 2023). While researchers have examined the effectiveness of default policies in other fields (e.g., automatic voter registration [Garnett, 2022] and automatic enrollment into retirement savings [Madrian & Shea, 2001]), we have little information on how default policies function in K-12 education.

Additionally, we contribute to the literature on EL reclassification by evaluating the ability of two commonly used reclassification procedures to reclassify eligible students. Prior research identifies reclassification as a critical but elusive juncture in EL students' educational trajectory, with a substantial population of reclassification-eligible students remaining classified as ELs and lacking access to mainstream coursework. However, no studies have investigated whether shifting to an automatic reclassification process leads to more students who meet reclassification criteria exiting EL status. Moreover, existing research on EL reclassification uses data from before ESSA's (2015) implementation, which included a push for greater standardization of state reclassification procedures. This study uses post-ESSA data, which may apply to states' current reclassification contexts.

Background

In what follows, we discuss the prior literature on the significance of timely EL reclassification and the factors influencing reclassification, as well as the role of default policies. This review will examine the adverse effects of prolonged EL status for students demonstrating English proficiency, the variability in state and district reclassification policy implementation, and the complexities of implementing manual reclassification procedures. We then discuss the impact of default policies across various contexts and highlight their potential as effective policy tools in education, particularly for improving EL reclassification outcomes.

English Learner Reclassification

Timely reclassification is vital as prolonged EL status may inadvertently lead to adverse social and academic consequences for students. For those with relatively advanced English proficiency, the EL label itself and resulting barriers to core academic coursework can adversely affect academic achievement (Umansky, 2016). EL classification has also been linked to restricted access to core academic content and limited opportunities to interact with non-EL peers (Umansky, 2018), limited access to honors and college preparatory coursework at the secondary level, higher dropout rates, and decreased rates of college enrollment (e.g., Carlson & Knowles, 2016). Ultimately, EL services are intended to benefit students as they develop English proficiency, and reclassification upon demonstrating English proficiency ensures students have equitable access to challenging and appropriate coursework.

State-specific reclassification criteria and processes determine whether and when EL students reclassify (Morales & Lepper, 2024). Although states establish standardized criteria for reclassification eligibility, meeting reclassification criteria does not guarantee that a student will reclassify. Existing studies have found that in many districts, a substantial number of eligible students are not reclassified due to variations in local policy interpretation and implementation

(Cimpian et al., 2017; Estrada & Wang, 2018; Mavrogordato & White, 2020). For example, in a mixed-methods case study of reclassification procedures and outcomes in two California school districts, Estrada and Wang (2018) report that while one district reclassified nearly all students who met the criteria, another district reclassified only 67% of eligible students. Ultimately, the authors conclude that several factors drive differences in reclassification likelihood for eligible students, including excessive administrative burden on school district leaders to formally reclassify a student (e.g., requiring signature forms from parents, errors in applying criteria, lack of district monitoring of students and procedures) and staff perceptions of the benefits or drawbacks of reclassification compared to remaining EL-classified for specific students.

To better understand between-district variation in reclassification outcomes for students who meet reclassification criteria, Cimpian and colleagues (2017) compare two states using a regression discontinuity design. Across both states, the authors report substantial between-district and between-grade variation in how meeting test-based reclassification criteria predicts a student's likelihood of reclassifying. For example, among districts with below-average reclassification rates in one state, meeting the reclassification criteria did not influence a student's likelihood of reclassifying. In contrast, meeting the same criteria in districts with above-average reclassification rates significantly increased students' likelihood of reclassifying. These studies suggest a complex interplay between reclassification criteria and outcomes, and they indicate that several factors, including district characteristics, district-level policies, procedural burden, and staff attitudes or knowledge, contribute to variation in reclassification likelihood among students who meet the established criteria.

Other studies identify heterogeneity in eligible students' likelihood of reclassification based on their grade level. For example, Robinson (2011) finds that as EL students in California progress through school, they are less likely to reclassify upon meeting the criteria. Specifically, the reclassification rate for fourth graders meeting all criteria was 91%, compared to 64% for tenth graders, signifying greater teacher discretion in the reclassification process in high school. In contrast, using data from one large California school district, Umansky and Reardon (2014) report that 12% of ELs who meet test-based reclassification criteria do not reclassify in 5th grade. However, more ELs reclassify in 11th grade than qualify, suggesting educators are more likely to perceive an urgent need to reclassify students in later grades (Umansky & Reardon, 2014).

Finally, educators' perceptions of EL students from different racial or ethnic backgrounds may also inform their reclassification decisions. For example, Umansky and colleagues (2020) report a higher likelihood of reclassification among Chinese-origin than Latinx ELs, even when Chinese-origin ELs do not meet reclassification criteria. Furthermore, Mavrogordato and White (2017) find that reclassification-eligible students who speak languages other than Spanish are five percentage points more likely to be reclassified than their Spanish-speaking peers. In summary, findings regarding heterogeneity in the reclassification rates of eligible students suggest that the reclassification processes can be influenced by various factors, including grade level, teacher discretion, and perceived urgency, highlighting the complexity of manual reclassification decisions. Given these collective findings, researchers suggest that reliance on more objective and standardized reclassification policies can improve discrepancies in reclassification for eligible students (Estrada & Wang, 2018; Okhremtchouk et al., 2018).

Default Policies

An automatic or default policy is a standard or predetermined choice automatically applied to an action if no alternative option is chosen (Herd et al., 2013). Policymakers establish defaults that reflect their preferred choice, and individuals must take deliberate action to opt for something different. Ultimately, policymakers implement these preselected choices because they are a subtle but powerful way to influence decisions and increase the uptake of a preferred option (Jachimowicz et al., 2019).

In practice, default policy options provide straightforward means of implementation. For example, a well-known default policy is organ donation registration. This policy is often implemented by stating, "You are currently registered as an organ donor. Do you *not* want to be an organ donor?" By default, individuals are enrolled to be organ donors and must take deliberate action if they wish to opt out (Johnson & Goldstein, 2003). Despite their simplicity, default policies substantially influence individuals' decision-making. Defaults have proven to be practical policy tools across a wide range of social issues, increasing organ donation rates (e.g., Johnson & Goldstein, 2003), voter registration (e.g., Garnett, 2022), and retirement savings (e.g., Madrian & Shea, 2001). Generally, the literature on default effects finds that decision-makers are likelier to choose the default option than an alternative (Jachimowicz et al., 2019).

Research studying the efficacy of default effects in education policy is small but growing and, thus far, has focused on higher education rather than K-12 education policy. Behlen and colleagues (2023) investigate the impact of defaults on universities' final exam sign-up procedures in contexts that require students to register for final exams. They find that under default enrollment in final exams, students are likelier to participate in and succeed in final exams, underscoring the effectiveness of defaults in education. Additionally, Cox and colleagues (2020) provide insights into the factors that affect student loan borrowers' decisions to opt for income-driven repayment plans instead of other loan repayment plans that may increase their chances of defaulting. Their findings highlight the importance of defaults and information provision in decision-making processes in this area and demonstrate that implementing income-driven repayment plans as the default can substantially decrease students' likelihood of choosing riskier repayment plans. These studies indicate that defaults can be promising policy levers to increase students' short- and long-term outcomes.

Specific to EL reclassification policy, automatic procedures require districts to opt out students who meet reclassification criteria, while manual procedures require districts to actively opt in students to be reclassified. Many states currently employ a manual reclassification procedure in which school districts are responsible for identifying and reclassifying eligible students in district data systems. A handful of states have moved to an automatic reclassification procedure, which leverages state administrative data systems to automatically identify and reclassify eligible students upon receiving standardized test scores. Automatic procedures eliminate the burden on districts to identify and complete reclassification paperwork for eligible students, factors that contribute to disparities in the reclassification of eligible students (Estrada & Wang, 2018). In addition, they implicitly convey to districts that reclassification is the status quo for students who meet the criteria.

Consistent with existing literature examining default policy effects, we hypothesize that an automatic reclassification policy will increase reclassification rates among eligible students compared to a manual reclassification policy. Additionally, because automatic reclassification is based solely on students' test scores and completed via state data systems, we anticipate that automatic policy will close gaps in reclassification rates across subgroups of eligible EL students whom prior research has identified as less likely to reclassify when eligible, particularly Spanish speakers (Mavrogordato & White, 2017; Umansky et al., 2020).

Michigan Policy Context

Michigan serves a linguistically diverse and growing number of EL students. In the past ten years, the Michigan EL population has nearly doubled, and in 2023, ELs comprised 98,771 students, roughly 6.9% of Michigan's K-12 student population.

Following federal requirements that states annually assess ELs' English proficiency growth, Michigan and 40 other states use the WIDA ACCESS 2.0 (hereafter, WIDA) English proficiency assessment to evaluate ELs. All Michigan ELs take the WIDA assessment and other statewide standardized tests each spring. WIDA consists of four domains (listening, speaking, reading, and writing). Students receive a scale score for each domain and an overall scale score ranging from 100 (lowest score) to 600 (highest score; WIDA, 2024). Ultimately, scale scores correspond to an interpretive "proficiency level" score ranging from 1.0 (low) to 6.0 (high; WIDA, 2024). Scale score interpretations vary across grades, while proficiency levels can be compared across grades. We detail the history of Michigan's reclassification criteria and procedures below.

Manual Reclassification, 2016-17 through 2018-19

Before the 2019-20 school year, Michigan required students to meet multiple test-based criteria to qualify for reclassification. In addition, districts were responsible for manually reclassifying students in the state data reporting system. To qualify for reclassification, students needed to attain (1) a WIDA overall score of 4.5, (2) a WIDA reading domain score of 4.0, (3) a WIDA writing domain score of 4.0, and (4) score "proficient" in a locally chosen reading assessment. Regarding the locally chosen reading assessment criterion, districts were permitted to choose from several pre-approved options (e.g., NWEA, AIMSWeb, DIBELS Next, iReady Diagnostic, Star Early Literacy, and the state standardized reading assessment, M-STEP) and define the minimum score required to be considered "proficient," creating variation across districts in the score needed to demonstrate proficiency and the overall difficulty of the chosen assessment (Personal communication with MDE, 2023). For example, some districts set their proficiency threshold at the 25^{th} percentile of their locally chosen assessment, while others set it at the 75^{th} percentile (Personal communication with MDE, 2023). Districts were not required to submit proficiency thresholds or student performance data on locally chosen reading assessments to the state. However, the proficiency threshold was generally lower on locally chosen reading assessments than on statewide standardized ELA assessments (Personal communication with MDE, 2023). During the manual reclassification process, school district administrators identified and reclassified EL students who met state reclassification criteria using a multi-step process depicted in Figure 1.

[Figre 1 about here]

Automatic Reclassification, 2019-20 through 2021-22

Acknowledging disparities between the number of students meeting reclassification criteria and those actually reclassifying, Michigan shifted from a manual to an automatic reclassification policy in the fall of 2019. Changes to the reclassification protocol in the fall of 2019 applied to the 2019-20 school year[2] and beyond.

In shifting to an automatic process, MDE assumes responsibility for reclassifying students through state administrative data systems when the state receives annual WIDA scores from the WIDA consortium. While districts have the opportunity to review students identified for reclassification and can override reclassification decisions if they feel a student is unprepared to reclassify despite meeting criteria, the procedure automatically reclassifies eligible students. Importantly, this process removes all responsibility from districts to reclassify students in district and state data systems.

Beyond the shift to automatic reclassification, MDE made two significant changes to the reclassification criteria. First, MDE simplified the criteria to require only one assessment score as evidence of English proficiency. The revised reclassification criteria eliminated the WIDA reading, WIDA writing, and local ELA assessment proficiency thresholds. Second, the state raised the overall WIDA performance level required to qualify for reclassification from 4.5 to 4.8 out of 6.0. Table 1 outlines the differences in each period's reclassification criteria and procedures.

[Table 1 about here]

Data

The data for our analyses come from the Michigan Education Data Center and include observations for all 3rd through 8th grade EL students with valid WIDA scores between academic years 2016-2017 through 2021-2022 (N = 300,180). We restrict our sample to include only EL students who met at least a 4.0 reading and 4.0 writing performance level on the WIDA assessment and have a valid state standardized ELA test score (N = 43,543). We make this restriction to facilitate comparisons across policy periods, as these students would have been relatively close to the reclassification cutoff in both the manual and automatic reclassification periods. In addition, this restriction accounts for students needing to meet a 4.0 reading and 4.0 writing WIDA performance level and score proficient on a reading assessment to qualify for reclassification in the manual reclassification period.

Table 2 presents summary statistics for the primary analytic sample by policy period (manual reclassification = pre-period; automatic reclassification = post-period). About 34 and 25 percent of students reported speaking Spanish or Arabic as their home language, respectively. Approximately 4 percent of students were also classified as students with disabilities (SWDs), and roughly 71 percent were identified as low-income. About 76 percent of students in the sample were enrolled in elementary grades (third through fifth), and about 24 percent were in middle grades (sixth through eighth). Students' overall, reading, and writing performance level scores were similar across policy periods. This provides some evidence that students in our sample were similar academically before and after the policy change. WIDA overall and subdomain performance levels are similar across policy periods.

To compare scale scores across grades in our analysis, we recenter the scale scores around 0, and 0 represents the minimum scale score required to qualify for reclassification in a given grade. As an example, a third-grade student who attained an overall scale score of 357 in the manual reclassification period would have a value of 1 (the reclassification threshold is 356 for third graders), and a fourth-grade student who attained an overall scale score of 367 in the manual reclassification period would also have a value of 1 (the reclassification threshold is 366 for fourth graders). Recentering the standardized scale scores allows us to estimate the effects of changing policy procedures for the total sample of students.

Recentering within grades also facilitates comparisons between the manual and automatic reclassification policy periods. MDE increased the overall WIDA score needed to qualify for reclassification between periods. To facilitate comparisons across policy periods, we recenter students' scores around the post-period reclassification threshold. On average, ELs in our sample scored 2.4 points above the post-period reclassification threshold. To address changes made to the overall scale score reclassification threshold, we simulate raising the reclassification threshold in the pre-period to match the post-period threshold in our primary analysis. We discuss methods used to simulate raising the threshold in detail in the section titled "Endogeneity Issues."

Of note, the state did not collect data on locally chosen reading assessments, which comprised one additional component of reclassification criteria during the manual reclassification period. To account for this, we use standardized, statewide ELA assessments (M-STEP ELA or PSAT Reading) as a proxy for students' ELA proficiency on locally chosen assessments. Statewide ELA assessments were likely to be as difficult or more difficult than locally chosen reading assessments, so they should serve as a strong indicator for students' performance on local reading assessments (Personal Communication with MDE, 2023). Due to the COVID-19 Pandemic, statewide standardized ELA assessments were not administered during the 2019-20 school year. As a result, we do not include observations from 2019-20 in our main sample[3].

We consider students "qualified for reclassification" if they meet the WIDA overall performance thresholds. The sample in Table 2 reflects the set of students who met WIDA reading and writing performance thresholds of 4.0 (criteria for reclassification eligibility in the manual reclassification period). The manual reclassification period shows large gaps between the number of students eligible for reclassification and those who were reclassified (80 percent versus 51 percent). In contrast, we see similar percentages of students qualifying and reclassifying in the automatic reclassification period (54 percent versus 53 percent). The overall percent of students qualifying for reclassification likely shrunk between policy periods due to raising the overall score needed to qualify for reclassification. Our primary outcome of interest is whether an EL student reclassifies upon meeting reclassification criteria. Table 2 displays the percentage of students in our main sample who met each reclassification threshold under manual and automatic reclassification policies.

[Table 2 about here]

Research Methods

We use two approaches to estimate the effect of qualifying for reclassification on reclassifying during the manual versus automatic reclassification periods. First, we use a sharp RD analysis to estimate the intent-to-treat (ITT) effect of qualifying for reclassification on reclassifying during each of the two policy periods. The ITT effect estimates the impact of meeting or exceeding the reclassification threshold (recentered WIDA scale score) on the outcome (reclassification). Then, we use a DiRD approach to compare the two ITT effect estimates (e.g., Robinson-Cimpian & Thompson, 2016). The difference obtained from the DiRD framework provides a plausibly causal estimate of the effect of shifting from a manual to an automatic reclassification process.

Substantial research indicates that meeting reclassification criteria differentially impacts students' likelihood of reclassifying based on their grade level (e.g., Robinson, 2011; Umansky & Reardon, 2014). To address this, we estimate all RD models separately for grade-level subsets. We present results for grade-level subsamples of students as well as a weighted average effect of thte policy for all students in the sample.

Sharp RD Estimates

We first estimate the ITT effect of meeting the overall WIDA reclassification threshold in the "pre" period (manual reclassification) for student *i* in grade *g*:

$$Y_{ig}^{pre} = \beta_0^{pre} + \beta_1^{pre} C_{ig} + f(M_{ig})^{pre}\ [+X_{ig}\beta^{pre}] + v_{ig}^{pre} \tag{1a}$$

This RD model predicts student *i*'s likelihood of reclassifying *Y* as a function f[4] of their recentered and standardized overall WIDA scale score *M*, an indicator for whether or not that score is above the recentered reclassification threshold *C*, and in some specifications, a vector *X* of additional covariates (recent immigrant status, special education status, low-income status, gender, home language, prior year overall WIDA score).

We restrict the bandwidth of WIDA scale scores used in the analysis to limit the influence of outlier students with very high or very low WIDA scores using the *rdrobust* command, which implements a data-driven process to determine an optimal bandwidth (Calonico et al., 2014). We report results from the optimal bandwidths chosen by the *rdrobust* command for each grade but also report results for ½ and twice the size of the optimal bandwidths as robustness checks. Our preferred model uses a triangular kernel, but we also report results using a uniform kernel as a robustness check. For all RD models, we cluster standard errors at the school district level because districts were responsible for manually reclassifying students in the pre-period and overriding automatic reclassification in the post-period. The *rdrobust* command also adjusts for mass points in determining the bandwidth, meaning that it accounts for the running variable being less than fully continuous, as is the case with most education studies.

In equation (1a), β_1^{pre} represents the ITT effect of just barely qualifying for reclassification on reclassifying in the pre-period (manual reclassification). Equation (1b) is a corresponding analysis for the "post," automatic reclassification period:

$$Y_{ig}^{post} = \beta_0^{post} + \beta_1^{post} C_{ig} + f(M_{ig})^{post}\ [+X_i g \beta^{post}] + v_{ig}^{post} \tag{1b}$$

The ITT estimates for each policy period and grade-level subsample give the impact of qualifying for reclassification on reclassifying in each policy period.

Difference-in-Regression Discontinuities

Next, we use a DiRD approach in an attempt to estimate the impact of shifting from manual to automatic reclassification on eligible students' likelihood of reclassifying. The DiRD approach will estimate the difference in ITT effects from equations (1a) and (1b). This estimate can inform policy by indicating whether the change altered eligible students' likelihoods of reclassifying. Equation (1c) estimates the DiRD separately for each grade level using the sample of students included in the optimal bandwidths[5]. Estimates can be interpreted as the causal effect of shifting from a manual to an automatic reclassification process if there are no other confounding factors and are obtained by subtracting the post- and pre-period ITT effects:

$$DiRD = \beta_1^{post} - \beta_1^{pre} \tag{1c}$$

Where $DiRD$ is the coefficient estimate of the difference between ITT estimates ($\beta_1^{post}, \beta_1^{pre}$) in the post- and pre-periods.

Although we estimate $DiRD$ separately for each grade level, we also report estimated effects for the full sample of students. These estimates are precision-weighted by grade level estimate. To do this, we calculate a weighted average estimate inversely proportional to the standard error of each grade-level estimate. Equation (2) calculates the weighted average estimate of the effect of qualifying for reclassification on reclassifying[6]:

$$Weighted\ Average\ DiRD = \frac{\Sigma_i(\frac{1}{var(DiRD)_g} \times Grade{-}Level\ Estimate_i)}{\Sigma_i \frac{1}{var(DiRD)_g}} \tag{2}$$

The weighted average DiRD is the combined estimate for all grades with weights inversely proportional to the variance; $var(DiRD)_i$ represents the variance of the estimate for grade level g; $Grade - Level\ Estimate_i$ is the estimate for grade level g. The weighted average effect provides an overall estimate of the impact of the policy change on all 3rd to 8th-grade students (e.g., a weighted average effect of 0.26 corresponds to a 26 percentage-point increase in an eligible student's likelihood of reclassifying after the policy change).

Subgroup Analyses

We next use a difference in DiRD (DiDiRD) framework to evaluate differential changes in reclassification rates related to the policy change for subgroups of ELs reporting different home languages. The estimate obtained from the subgroup analyses provides preliminary evidence of the ability of the policy change to ameliorate differential reclassification outcomes unrelated to English proficiency level. Equation (3) estimates the DiDiRD separately for each grade level subgroup of students[7]. Estimates can be interpreted as the effect of shifting from a manual to an automatic reclassification process for a given subgroup of students and are obtained by subtracting the DiRD estimates for subgroups of students:

$$DiDiRD = DiRD_{subgroup\ 1,g} - DiRD_{subgroup\ 2,g} \tag{3}$$

Where $DiDiRD$ is the estimate of the difference in DiRD estimates for a subgroup of students in grade g. For example, if testing for a differential effect of the policy across home language, we would subtract the DiRD estimate for Spanish speakers from that of Arabic speakers in grade g.

Endogeneity Issues

Our estimation strategy faces two primary endogeneity threats. First, when MDE shifted from manual to automatic reclassification, they also eliminated three components of the reclassification criteria (see Table 1). Second, MDE increased the overall WIDA score needed to qualify for reclassification upon shifting to an automatic reclassification process. This section discusses how our analysis accounts for these endogeneity threats.

Accounting for Eliminating Reclassification Criteria

To address the removal of WIDA reading, writing, and local reading assessment scores from state reclassification criteria, we apply a "frontier RD" approach (Reardon & Robinson, 2012). Frontier RD models subset the sample of students used in the analysis to those with scores above or below the cutoff score on all dimensions but one, then model the RD along only one cutoff score (Reardon & Robinson, 2012). Here, we subset the sample to students with at least a 4.0 performance level on the WIDA reading and writing subdomains required to qualify for reclassification during the pre-period. We also include a covariate for students' recentered and standardized M-STEP ELA or PSAT reading scores in equations (1a) and (1b) as a proxy for proficiency on local reading assessments. The resulting frontier RD sample produces results that are easily interpretable, reduces a multidimensional problem (shifting multiple reclassification criteria simultaneously) to a single dimension (only estimating the effect of shifting manual to automatic reclassification policies by accounting for other factors through sample selection), and isolates the effect of shifting from manual to automatic reclassification (e.g., Reardon & Robinson, 2012).

Accounting for Raising the Reclassification Eligibility Threshold

The second endogeneity threat concerns the change to the reclassification eligibility threshold for students' overall WIDA scale scores. MDE increased the overall scale score required to qualify for reclassification across policy periods. Students were eligible for reclassification if they achieved a WIDA overall performance level of 4.5 during the pre-period and 4.8 during the post-period. To ensure comparisons between similar students who would have met reclassification criteria in either period, we simulate raising the reclassification threshold in the pre-period to match the post-period threshold. To create a sharp RD cut point at the higher reclassification threshold in the pre-period models, we assume that all students who did not meet the post-period reclassification threshold were not reclassified. In reality, some of the students below the simulated threshold were reclassified (as demonstrated in Figure 2, Upper Bound ITT Estimate). As a result, our simulated sample (Figure 2, Lower Bound ITT Estimate) will provide a *lower-bound estimate of the effect of shifting reclassification procedures* because the simulated

ITT effect in the pre-period is larger than in reality. Formally, this implies that we inflate the pre-period ITT effect estimate. This inflated estimate is subtracted from the post-period estimate such that $DiRD = \beta_1^{post} - \beta_1^{pre}$. We report results for estimates using the simulated 4.8 reclassification threshold for comparison across policy periods.

[Figure 2 about here]

Internal Validity of Estimates

Our DiRD design relies on several assumptions to produce a causal estimate of the effect of shifting from manual to automatic reclassification. First, we assume that the running variable is not manipulated at the cutoff. Because educators and EL students know the cutoff score required to qualify for reclassification, they may act to manipulate student scores to either retain or reclassify students from EL status, potentially threatening the validity of our estimates. We test the assumption that the running variable is not manipulated at the cutoff using a McCrary test. McCrary (2008) suggests there should be no discontinuity in observations at the cutoff for this assumption to hold. A spike in observations on either side of the cutoff may indicate score manipulation. We report results from McCrary tests in Online Appendix Table A1 to demonstrate no discontinuities in recentered scale scores at the cutoff using triangular and uniform kernels, confirmed using the *rddensity* command in Stata.

Next, the RD assumes that only treatment and outcomes change discontinuously at the cutoff. In other words, although treatment status should change at the cutoff, students must be otherwise similar on either side of the cutoff. If this assumption holds, the RD design produces causal estimates of the effect of the policy change as the groups of students on either side can be used as counterfactuals for one another. Although we cannot test this assumption for unobservable student characteristics, we conduct tests for observable factors (such as gender, special education status, low-income status, and home language). We tested these factors by running separate RDs by grade level and policy period for the analytic sample, each time substituting a different variable as the outcome of interest. Results indicate that no observable student characteristics vary discontinuously at the cutoff other than reclassification likelihood. Online Appendix Table A2 displays the results of these tests.

Finally, for the DiRD to be interpreted as a causal effect, there must be no other cooccurring change—other than those already addressed above, such as the shift in the threshold—that changed and could account for the manual-to-automatic reclassification effect estimate. Ideally, we could address this assumption through the use of an unaffected comparison group via a DiDiRD approach, but there is no unaffected group (e.g., never-ELs) who take the WIDA assessment and are not affected by either one of these policies. As such, there is no comparison that can be used to remove any secular trend from the reclassification policy change. While noting this caveat, it is also worth noting that reclassification rates are fairly stable from year to year in Michigan, as we demonstrate in Table 3. Additionally, there are no other policy changes that we are aware of that could produce a discontinuous change in reclassification rates at the threshold in either the pre- or post-period. Thus, a sizable change from the pre- to post-period in the DIRD is plausibly attributable to the change from manual to automatic reclassification.

[Table 3 about here]

Results

Effects of Automatic Policy on the Likelihood of Reclassification

In this section, we will focus on the results of our lower-bound estimates, which provide the strongest test of the policy change and most directly address credible threats to internal validity of the design. We note that the patterns of significant effect estimates reported here for the lower-bound estimates also—and unsurprisingly—hold for the upper-bound estimates of the effect. See Online Appendix Table A3 for the upper-bound estimates.

We now focus on our preferred model, which we argued yields a lower-bound estimate of the effect. Table 4 presents the results from this preferred model for reclassification likelihood by grade and policy period for the main model specification and an overall effect of the policy change by grade. Online Appendix Table A4 displays results by alternative bandwidth and kernel specifications. Overall, we estimate a significant discontinuity ($p < 0.001$) in students' likelihood of reclassifying upon meeting the reclassification threshold in both the pre- and post-period. This finding holds for all grade levels and weighted average effects in each policy period. Of note, the magnitude of the jump in eligible students' likelihood of reclassifying varies by grade level. For example, during manual reclassification, eligible third graders were the most likely to reclassify (β_1^{pre}= 0.805, $p < 0.001$). In other words, a third grader just above the reclassification threshold experienced an 80.5 percentage-point increase in their likelihood of reclassifying compared to a peer just below the threshold. In contrast, eligible fifth graders were least likely to reclassify (β_1^{pre}= 0.393 [39.3 percentage points], $p < 0.001$). Although a fifth grader just above the reclassification threshold was more likely to reclassify than a peer just below the threshold, the effect of meeting the reclassification threshold was substantially smaller than in other grades. During manual reclassification, we estimate a weighted average effect of 0.630 ($p < 0.001$), meaning, on average, meeting the reclassification threshold increased a student's likelihood of reclassifying by 63.0 percentage points.

During the automatic reclassification period, students just above the reclassification threshold experienced much greater reclassification likelihood than those just below. On average, students just above the cut point were 97.5 percentage points more likely to reclassify than those just below. Further, there is less grade-level variation in eligible students' likelihood of reclassifying during the automatic period. For example, eligible fourth graders are the most likely to reclassify of all grade levels (β_1^{post}= .979 [97.9 percentage points], p < 0.001), and eligible fifth graders are the least likely to reclassify (β_1^{post}= .963 [96.3 percentage points], p < 0 .001).

Estimated ITT effects across policy periods imply that manual and automatic reclassification features differentially impacted eligible students' likelihood of reclassifying. Across all grade levels, we find a statistically significant DiRD estimate. The DiRD estimate is largest in fifth grade ($DiRD$ = .570, p < 0.001), where shifting from manual to automatic reclassification increased eligible students' likelihood of reclassifying by 57 percentage points. This finding suggests that of all grade levels, automatic reclassification "leveled the playing field" the most for fifth graders, who were the least likely of all grade levels to reclassify in the pre-period. In contrast, the DiRD effect is smallest in the third grade ($DiRD$ = .162, p < 0.001), meaning shifting from manual to automatic reclassification increased eligible third graders' likelihood of reclassifying by roughly 16.2 percentage points. Although the DiRD estimate is the smallest in third grade, the policy change still resulted in a statistically significant increase in eligible students' likelihood of reclassifying. The weighted average effect of the policy change is .353 for all students. On average, the policy change across the sample resulted in a 35.3 percentage point increase in any third through eighth-grade student's likelihood of reclassifying upon meeting state reclassification criteria.

Differences in Policy Change Effects Across Language Subgroups of ELs

Prior research has found that eligible students' likelihood of reclassifying varies based on students' racial and linguistic backgrounds (e.g., Mavrogordato & White, 2017; Umansky et al., 2020). In other areas of education policy, automatic procedures are being implemented to ensure greater racial equity in service provision (e.g., automatic advanced course enrollment for students who score at the top of a standardized test distribution, Berg & Plucker, 2023). Where sample sizes allow, we evaluate the effect of automatic reclassification policy across subgroups of ELs to test its ability to increase equity in reclassification outcomes using a DiDiRD framework. We find that the shift to automatic policy had a larger effect on students reporting Spanish as a primary language compared to students reporting other primary languages. This implies that under manual reclassification, eligible Spanish speakers were less likely to reclassify than students speaking other primary languages, and automatic reclassification ameliorated some of the difference in reclassification rates for students reporting different home languages.

Among students near the reclassification threshold, we estimate that shifting from a manual to an automatic reclassification policy affected Spanish speakers more than students reporting another home language. Table 5 presents the DiRD point estimates of shifting from manual to automatic reclassification for students reporting Spanish and Arabic as their home languages. We compare the effects of the policy change for Spanish and Arabic speakers because these are the two most commonly spoken languages among Michigan ELs. We report comparisons to students reporting another primary language in Online Appendix Table A4. Results are consistent with those reported here.

[Table 5 about here]

Results from the weighted average effects indicate that during manual reclassification, Spanish speakers just above the reclassification threshold were roughly 18.7 percentage points less likely to reclassify than Arabic speakers (.496 vs. .683). In the automatic reclassification period, the weighted average effects of qualifying for reclassification on reclassifying are more similar across subgroups, with (.977 for Spanish speakers vs. 1.000 for Arabic speakers). Overall, leveling out across subgroups' likelihood of reclassifying under automatic reclassification implies that the policy had a greater impact on Spanish speakers (.426) than Arabic speakers (.313). Table 6 presents DiDiRD estimates of the effect of the policy change for Spanish and Arabic speakers. The DiDiRD estimates of the policy change confirms this finding, with the shift to automatic policy having an 11.3 percentage-point greater impact on Spanish speakers than on Arabic speakers.

[Table 6 about here]

Notably, the difference in DiRD estimate appears to be primarily driven by differences in third-grade Spanish reclassification rates compared to Arabic. During manual reclassification, third-grade Spanish speakers just above the reclassification threshold experienced a roughly 63.5 percentage-point increase in their likelihood of reclassifying. In contrast, similar Arabic speakers experienced a roughly 84 percentage-point increase in reclassification likelihood. Other grade-level ITT effect estimates are similar across these linguistic subgroups during manual reclassification. For example, eligible fifth graders reporting Spanish as a home language experience a nearly 34.6 percentage point increase in likelihood of reclassification, compared to an increase of nearly 38.7 percentage points at the threshold for students reporting Arabic as a home language.

During automatic reclassification, third-grade Spanish speakers just above the threshold continue to experience a lower likelihood of reclassification (.955) compared to Arabic speakers (1.000), but ITT effects are more similar. Weighted average effects during the automatic reclassification period suggest that eligible Spanish speakers (.977) continue to experience a lower likelihood of reclassification than Arabic speakers (1.000). However, the difference in weighted average effects across groups is much smaller than under manual reclassification.

Robustness Checks

We conduct several robustness checks to estimate the ITT effect of qualifying for reclassification on reclassifying across grade levels and policy periods. We test the sensitivity of our main analytic models to different models, including alternative kernels, bandwidths, and clustered standard errors. We also estimate each model with and without the inclusion of covariates. Finally, our main models include a control for standardized ELA scores as a proxy for achievement on a local ELA assessment (one component of reclassification criteria during manual reclassification). This excludes data from 2019-20, as standardized tests were not administered during the COVID-19 Pandemic. In some alternative models, we exclude the standardized ELA score control and include reclassification data from 2019-20. Results from the robustness checks are comparable to results presented in the main findings and indicate that shifting to automatic reclassification had a large, positive effect on reclassification rates among eligible students.

Conclusion and Policy Implications

Using administrative data from Michigan, this study finds that automatic or default procedures can (1) substantially increase adherence to statewide standardized EL reclassification policy and (2) reduce linguistic or other disparities in access to reclassification compared to manual procedures. We find statistically significant, substantial effects of shifting from a manual to an automatic reclassification policy on reclassification rates of eligible EL students. The effects of shifting to an automatic reclassification policy are larger for specific subgroups of ELs, namely Spanish speakers. These findings have implications for EL policy in light of ESSA's (2015) mandate that states establish standardized EL reclassification protocols. These findings also have implications for education policy more broadly as states and school districts look to increase equity in students' access to specialized programs, such as advanced coursework.

Implications for EL Policy

From an EL policy perspective, these findings corroborate earlier pre-ESSA research that finds substantial discrepancies between the population of students qualifying for reclassification and those reclassifying (Cimpian et al., 2017; Estrada & Wang, 2018). We extend this research base by highlighting automatic reclassification procedures as a mechanism to reduce these discrepancies. Recognizing that many eligible ELs were not reclassifying on time under manual procedures, Michigan implemented an automatic reclassification policy. Under manual reclassification, we confirm significant disparities in reclassification rates of eligible students. This suggests that reclassification decisions may have been based on factors other than reclassification criteria. Prior literature highlights several features of manual reclassification procedures that may contribute to disparities in reclassification rates, including excessive administrative burden on school districts, educators' beliefs about the merits of reclassification, and EL students themselves, differences in state reclassification policy interpretation (Estrada & Wang, 2018; Mavrogordato & White, 2017), and variation in policy implementation across districts (Cimpian et al., 2017). This study provides causal evidence that shifting to an automatic procedure can create much greater parity in reclassification rates of eligible students. Of note, some eligible students still do not reclassify under automatic procedures. This is likely because districts can override or opt out of automatic reclassification, a key feature of default policies.

Further qualitative research is needed to understand the circumstances in which districts override reclassification for eligible students.

As the EL population continues to grow and diversify rapidly, it is vital to consider how reclassification policies impact students within the EL subgroup differently. We provide the first causal evidence of the effect of a state's choice of reclassification procedures on reclassification outcomes for ELs as a whole and among subgroups. Overall, we find that manual procedures impact subgroups of ELs differently. First, we find that eligible ELs are more or less likely to reclassify based on their grade. For example, under manual reclassification, roughly 20 percent of third-grade students who were eligible for reclassification did not reclassify. This research parallels Umansky and Reardon's (2014) conclusion that in early grades, more students meet reclassification criteria than reclassify. However, Umansky and Reardon (2014) find that this trend reverses in middle school, with more students reclassifying than meeting eligibility criteria. In contrast, we find that under manual procedures, a significant proportion of 6th through 8th grade students who met reclassification criteria were not reclassified. Under automatic reclassification, these gaps close, and nearly all eligible 3rd through 8th grade students reclassify. Taken together, these findings suggest that between-grade variation in eligible students' reclassification likelihood exists under manual procedures. Automatic procedures may be more effective at standardizing reclassification outcomes for students, a key goal of ESSA (2015).

In addition, research has identified subgroups of ELs that are less likely to reclassify upon meeting reclassification criteria, particularly ELs reporting Spanish as a home language (Mavrogordato & White, 2017; Umansky et al., 2020). Our estimates align with this research. Under manual reclassification procedures in which districts are responsible for reclassifying ELs, we find that eligible ELs who report Spanish as a home language are substantially less likely to reclassify than ELs who report other home languages. However, this discrepancy largely dissipates upon shifting to automatic reclassification, in which state data systems reclassify eligible ELs. This finding suggests potential bias against Spanish speakers under manual reclassification procedures.

Implications for Education Policy

Nationwide, state education agencies are grappling with the most effective ways to increase representation and enrollment in specialized educational services such as advanced coursework and gifted education (Blad, 2020). Many state education agencies are moving towards automatic enrollment to ensure students are served in a developmentally appropriate environment (Plucker, 2021). In light of this movement, rigorous causal evidence is needed to evaluate the ability of automatic policy to increase equity and representation in educational service enrollment. The present study offers the first evaluation of the effects of automatic policy in K-12 educational settings, finding it can increase students' likelihood of being served in a developmentally appropriate environment (e.g., by reclassifying upon demonstrating English proficiency).

This study faces several notable limitations. First, reclassification often entails a significant change in students' instructional environment and has important implications for their short- and long-term outcomes. The present study focuses on evaluating the efficacy of automatic procedures in increasing adherence to state policy and ESSA guidance rather than assessing the effects of reclassification on students' outcomes. Future research may explore outcomes for "compliers," or eligible students who would reclassify under automatic procedures and not under manual procedures, to determine whether shifting the policy had positive or negative effects on student outcomes. Moreover, this study does not identify the mechanisms that caused lower reclassification of eligible students under manual reclassification. Future qualitative research may explore why manual reclassification procedures resulted in a substantially lower likelihood of reclassification among eligible students than automatic procedures. In addition, there are several limitations of our DiRD design worth noting. The generalizability of our estimates is restricted to students just above or below the reclassification threshold, and this limits the applicability of our conclusions. As the threshold determines reclassification decisions, this study provides policy-relevant findings but cannot speak to students whose reclassification decisions are overridden (e.g., students who do not "comply"). Finally, these findings will not be generalizable to all states. Many states include subjective measures in their reclassification criteria (e.g., teacher recommendation, student grades), and subjective criteria are not collected by state data systems. As such, results and implications should be considered in a state with reclassification criteria captured by state administrative data systems.

Rigorous research is needed to examine the ways policy can expand or constrain educational opportunities for the growing and diversifying EL population in US schools. Reclassification is one mechanism through which ELs gain access to the full range of academic coursework, and thus, policymakers should prioritize reclassifying students who demonstrate eligibility by meeting state reclassification criteria. This study offers one potential mechanism, automatic policy, that policymakers may consider to ensure greater equity in reclassification decisions among eligible students.

Notes

1. Although ESSA (2015) requires states to establish standardized within-state EL exit criteria, states have discretion to determine their exit criteria. Reclassification criteria

vary across states and may include student grades, teacher recommendations, standardized test scores, and other factors, but all states must include an assessment-based measure of English proficiency set by policymakers within that state (Linquanti & Cook, 2015).

2. Schools first closed for the COVID-19 Pandemic in spring 2020. The window of time in which schools administered WIDA assessments occurred largely before school closures, unlike many standardized assessments. Therefore, we consider the first year of automatic policy implementation to be unaffected by the COVID-19 Pandemic. Further, although fewer students participated in WIDA assessments during virtual learning (2020-21), we find no substantive differences in WIDA performance between the pre- and post-periods, suggesting that schools did not systematically test higher performing students during virtual learning.
3. Further, fewer assessments than usual were administered during the 2020-21 school year because many students attended virtual school due to the COVID-19 Pandemic and thus did not participate in standardized testing. We pool observations across policy periods for our main analyses.
4. Our preferred model uses a first-order polynomial because we anticipate that a student's WIDA score has a linear relationship to their likelihood of reclassifying (students with a perfect score being most likely to reclassify under manual reclassification procedures). Our results are robust to higher-order polynomials.
5. After obtaining a DiRD estimate for each grade level, we compute standard errors for each grade-level DiRD estimate:

$$SE(DiRD) = \sqrt{var(\beta_1^{pre}) + var(\beta_1^{post})}$$

Where $SE(DiRD)$ represents the standard error of the DiRD estimates. $var(\beta_1^{pre})$ and $var(\beta_1^{post})$ represent the variance of the estimated treatment effects obtained separately for the manual and automatic reclassification periods. This formula accounts for uncertainty in both pre- and post-period effect estimates when calculating the standard error of $DiRD$. We assume covariance between β_1^{pre} and β_1^{post} is zero or positive, thus if covariance is included in this equation, $SE(DiRD)$ will be smaller than is presented in the results.

6. The variance estimate of the precision-weighted average effect is computed using the following formula:

$$Var(Weighted\ Average\ DiRD) = \frac{1}{\Sigma \frac{1}{var(DiRD)_i}}$$

The corresponding standard error is calculated using the following equation:

$$SE_{Weighted\ Average\ DiRD} = \sqrt{var(Weighted\ Average\ DiRD)}$$

7. After obtaining a difference in DiRD estimate for each grade level, we compute standard errors for each grade-level DiRD estimate:

$$SE(DiDiRD) = \sqrt{var(DiRD_{subgroup\ 1,g}) + var(DiRD_{subgroup\ 2,g})}$$

Where $SE(DiDiRD)$ represents the standard error of the difference in DiRD estimates.

References

Behlen, L., Himmler, O., & Jäckle, R. (2023). Defaults and effortful tasks. *Experimental Economics*, *26*(5), 1022-1059. https://doi.org/10.1007/s10683-023-09808-8

Berg, B., & Plucker, J. (2023, November 6). Why your state should automatically enroll top math students in advanced classes. The 74 Million. https://www.the74million.org/article/why-your-state-should-automatically-enroll-top-math-students-in-advanced-classes/

Blad, E. (2020). *The simple policy change that's getting more students of color in advanced courses.* Education Week. https://www.edweek.org/leadership/the-simple-policy-change-thats-getting-more-students-of-color-in-advanced-courses/2020/03

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, *82*(6), 2295-2326. https://doi.org/10.3982/ECTA11757

Carlson, D., & Knowles, J. (2016). The effect of English language learner reclassification on student ACT scores, high school graduation, and postsecondary enrollment: Regression discontinuity evidence from Wisconsin. *Journal of Policy Analysis and Management, 35,* 559-586. https://doi.org/10.1002/pam.21908

Cimpian, J. R., Thompson, K. D., & Makowski, M. B. (2017). Evaluating English learner reclassification policy effects across districts. *American Educational Research Journal*, *54*(1), 255S-278S. https://doi.org/10.3102/0002831216635796

Cox, P., Fox, J., & Tutt, S. (2020). Forgotten borrowers: protecting private student loan borrowers through state law. *UC Irvine L. Rev.*, *11*, 43.

Estrada, P., & Wang, H. (2018). Making English learner reclassification to fluent English proficient attainable or elusive: When meeting criteria is and is not enough.

American Educational Research Journal, *55*(2), 207-242. https://doi.org/10.3102/0002831217733543

Every Student Succeeds Act. (2015). Pub. L. No. 114-95, S. 1177—114th Congress Stat.

Garnett, H. A. (2022). Registration innovation: The impact of online registration and automatic voter registration in the United States. *Election Law Journal: Rules, Politics, and Policy*, *21*(1), 34-45. https://doi.org/10.1089/elj.2020.0634

Herd, P., DeLeire, T., Harvey, H., & Moynihan, D. P. (2013). Shifting administrative burden to the state: The case of medicaid take-up. *Public Administration Review*, *73*(s1), S69-S81. https://doi.org/10.1111/puar.12114

Jachimowicz, J. M., Duncan, S., Weber, E. U., & Johnson, E. J. (2019). When and why defaults influence decisions: A meta-analysis of default effects. *Behavioural Public Policy*, *3*(2), 159-186. https://doi.org/10.1017/bpp.2018.43

Johnson, E. J., & Goldstein, D. (2003). Do defaults save lives?. *Science*, *302*(5649), 1338-1339. https://doi.org/10.1126/science.1091721

Madrian, B. C., & Shea, D. F. (2001). The power of suggestion: Inertia in 401 (k) participation and savings behavior. *The Quarterly journal of economics*, *116*(4), 1149-1187. https://doi.org/10.1162/003355301753265543

Mavrogordato, M., & White, R. (2017). Reclassification variation: How policy implementation guides the process of exiting students from English learner status. *Educational Evaluation and Policy Analysis*, *39*(2), 281-310. https://doi.org/10.3102/0162373716687075

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, *142*(2), 698-714. https://doi.org/10.1016/j.jeconom.2007.05.005

Morales, C., & Lepper, A. (2024, April 25). The changing landscape of states' English learner reclassification policies. Evanston, IL: Northwestern Center for Education Efficacy, Excellence, and Equity. https://e4.northwestern.edu/2024/04/25/the-changing-landscape-of-states-english-learner-reclassification-policies/

National Center for Education Statistics (NCES). (2022). English language learners in public schools. *Condition of Education*. U.S. Department of Education, Institute of Education Sciences. Retrieved May 13, 2022, from https://nces.ed.gov/programs/coe/indicator/cgf.

Okhremtchouk, I., Levine-Smith, J., & Clark, T. (2018). The web of reclassification for English language learners—a cyclical journey waiting to be interrupted: Discussion of realities, challenges, and opportunities. *Educational Leadership Administration: Teaching and Program Development*, *29*(1), 1-13. https://files.eric.ed.gov/fulltext/EJ1172216.pdf

Plucker, J. A. (2021). Automatic enrollment is a no-brainer. American Consortium for Equity in Education. https://www.ace-ed.org/jonathan-a-plucker-automatic-enrollment-is-a-no-brainer/

Reardon, S. F., & Robinson, J. P. (2012). Regression discontinuity designs with multiple rating-score variables. *Journal of Research on Educational Effectiveness*, *5*(1), 83-104. https://doi.org/10.1080/19345747.2011.609583

Richman, T. (2023, October 3). *How Texas plans to make access to advanced math more equitable.* The Hechinger Report. https://hechingerreport.org/how-texas-plans-to-make-access-to-advanced-math-more-equitable/

Robinson, J. P. (2011). Evaluating criteria for English learner reclassification: A causal-effects approach using a binding-score regression discontinuity design with instrumental variables. *Educational Evaluation and Policy Analysis*, *33*(3), 267-292. https://doi.org/10.3102/0162373711407912

Robinson-Cimpian, J. P., & Thompson, K. D. (2016). The effects of changing test-based policies for reclassifying English learners. *Journal of Policy Analysis and Management*, *35*(2), 279-305. https://doi.org/10.1002/pam.21882

Umansky, I. M. (2016). Leveled and exclusionary tracking: English learners' access to academic content in middle school. *American Educational Research Journal*, *53*(6), 1792-1833. https://doi.org/10.3102/0002831216675404

Umansky, I. (2018). According to plan? Examining the intended and unintended treatment effects of EL classification in early elementary and the transition to middle school. *Journal of Research on Educational Effectiveness*, *11*(4), 588-621. https://doi.org/10.1080/19345747.2018.1490470

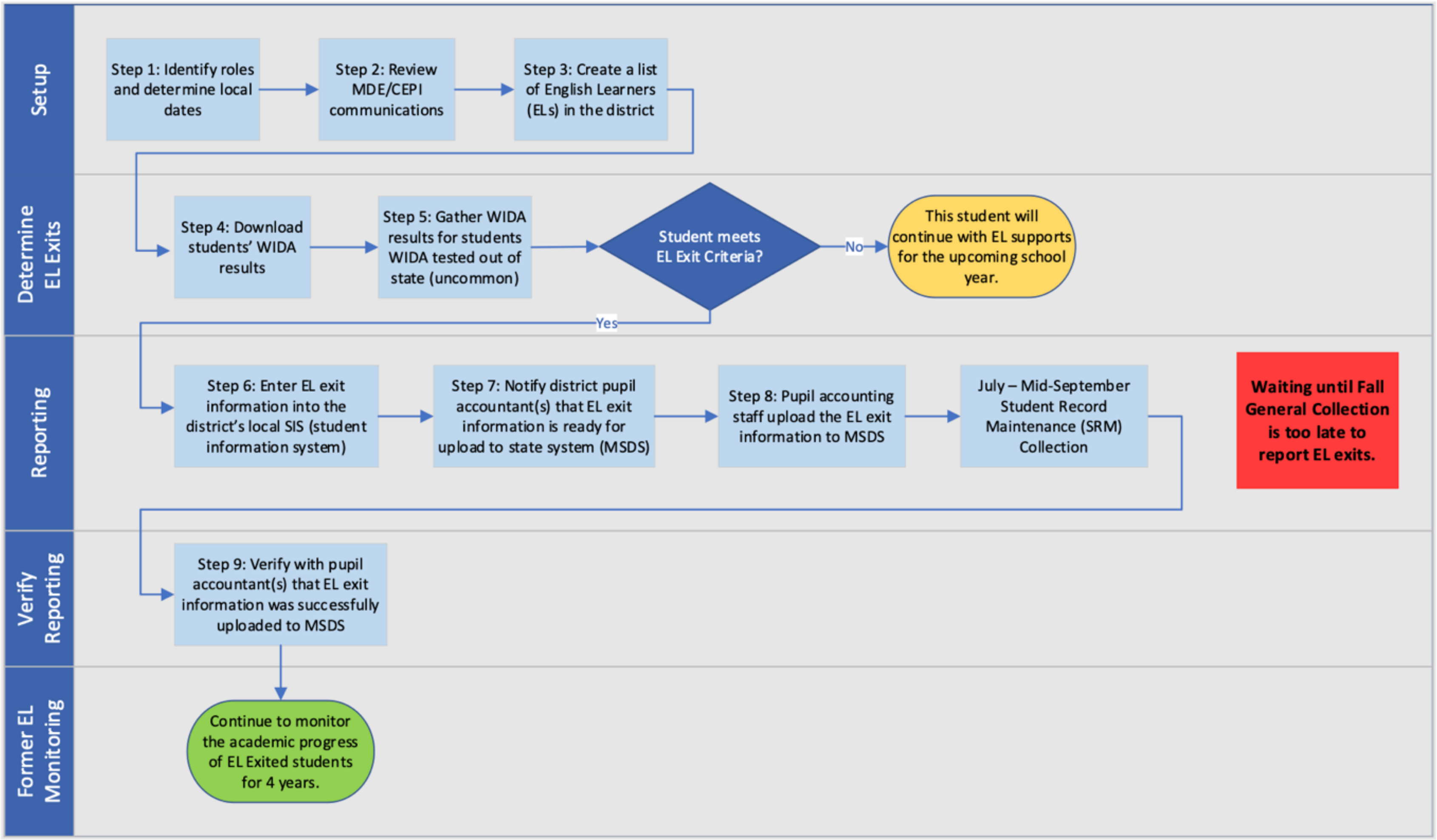
Umansky, I. M., Callahan, R. M., & Lee, J. C. (2020). Making the invisible visible: Identifying and interrogating ethnic differences in English learner reclassification. *American Journal of Education*, *126*(3), 335-388. https://doi.org/10.1086/708250

Umansky, I. M., & Reardon, S. F. (2014). Reclassification patterns among Latino English learner students in bilingual, dual immersion, and English immersion classrooms. *American Educational Research Journal*, *51*(5), 879-912. https://doi.org/10.3102/0002831214545110

WIDA. (2024). *ACCESS for ELLS: Interpretive guide for score reports, grades K-12, spring 2024.* https://wida.wisc.edu/sites/default/files/resource/Interpretive-Guide.pdf

Figure 1

Manual Reclassification Procedure in Michigan, 2016-17 through 2018-19



Source: Personal Communication with MDE, 2023

Table 1

Reclassification Criteria and Procedures by Policy Period

	Manual Reclassification (2016-17 through 2018-19)	Automatic Reclassification (2019-20 through 2021-22)
Criteria	• Overall WIDA performance level of 4.5 or greater • WIDA Reading Performance level of 4.0 or greater • WIDA writing performance level of 4.0 or greater • "Proficient" in a locally chosen reading assessment	• Overall WIDA performance level of 4.8 or greater
Procedure	• District convenes a team and completes a nine-step reclassification process for qualified students	• State data system automatically reclassifies qualified students • Districts have the option to override automatic reclassification

Table 2

Descriptive Statistics by Policy Period

	Reclassification Period		
	Manual Period (2016-17 through 2018-19)	Automatic Period (2019-20 through 2021-22)	Full Sample
Covariate means (% of sample)			
Female	54	55	54
Primary Language: Spanish	36	31	35
Primary Language: Arabic	25	26	26
Primary Language: Other	37	42	38
SWD	4	4	4
Low-income	72	70	72
Grade (% of sample)			
3	19	17	19
4	28	35	30
5	25	33	27
6	9	4	8
7	9	5	8
8	9	6	8
Test Scores			
WIDA Overall Scale Score Recentered Around Post-Period Reclassification Threshold	2.19 (15.49)	2.23 (15.95)	2.19 (15.49)
WIDA Overall Performance Level	4.87 (0.52)	4.87 (0.53)	4.87 (0.52)
WIDA Reading Performance Level	5.58 (0.56)	5.53 (0.58)	5.57 (0.57)
WIDA Writing Performance Level	4.35 (0.33)	4.37 (0.40)	4.35 (0.35)
Reclassification			
Qualified for Reclassification	80%	54%	62%

Reclassified	51%	53%	51%
N	32,126	11,417	43,543

Note: Data for this analysis come from the Michigan Education Data Center. EL, English learner; SWD, student with disabilities. WIDA scale scores are recentered around their respective grade-level reclassification threshold for interpretation across grades (e.g., a recentered scale score of -1 can be interpreted as meaning the student missed the reclassification threshold in their grade and policy period by 1 point). The WIDA assessment reports scale scores between 100-600, and performance level scores are reported on a scale of 1.0-6.0. State ELA assessment = M-STEP ELA or PSAT reading. A student is considered "qualified for reclassification" if they met the grade-level WIDA overall scale score threshold for reclassification in the post-period.

Figure 2

Upper- and Lower-Bound Pre-Period ITT Estimates

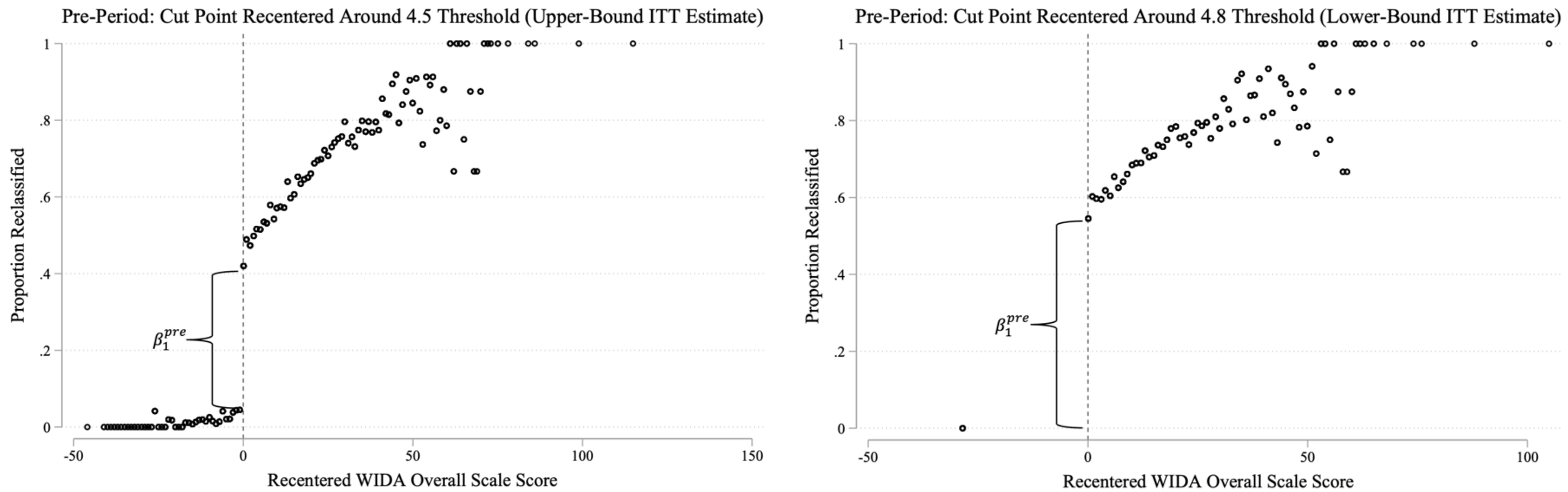


Table 3

Reclassification Rates for Eligible ELs by Year

Year	Students Eligible to Reclassify (N)	Students Reclassified (N)	Percentage of Eligible Students Reclassified
2016-17	7,562	4,239	56.06
2017-18	8,819	5,811	65.89
2018-19	9,164	6,119	66.77
2019-20	2,156	2,127	98.65
2020-21	3,980	3,905	98.12

Note: Data come from the Michigan Education Data Center. Automatic reclassification was implemented in the 2019-20 school year.

Table 4

Lower-Bound Estimated Effect of Qualifying for Reclassification on Reclassifying Across Grades and Policy Periods

	Manual Reclassification 2016-17 through 2018-19	**Automatic Reclassification 2019-20 through 2021-22**	**Policy Change (DiRD)**
3	0.805***	0.968***	0.162***
SE	(0.024)	(0.018)	(0.030)
BW	[-50, 15]	[-11, 8]	
N	5317	950	
4	0.562***	0.979***	0.418***
SE	(0.036)	(0.008)	(0.037)
BW	[-48, 25]	[-18, 21]	
N	7976	3095	
5	0.393***	0.963***	0.570***
SE	(0.035)	(0.021)	(0.041)
BW	[-48, 26]	[-11, 21]	
N	8076	2608	
6-8	0.580***	0.973***	0.393***
SE	(0.026)	(0.019)	(0.032)
BW	[-57, 18]	[-7, 12]	
N	8279	792	
Weighted Average	0.630***	0.975***	0.353***
SE	(0.014)	(0.007)	(0.017)
Standardized ELA Score Control		X	
Local Polynomial		1	
Bandwidth		Optimal	
Kernel		Triangular	

Note: * $p < .05$, ** $p < .01$, *** $p < .001$ Robust standard errors clustered at the school district level appear in parentheses below the point estimates.

Table 5

DiRD Estimates Across Language Subgroups of ELs

	Primary Language: Spanish			**Primary Language: Arabic**		
	Manual Reclassification 2016-17 through 2018-19	**Automatic Reclassification 2019-20 through 2021-22**	**Policy Change (DiRD)**	**Manual Reclassification 2016-17 through 2018-19**	**Automatic Reclassification 2019-20 through 2021-22**	**Policy Change (DiRD)**
3	0.663***	0.953***	0.290***	0.844***	1.000***	0.156***
SE	(0.064)	(0.044)	(0.053)	(0.029)	0.000	(0.029)
Bandwidth	[-50, 8]	[-11, 9]		[-44, 9]	[-41, 2]	
N	1283	208		1274	370	
4	0.527***	0.994***	0.467***	0.543***	0.988***	0.445***
SE	(0.053)	(0.017)	(0.056)	(0.049)	(0.011)	(0.050)
Bandwidth	[-40, 19]	[-42, 8]		[-42, 19]	[-6, 16]	
N	2765	729		1978	568	
5	0.358***	0.936***	0.578***	0.396***	0.978***	0.582***
SE	(0.045)	(0.044)	(0.063)	(0.064)	(0.025)	(0.068)
Bandwidth	[-48, 21]	[-55, 22]		[-44, 23]	[-8, 12]	
N	2886	1242		1864	512	
6-8	0.575***	0.931***	0.356***	0.596***	1.002***	0.407***
SE	(0.040)	(0.042)	(0.058)	(0.040)	(0.003)	(0.040)
Bandwidth	[-52, 13]	[-44, 12]		[-51, 14]	[-5, 4]	
N	3298	551		1933	100	
Weighted Average	0.515***	0.977***	0.411***	0.687***	1.000***	0.306***
SE	(0.024)	(0.014)	(0.029)	(0.020)	0.000	(0.020)
Standardized ELA Score Control		X			X	
Local Polynomial		1			1	

Bandwidth	Optimal	Optimal
Kernel	Triangular	Triangular

Note: * $p < .05$, ** $p < .01$, *** $p < .001$ Robust standard errors clustered at the school district level appear in parentheses below the point estimates. When point estimate and standard error equal 1 and 0, respectively, estimates indicate that every student who qualified for reclassification was reclassified. This could be interpreted as a sharp RD. Because the third-grade post period point estimate for Arabic speakers is interpreted as a sharp RD, the precision weighted average is exactly the value of the sharp RD.

Table 6

DiDiRD Estimates Across Language Subgroups of ELs

	DiRD: Spanish	DiRD: Arabic	DiDiRD
Weighted Average Estimate	0.411***	0.306***	0.105***
SE	(0.029)	(0.020)	(0.035)