The Relationship Between Special Education Placement and High School Outcomes

Abstract

Propensity score matching was used to create intervention and comparison groups to compare academic outcomes of students with disabilities in Indiana placed in more inclusive settings with those placed in less inclusive settings. Student and school demographic and outcome data were analyzed from the eighth grade in 2013 through graduation in 2018. Students with disabilities spending 80% more time in inclusive classrooms did better in reading and math than peers spending more time in special education classrooms. The study shows differences in diploma types of students in more inclusive settings than those in less inclusive settings, indicating that students in inclusive settings engage in more rigorous course of study and are more prepared for successful post-secondary educational and employment opportunities.

Keywords: Inclusion, achievement, propensity score matching, high school

The Relationship Between Special Education Placement and High School Outcomes

Whether inclusion is a better placement decision for the provision of services to meet the needs of children with disabilities is still debated among proponents and advocates. However, this debate often is about whether every child, regardless of their disability is in the appropriate placement. This debate is anchored in two of the basic aspects of the Individuals with Disabilities Education Act (IDEA): 1) the requirement of receiving services in the Least Restrictive Environment (LRE) and 2) being provided Free and Appropriate Services (FAPE). These two aspects are often misconstrued in the reasoning that either side employs. Rather than arguing that all children must be included, a clear violation of FAPE in pursuit of LRE, the better course is a discussion of whether more students can be and should be included, and if so, how to do so. This position allows a pursuit of LRE placements without violating FAPE and maintains the spirit of special education placements for services reflecting the Individual Education Plan (IEP) requirements of an individual student.

This study is the second phase of the author's published study in 2020, which focused on the relationship of placement to the academic outcomes of students in Indiana grades 3-8 for students with primary disabilities of Cognitive Disability, Learning Disability, Autism Spectrum Disorder, Emotional Disability, Other Health Impaired, Blind/Low Vision and Deaf/Hard of Hearing.

Given the findings in favor of high inclusion in the first study, it became important to assess whether the same pattern existed for high school students with disabilities. This study which conducts a statewide comparison of student achievement in Indiana for high school students in low and high inclusive placements was undertaken to help resolve the issue of whether more students should be included to address their academic needs, and thus be

consistent with the requirements IDEA and its foundational concepts of FAPE and LRE. Moreover, because diploma type often reflects the curriculum path of high school students, this study was undertaken to investigate relative differences in preparedness for post-secondary transitions related to placement. Finally, as noted by Francisco (2020), additional research on the academic effectiveness of inclusion is needed.

Literature Review

Special Education as a Legal Mandate

Following several landmark court cases concerning individuals with disabilities, Public Law 94-142 was passed in November of 1975. It was followed in 1990 with the Individuals with Disabilities Education Act (IDEA) and the Americans with Disabilities Act (ADA). In 2004, the IDEA was amended and reauthorized as the Individuals with Disabilities Education Improvement Act and followed previous legislation with provisions that gave students the right to a Free and Appropriate Public Education (FAPE) in the Least Restrictive Environment (LRE) (Francisco et al.,2020)

The Individualized Education Program (IEP) is a key part of the IDEA and is intended to provide the goals, objectives, and necessary services to meet the student's needs. The IEP is one of the most important documents in special education and yet, because it is developed and implemented by various individuals, its quality varies (Francisco et al., 2020).

One of the most debated constructs of IDEA is the "least restrictive environment". It is described in the law as educating students with disabilities in a general education classroom as much as appropriate. LRE is often conceptualized as a continuum of services, with an assumption that the students who need the most intensive services should be placed in a segregated classroom, which focuses on place, rather than service. The definition is broad and

vague and has resulted in different interpretations across states and schools. There is strong evidence that placement decisions may be heavily influenced by the state in which a student lives (Anderson et al., 2020). Often factors such as financial resources, school organization, advocacy, knowledge, values, and local context greatly influence placement decisions for students with disabilities and may deny the student the right to FAPE (Francisco, et al., 2020).

Student Outcomes

The initial legislation outlining policies, procedures, and programming for students with disabilities is now more than 4 decades old and over time, progress has been made in ensuring that students with disabilities have access to the general education classroom. In 2019, Williamson et al., reported a trend over two decades that showed more students with disabilities are being educated in general education settings for most of the day. Yet, according to the most recent Report to Congress, only 64% of students with disabilities are in the general education classroom 80% or more of their day (U.S. Department of Education, 2021).

There has been consistently poor achievement of students with disabilities in reading and math proficiency. In a 2014 memo to states, the U.S. Department of Education noted that from 2009-2013, proficiency levels on the National Assessment of Education Progress (NAEP) decreased for students with disabilities while they increased for their non-disabled peers. Only 10–12% of students with disabilities read at a proficient level, and that number decreases to approximately 6-8% for math proficiency (National Center for Educational Statistics, 1992–2017). In the most recent Report to Congress, it was reported that in 2017-18 school year, the median percentages of students with disabilities who were found to be proficient on state assessments in math ranged from 7-23.5 percent; reading percentages ranged from 10.9-18.9 percent. These unfavorable outcomes are also reflected in studies of the experiences of students

with disabilities in the all-important high school years. Consistently unsuccessful outcomes for students with disabilities, including high drop-out rates, low rates of graduation and low rates of enrollment in post-secondary institutions of education, show disparities between youth with and without disabilities (Mazzotti, et al., 2021). Because of their consistent lack of success on the measures of academic success, the education of students with disabilities continues to be scrutinized.

Impact of Inclusion

According to Francisco et al., (2020), the issues and concerns regarding the implementation of inclusion are due in part to the lack of consistent evidence in empirical studies on the effectiveness of inclusion. This inconclusive evidence about placement for students with disabilities confounds the decision-making process for initial placement of students with disabilities and the intent of FAPE and LRE. Some studies that research the relationship of inclusion with different disabilities and school outcomes have found that inclusion has yielded mixed and even negative results (Cook & Cook, 2020; Afacan & Wilkerson, 2021). A study by McKibbin (2020) found that program placement (regular or specialized program) did not predict academic achievement. Conversely, large-scale academic studies found that inclusive education produces improved social and developmental outcomes, improved quality of life and improved academic outcomes (Barrett et.al., 2020; Wehmeyer et al., 2021). Using data from the National Longitudinal Transition Study-2, Rojewski et.al., (2015) found significant causal effects for inclusive placement in high school and post-secondary outcomes for students with high incidence disabilities. Lombardi et.al., (2013) also used data from the National Longitudinal Transition Study-2 to review the relationship between placement and post -

secondary participation. They found that students with disabilities were 2 times more likely to participate in a two- or four-year college program.

Though there is research indicating that many students with disabilities benefit academically and socially from inclusive educational services (Barrett, et al., 2020) self-contained classrooms for students with disabilities continue to exist in public schools across the k-12 grade level continuum (Causton-Theoharis, et al., 2011). The highly structured, and controlled environment where teachers can provide specialized instruction is used by some to justify placement in special education classrooms without general education peers. From the standpoint of instructional effectiveness, the proponents contend that providing instruction in small, homogeneous instructional groups may enable instruction more tailored to the needs of students with disabilities (Landrum et al., 2003).

However, the potential benefits of self-contained settings may not be realized in schools. Causton-Theoharis et al. (2011) found "a disconnect between the rationale for self-contained special education and reality" (p. 73) with little evidence to suggest that the supports offered in self-contained classrooms could not be integrated into more inclusive settings. Research has suggested the majority of practices and supports observed in self-contained settings could be implemented in more inclusive settings without compromising the effectiveness of instruction (Causton-Theoharis, et.al, 2011; Barrett, et al., 2020).

Results Driven Accountability

In 2014, the U.S. Department of Education launched the Results-Driven Accountability (RDA) framework for monitoring a state's delivery of special education services. This accountability system shifts the accountability from an emphasis on compliance to a framework focused on improved student outcomes based on performance assessments. The shift was made

based on a need for greater emphasis on improving academic results for students with disabilities. In the six years of state determinations based on this framework, states have made little or no progress in improving their ratings or their outcomes for students with disabilities. In 2019, 37 states were in the "needs assistance" category and 3 states were in "need of intervention" (U.S. Department of Education, 2020).

Purpose of Study

This study investigates the relationship between inclusion and academic outcomes. With the evolution of state and federal accountability systems focused on results, it is an opportune time to use research on inclusion and student outcomes to guide policy and practice. The study's design incorporates the most significant school and student level variables found in the research that relate to student, referral, placement, and outcomes. By including these student and school variables the results of the study enable their impact to be considered in policy and practice concerning decision-making for placement to provide services to meet the needs of students identified with disabilities.

The study's statistical methodology also accounts for, in part, the unique and unidentified aspects of the referral, identification, and the placement process generally not included in state reporting requirements, like parental involvement, teacher attitudes, and individualized education program (IEP) instructional strategies. Individually and collectively, the aspects of the referral, identification, and placement process could influence how decisions are made concerning the provision and the quality of services. Addressing them in research investigating the impact of placement upon student outcomes is critically important. Ruling out the potential impact, if any, of these and other variables not incorporated into the analyses in interpreting the study's results allows for meaningful and valid interpretations of the findings.

Method

The data, methods, and analytical procedures for the current study replicate those of the author's prior study that looked at placement and academic outcomes for students in grades 3-8 using statewide data from Indiana. That study used a propensity matching process to create matched samples of students in high inclusion and low inclusion placements to affect a randomized design for the intervention and comparison groups, and to account for relationships between an array of student, and school variables including primary disability category related to special education placement and academic outcomes. In that study, students experiencing more inclusion demonstrated significantly higher achievement on the state assessment than students experiencing less inclusion regardless of the disability category or any of the unique school characteristics incorporated into the study's design.

Use of propensity score matching for the current study that extends the research of the impact of placement to high school students is presented in greater detail in the following sections. An important aspect and value of using propensity matching as a method to create treatment and control groups to approximate the random control and experimental groups of an experimental design is conducting sensitivity analyses to assess the impact of matching to determine that the groups are statistically equivalent on the variables used in the matching process. Additionally, sensitivity tests can also be used to determine the impact, if any, of unobserved variables that might be related to the matching process upon the outcomes. For the current study these sensitivity tests enable the measurement of any systematic bias from variables not in the data base and unable to be included in the research design such as the inconsistencies in IEP development and in other areas identified in the literature involving the

accuracy and consistency of how assessments are used, levels and extent of parent involvement, and differences in educator role and influence in the IEP development process (Frey, J.R., 2019).

Propensity Score Matching

Briefly explained, a propensity score is a conditional probability that expresses how likely a participant is to be found in the intervention condition given a set of observed baseline characteristics (Austin, 2011). The primary aim of conditioning on the propensity score is to achieve balance on the observed covariates and to approximate an indiscriminate group assignment that would have arisen from a randomized design. The use of observational data sets is not unusual in educational research. Because they consist of already existing data, establishing cause and effect for outcomes associated with strategies, placements and measures is challenging. Factors that determine for whom the strategies are used and under what conditions are influenced by characteristics of the subjects, or cases, as well as differences across educational environments and how professionals engage in the decision-making processes. All are especially relevant in research concerning students with disabilities and their placement for services because the nature of decision-making, though required to follow the guidelines specified in IDEA, is individualized for each student and unique through the referral, identification, and placement process. Using the observational data of educational research, subjects in what are designated as the treatment and control groups can differ in systematic ways on a number of observable or unobservable characteristics. These differences must be accounted for to rule out systematic bias within and across students, interventions, and in and out of school demographic and environmental characteristics. Propensity matching has been chosen by researchers as a method to account for these differences in base line

characteristics between subjects in treatment and control groups and eliminate the possible confounding effects of these characteristics in observational data (Austin, 2011).

To elaborate further on the need and usefulness of propensity matching, it can be used with existing data sets where the creation of randomized treatment and control groups can't be accomplished in the research design because it is a matching process that allows a balancing of the observed data, i.e., covariates, in the data across the statistically created treatment and control groups. Using this process equates them as far as these covariates are concerned as if the groups had been created through a randomized treatment and control group design and do not influence the investigated outcomes. Balance on covariates across the treatment and control groups is desirable because a balanced covariate will not skew the outcomes and the estimate of the treatment effect **(**Thoemmes & Kim, 2011). The primary aim of conditioning or establishing likelihood of assignment to a treatment or control on the propensity score is to achieve balance on the observed covariates and to approximate an indiscriminate group assignment that would have arisen from a randomized design. Balance on covariates, which means that covariates are uncorrelated with group assignment, is desirable in propensity score matching because a balanced covariate cannot itself bias the estimate of an intervention effect, even if the covariate is related to an outcome variable (Brookhart, et al., 2006).

A key assumption in propensity matching is that all relevant covariates have been incorporated into the matching process. Under this assumption, a propensity score analysis can produce unbiased causal effect estimates (Thoemmes & Kim, 2011). However, it is possible that unobserved covariates not used in the matching process can influence the treatment effect of intervention outcomes. To rule out the impact of these unobserved covariates, additional sensitivity tests can be used in research designs incorporating propensity matching in defining

treatment and controls groups to determine how strongly an unobserved covariate would need to impact group placement to negate the observed treatment effect. Knowing how sensitive the estimated treatment impact is to both the observed and accounted for variables as well as those not observed and accounted for is critical to establishing the validity of the findings.

If there are unobserved variables that affect assignment into treatment and the outcome variable simultaneously, a hidden bias might arise to which matching estimators are not robust. This problem can be addressed with a statistical determination of the strength of the impact of the unmeasured variable upon the creation of the treatment and control groups necessary to invalidate the treatment implications resulting from the matching analysis. This technique of calculating the rbounds has enjoyed considerable usage in research with an observational data base (Rosenbaum, 2002; The Stata Journal, 2007). This study applies such sensitivity analyses further described below in the section on robustness checks.

Another key assumption is that no propensity scores with the two extreme values of 0 and 1 are observed in the matching process, or in other words no unit has a guaranteed or necessary placement in one group versus another. This is important because a guaranteed placement in either group eliminates the possibility of finding a match in both the treatment and control group. This feature of the propensity matching process known as the "positivity or experimental treatment assumption" and its requirement allows a determination that for all combinations of the matching variables there will be matches across the treatment and control groups for the subject sample in the study (Westreich & Cole, 2010). This study observed these parameters in its matching process.

Outcome and Matching Variables

For the time of this study the Indiana State Test of Educational Progress (ISTEP+) administered grades 8-10 was the state assessment used for all students except those identified for the state's alternate assessment (approximately 1% of students with disabilities). ISTEP+ is a valid and reliable assessment that is administered in the spring of each school year (Roeber & Briggs, 2016). ISTEP+ scores in 10th grade were used as outcome measures. Scores for the ISTEP+ are continuous and approximately normally distributed, with ranges varying by grade level. Because this study followed cohorts of students from grade 8 through graduation, matching for all analyses was performed using 8th grade ISTEP scores. Since eighth grade ISTEP scores are used in the matching process resulting in no differences between low and high inclusion groups' eighth grade ISTEP scores, the ISTEP scores for this grade level are not part of the analyses.

The state of Indiana confers one of two types of diplomas upon graduation, a general or a Core 40 diploma. Each of these two diplomas can be awarded in one of two ways: passing the state assessment or through an evidence-based waiver. Accordingly, we operationalize four different graduation outcomes. We refer to the first outcome as *graduation*. Graduation is simply defined as having a graduation record by the Indiana Department of Education (IDOE). The second outcome (*general diploma*) regards whether a student is awarded a general diploma via passing the state assessment versus attainment via a waiver. The third outcome (*core 40*) regards whether a student is awarded a core 40 diploma via passing the state assessment versus attainment via a waiver. The fourth outcome (*waiver*) is whether a student was awarded any diploma (general or core 40) via a waiver versus by passing the state assessment. To explore the role of placement most proximal to the time of graduation, additional graduation analyses were conducted matching on 10th grade ISTEP scores.

The following student level variables were used in the matching process: 1) *Primary disability*, or the disability diagnosed as the main reason for referral, identification and placement in special education (As previously stated, because a high percentage of students with a primary disability of speech are already or generally included in general education for a high percentage of their time, they were not included in the samples or the matching process that created the control and treatment groups); 2) *Student Indiana State Test of Educational Progress (ISTEP+)* scores for 8th grade English/Language Arts and math; 3) Student *eighth grade attendance records*, in days, to ensure that students were paired with students with similar attendance records; 4) Student *gender* on record by the IDOE; 5) *Student race/ethnicity* on record by the IDOE; 6) A binary indicator of whether a student has ever been *suspended or expelled* was created for the matching process; 7) Student's status as an *English language learner* on record by the IDOE; and 8) IDOE records of whether the student receives *free lunch, reduced price lunch, or paid lunch* (free and reduced lunch have been combined in these analyses). School level variables used in the matching process included 1) the proportion of students receiving free or reduced lunch; and 2) the student body's racial and ethnic distribution. Four main categories are matched on: African American, Asian, Hispanic, and White.

We used a two-stage propensity matching approach. Stage one utilizes logistic regression to estimate each participant's propensity to be assigned to the intervention group based on a set of matching variables theoretically and empirically relevant to group assignment. The matching variables we have used here have been theoretically and empirically shown to be related to special education placement (Hibel et.al., 2010). Once propensity scores are estimated, the second stage matches individuals across intervention and comparison groups with similar assignment propensities, resulting in members of the intervention and comparison groups having

approximately equivalent propensity distributions for intervention assignment, thereby approximating a randomized control trial. The matched sample is then used to estimate an average treatment effect on the treated (ATET). In the absence of a randomized design, the ATET offers a better estimate of what this difference would have been had the comparison group exhibited similar (randomized) attributes as the intervention group. This difference is often referred to as the counterfactual (Austin, 2011; Imbens, 2004).

Factors most related to placement in high or low inclusion settings pertain primarily to students' individual characteristics, and secondarily how these characteristics may compare to other students at their school (Hibel et al., 2010). We include in the matching process student and school level variables, identified in the educational research literature as having a relationship with academic performance, and special education placement, teacher expectations, peer relationship, and school resources to ensure that matched students are alike with respect to personal and academic characteristics, while also coming from similar schools (Hibel, et al., 2010; Wells et al., 2016)

Intervention & Comparison Groups: Low vs. High Inclusion

Comparisons were made between low inclusion settings, defined as the intervention group, i.e., always receiving services in the general education classroom less than 80% of the time, and high inclusion settings where students always received services in the general education classroom 80% or more of the time. Because of the number of ways in which students may be placed in high- or low-inclusion settings across multiple years, there are no clear benchmarks for operationalizing and analyzing the effects of mixed inclusion. Therefore, mixed inclusion categories were not reported in this research.

The total eligible sample for the 10th grade English/Language Arts comparisons consisted of 23,796 students, 1,789 of whom were in the low inclusion group. For math comparisons, the total eligible sample was 23,940 students, 1,812 of whom were in the low inclusion group. Eligible sample sizes for graduation analyses varied on the type of diploma received and whether matching was done on 8th or 10th grade ISTEP scores. Matched sample sizes for all models are included in the results section.

Matching Details and Robustness Checks

In all cases, we used two-stage matching with the student and school level variables, with one-to-one nearest neighbor matching on Mahalanobis distance and a matching caliper of 0.1. A caliper of 0.1 requires that all paired individuals to be within 0.1 standard deviations of one another to be considered an acceptable match. Additionally, matching with replacement to minimize bias was used in the matched sample (Stuart, 2010). ATET estimates were used to contrast students in high and low inclusion settings (Imbens, 2004). Significance for all ATET estimates was adjusted using a Bonferroni adjustment. Rosenbaum bounds (rbounds; Rosenbaum, 2002) were used to assess the sensitivity of ATET estimates to unobserved variable bias. These were estimated via the rbounds package (Keele, 2015) in R version 4.0.3. Additionally, Kolmogorov-Smirnov (KS) tests using 500 bootstrapped samples were performed to confirm the homogeneity of matched distributions. Matching, ATET estimates, and KS tests were performed using the matching package in R (Sekhon, 2018). Unique ATET estimates are compared for each set of outcomes (10th grade ISTEP and graduation).

Sample and Analysis Plan

This study followed two cohorts of all students who receive special education services in Indiana who completed the Indiana state assessment (ISTEP+) and would have graduated

between 2013 and 2018. Together, the study's cohorts include students who were in the 6th and 7th grades in 2013. Data was obtained through a data share agreement with the IDOE; approval was received from Institutional Review Board (IRB). Students who receive special education services in Indiana receive a placement code each year corresponding with their approximate percentage of time to be spent in the general education classroom. These codes were used to identify students classified as high and low inclusion.

The sample cohorts were constructed by identifying all students who receive special education services possessing 8th and 10th grade math and English Language Arts ISTEP scores, then omitting the following students from said cohorts of students: 1) Students not receiving a placement code for one or more years; 2) Students alternating between high and low inclusion settings; 3) Any student who was in the same grade for two consecutive school years between grades 8-12; 4) Students classified as language or speech impaired for any grades 8-12; 5) Students missing any data on matching or outcome variables; and 6) Students with disabilities who took the Alternate State Assessment. Primary disability categories in this study include Cognitive Disability, Learning Disability, Autism Spectrum Disorder, Emotional Disability, Other Health Impaired, Blind/Low Vision and Deaf/Hard of Hearing.

The cohort began with 37,507 students who receive special education services with math scores across all years, and 37,521 students who receive special education services with English/Language Arts scores across all years. Following exclusions, ISTEP and graduation analyses matched on student's 8th grade records yielded a total of 23,940 students with math scores and 23,796 students with ELA scores. This resulted in a sample with only complete cases, i.e., there was no missing data analyzed. These drops in sample size were primarily the result of omitting students classified as language or speech impaired (N_{math} = 919; N_{ELA} = 920), followed

by the removal of students either not receiving placement for one or more years, or alternating between high and low inclusion settings (N_{math} = 8,391; N_{ELA} = 8,388). The remaining losses were due to missing matching variables (N_{math} = 4,257; N_{ELA} = 4,417). For graduation analyses matched on students' 10th grade records, after applying the same exclusion criteria 7,598 students with math scores and 7,617 with ELA scores remained.

While the same drops in sample size occurred from omitting students classified as speech impaired (N_{math} = 919, NELA = 920) and those either not receiving or alternating between high and low inclusion settings (N_{math} = 8,391; N_{ELA} = 8,388), the discrepancy between 8th and 10th grade complete samples comes from omitting students with missing data for matching variables (N_{math} = 20,599, N_{ELA} = 20,596). Specifically, a large number of grade 10 attendance records were missing. If we instead matched on 8th grade attendance, the number of complete cases returns to 24,085 for math analyses, and 24,050 for ELA analyses. Notably, grade 10 records were not used for matching ISTEP analyses (because grade 10 was the outcome and ongoing exclusion was the intervention). This discrepancy only applies to graduation analyses matched on 10th grade records and had no meaningful impact on the results[1].

Results

After matching, the distribution of propensity scores across low inclusion and high inclusion groups were confirmed for homogeneity. Low inclusion and high inclusion groups were made more alike across most matching variables. Using propensity score matching to control for selection bias on individual and school-level matching variables, this study found

[1] The Graduation (Any) column of Table 1 for 10th grade matching changes the math graduation rate to -.06 and the ELA graduation rate to -.04 (favoring inclusion by 6% and 4%, respectively). Math is significant after Bonferroni adjustments, and ELA is not. The sample sizes for math and ELA become 3,686 and 3,668 respectively. Gamma is unchanged. Effect sizes remain small and inconsistently significant across 8th and 10th grade matching, and thus the overall narrative remains the same.

strong evidence in support of greater inclusion in general education settings. Using a conservative p-value of 0.01 and a Bonferroni adjustment for 18 total analyses, comparisons of 10th grade ISTEP scores yield highly significant results favoring placement in the general education classroom for 80% or more time in both mathematics and English/Language Arts. For English/Language Arts comparisons, students in high inclusion settings had a higher average standardized test score of 24.3 points (ATET = -24.32, $t = -17.93$, $df = 3,569$, $p < .001$). Similar differences were seen with math scores, with high-inclusion students scoring an average of 18.4 points higher, (ATET = -18.42, $t = -13.33$, $df = 3,595$, $p < 0.001$).

Rosenbaum bounds (rbounds) for these comparisons reflect the robustness of the significance of statistical findings in the presence of influential omitted variables in the matching process. Analyses of 10th grade math scores yielded a rbound of 2.2, suggesting that an omitted variable would need to more than double the odds of being placed in low inclusion settings to negate our findings. For analyses of 10th grade ELA scores, the omitted variable would need to nearly triple the odds of placement in low inclusion settings before results were negated (rbound = 2.7). Both findings exceed the range of other rbounds commonly reported in education research (Calindo and Kopeinig, 2005; DiPrete & Gangl, 2004) thus suggesting relatively strong and robust support for our findings.

Overall findings of the graduation analyses also support high inclusion, though not all comparisons were statistically significant. After Bonferonni adjustments, differences were found in the types of diplomas awarded. Independent of whether they were matched on math or ELA scores, the probability of students in high inclusion settings to graduate with a core 40 by passing the state assessment rather than via waiver was an average of 22 percentage points higher than that of low inclusion students (ATET = -0.22, $p < 0.001$). Conversely, the probability of

graduating with any kind of waiver was between 13 (matched on 8th grade math scores) and 22 (matched on 8th grade ELA scores) percentage points higher for low inclusion students than for high inclusion students. These differences were 19 and 17 percentage points when matched on 10th grade math and ELA scores, respectively. Both findings remained highly significant ($p < .001$) even after Bonferroni adjustments. Strong differences were also found in the probabilities of receiving a general diploma via the state assessment when matched on 10th grade math scores (ATET = -0.19, $p = .002$), but these differences vanish when matching on 10th grade ELA scores and for all 8th grade matching. Findings for graduation analyses are summarized in Table 1.

Insert Table 1 here

Rbounds for the graduation analyses were all evaluated at 1.1. This suggests that they are noticeably more sensitive to the influence of omitted variables than our analyses of standardized test scores. This may be in part a product of varying sample sizes. Yet despite the sensitivity of these scores, they are within the range of other rbounds commonly reported in education research (Calindo & Kopeinig, 2005; DiPrete & Gangl, 2004).

Discussions and Implications

Special Education law and policy as it originated in Public Law 94142 and the various reauthorizations that have followed, specifically addresses whether placement and services are appropriate as it continues through requirements for Free and Appropriate Educational (FAPE) services in the Least Restrictive Environment (LRE). Indeed, they are addressed specifically with language that requires the provision of services in more restrictive settings only as a recourse when all other options are exhausted. One could argue that all things being equal the student should not do worse in the Least Restrictive Environment. This study shows that for those

students capable of being included in general education classrooms, the more restrictive environment appears to violate the principals of FAPE.

The results of this study provide additional supporting research for inclusive placement for students with disabilities and builds on the first phase of our research that focused on grades 3-8. As with the first study, the results call for a critical review of placement decisions for students with disabilities. Although one could expect that it would logically follow that students in general education most of the day would have higher academic achievement, the fact that this study found it to be true when comparing academic outcomes of students matched on a host of student and school variables associated with academic outcomes makes a clear statement for it as the preferable placement.

The findings of this study contribute to the transition literature by shedding light on the relationship between placement decisions and academic outcomes for students with disabilities and predict the high school graduation pathways for students with disabilities. These graduation pathways are associated with types of diplomas earned and how students with disabilities complete high school. How a student completes high school, in turn, influences the post-secondary experiences critical to successful life experiences. This data has clear implications for decisions made during transition planning. Specifically, it suggests that parents and students be given enough information about the differential outcomes for high/low inclusion placements, especially how they may relate to post-secondary options. The transition planning process should go beyond the immediate concerns of academics and placement and incorporate possible post-secondary possibilities and what it takes for students to be prepared for success beyond high school. For this to happen, training of administrators/special educators and monitoring of the transition planning process is critical.

By using diploma type as an outcome measure, this study forged linkages between inclusion and post-secondary options. To that end, the differential access to Core 40 diplomas is an important finding and while this study did not review data on the comparative rigor of the courses taken during a student's high school career, the course requirements for the Core 40 are different and more rigorous than the courses associated with other graduation pathways. This difference in rigor explains post-secondary choices and experiences because it would follow that completion of a college prep program in high school would ensure greater opportunity for post-secondary educational success. The findings related to diploma type also speak to differential expectations; students in high school who are included have greater access to courses with expectations for post-secondary education while students in non-inclusive courses were led to pathways with lower expectations resulting in either a general diploma or a diploma with a waiver. These different pathways are certainly related to the disparity in post-secondary choices and successful life experiences described in the research concerning students with disabilities and their general education peers.

The 10th grade ISTEP score results suggest the need for educators to review and revise curriculum and instruction in non-inclusive settings. If we accept that students are receiving those placements "legitimately," we are faced with data that indicate lower expectations/outcomes. Even more importantly, the results of this study highlight the impact of not providing inclusive placements in general education classroom which limits opportunity and creates a stereotypical attenuation of success over the course of their lives.

These findings highlight questions about the pedagogy in low inclusion settings. The need to ensure that placement in a more restrictive environment does not mean less effective instruction and services that result in a less desirable educational experience is critical when

deciding the best placement for delivering services to students with disabilities. It may be that rather than try to "put new wine", i.e., pedagogical excellence, into an old flask, i.e., separate classrooms, it is more effective to simply enhance the ability of students with disabilities to be included in general education classrooms with their peers. However, making this happen will require the evolution of policy and the continuous training of both general and special education teachers focused on the skills and dispositions necessary to teach students with disabilities in inclusive settings. While this study did not look at the nature of the high inclusion settings (e.g., co-taught, consult, etc.) it would seem that there must be qualitative differences between the instructional delivery in separate classrooms when compared to the pedagogy experienced by students with disabilities in the general education classrooms.

If being a part of a general education classroom results in better academic outcomes, then it is imperative that policies, practices, and resources are focused on how to not exclude them in the first place. Anderson et al. (2020) suggest that educational teams should make placement decisions that treat the general education classroom as the default classroom and should only remove a student when there is compelling evidence to do so. We would agree. It is important to note that the argument is not being made that every student across the spectrum of needs must be included regardless of resource inability or teacher preparation, although one could make the case that adhering to a moral imperative would be a reason to do so. We are not arguing that for some students the general education classroom may prove to be inappropriate or impracticable. However, for the students in this study, it appears that not pursuing this option results in a lower academic trajectory than they would otherwise have achieved. Thus, the challenge in the decision-making process is to establish a monitoring framework that orients decision-making toward keeping students with disabilities in the general education classroom.

This monitoring framework must also include the supports necessary for student success. Morningstar et.al., (2018) examined how college and career ready skills can be extended within the multitiered systems of supports (MTSS) to improve post-secondary outcomes. The use of a comprehensive array of student outcome data in a systematic fashion facilitates objective decision-making and evaluation of program and strategy effectiveness. Additionally, practices associated with how Fidelity of Implementation (FOI) of school-wide programs and strategies is monitored allows for programmatic adjustments in resources and processes to be made in a timely fashion. The criteria to determine fidelity of implementation also enable the decisions concerning the use of effective programs and strategies to broaden their focus to include systemic processes and resources while evaluating the appropriateness and capability of the chosen intervention to address individual student needs. Finally, by creating a data framework that includes both academic and non-academic measures in an MTSS system elevates a number of characteristics related to post-secondary success and not normally a part of a school improvement effort into the set of school priorities. By doing so, the programs and strategies necessary to address expanded outcomes become critical to the overall student and school success. However, developing and accessing valid and reliable measures will be critical to this integration of non-academic outcomes related to College and Career Readiness and transitions (Lombardi et.al., (2021).

Finally, we believe that the inclusion of students with disabilities is an issue of social justice. We would argue that “waiting until the system is ready” to include contradicts the concept of social justice. The search for social justice within the educational system is important because educational outcomes reflect societal inequities for those perceived as “different” (Annamma & Handy, 2020). We cannot, as a society, afford to continue to support policies and

practices that result in academic failure, limited post-secondary options and continued separation and marginalization based on disabilitys We can, however, accept the ambitious agenda to transform educational systems to create inclusive school environments, maximize student participation and increase the achievement of students with disabilities.

Limitations and Future Research

As noted in the discussion, the study does not address qualitative differences in inclusive versus separate classroom experiences. How and why teaching and learning experiences differ is important because it enables those instructional practices that facilitate student learning to be replicated while enabling educators to eliminate those that are not beneficial. Additionally, the study did not review, compare, and contrast the differences in the courses taken along the various diploma pathways. These are important research areas because what is learned from them can influence policy, placement decisions, professional development and support that creates a sense of efficacy in teachers for including students in general education classrooms. Further, there are some logical concerns about inclusive settings; preparation of general education teachers to teach students with disabilities, resource allocation, and effective leadership to name a few. Additionally, the use of reading and math scores could be acknowledged as a limitation because students receive services to address other academically related needs in IEP's that are unique to each student. However, reading and math are fundamental academic subject matter. Moreover, we matched on 8th grade ISTEP scores. As a result, the study's method is comparing students who have the same reading and math scores in time one and a different set in time two with the included students doing better. Reading and Math are core academic subjects and have a direct relationship with academic achievement and diplomas. Students who can't read or do math are not likely to get a general education diploma because there is a relationship between reading and

all subject matter required for graduation, and a relationship with math and some other critical subject matter like science. We believe that the FAPE placement should not result in lower reading and math achievement than the student is capable of.

One could also argue that the IEP process is differentiated in nature and unique to each student, and that these differences make drawing conclusions from the results spurious. However, the study accounted for differentiated needs specific to IEP's, i.e., accommodations, etc., through the propensity matching process using the rbounds statistic which is used for detecting bias from unmeasured but related variables. The rbounds computation indicated the results are free of any significant influence from unobserved and potentially confounding variables. The results of the rbounds sensitivity measure indicates that other IEP characteristics did not affect the outcomes in a systematic manner. The matching process accounted for the influence of any unaccounted-for variables, like IEP differences. However, not having IEP specific data is a limitation that could be addressed in future research with a different research design.

This study controlled for the influence of student and school characteristics upon academic outcomes in the propensity matching process. A different design, i.e., multi-level modeling, could look at the within and between school influences of these characteristics in a different way. However, it is our opinion that if differences within and between schools determine whether inclusion is successful it means that the educational personnel involved in placement decisions must address and remove how they become barriers to student success in the general education classroom.

Another area of critical importance in this research is to identify whether there are student, teacher, administrator, and school characteristics that predict placement. Ferreting out

these relationships will enable educators to be reflective in their decision-making so that the influence of these relationships is understood when making placement decisions. Understanding the influence of student and school characteristics upon their decision-making enables those involved to be intentional in overcoming the limitations they present. This informed decision-making will enable placement based upon expectations for student success and a belief in student capabilities rather than decisions limited by concerns unrelated to student needs. Overcoming the apprehension of placing students with disabilities in general education classrooms will allow students to have a k-12 academic experience that prepares them for successful post-secondary life outcomes.

Finally, the nature of the student sample in this study is a limitation in that low incidence, high need students with disabilities were not included. These students often present difficult and resource intensive challenges to their inclusion in the general education classroom. Understanding the relationship between inclusive placements and the 1% of students with disabilities who take the alternate assessment is an area worthy of research. In some respects, ensuring that these students are provided the intensive supports necessary for their success is the greatest challenge in k-12 education. This is also true of their post-secondary opportunities. How to best address the needs of these students with the most intensive needs in our schools and classrooms and enable them to be successfully integrated into our communities remains one of our greatest challenges.

References

Afacan, K. & Wilderson, K. (2021). Reading outcomes of students with intellectual disability on statewide assessments. *Journal of Intellectual Disabilities*. doi:10.1177/1744629521991409

Anderson, E.J. & Brock, M.E. (2020). Inclusion; District of Columbia, 8(3), 210-221.doi:10.1352/2326-6988-8.3.210

Annamma, S.A., & Handy, T. (2020). Sharpening justice through DisCrit: A contrapuntal analysis of education. *Educational Researcher, 50*(1), 41-50.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3), 399–424. https://doi.org/10.1080/00273171.2011.568786

Barrett, C., Stevenson, N.A., & Burns, M.K. (2020). Relationship between disability category, time spent in general education and academic achievement. *Educational Studies, 46*(4), 497-512.

Brookhart, M.A., Schneeweiss, S., Rothman, K.J., Glynn, R.J., Avorn, J., & Stürmer, T. (2006). Variable selection for propensity score models, *American Journal of Epidemiology*, *163*(12), 1149–1156.

Calindo, M. & Kopeinig, S. (2005). Some practical guidance for the implementation of propensity score matching. [IZA Discussion Paper No. 1588] Bonn, Germany: Institute for the Study of Labor. Retrieved from http://ftp.iza.org/dp1588.pdf

Causton-Theoharis, J., Cosier, M., Theoharis,G., & Orsati, F. (2011). Does self-Contained special education deliver on its promises? A critical inquiry into research and practice. *Journal of Special Education Leadership 24*, 61–78

Cook B.G. & Cook, L. (2020). An examination of highly cited research on inclusion. In On Educational Inclusion, Routledge: London.

DiPrete, T. & Gangl, M. (2004). Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. *Sociological Methodology, 34*, 271-310. https://journals.sagepub.com/doi/10.1111/j.0081-1750.2004.00154.x

Francisco, M., Hartman M, Wang Y. (2020). Inclusion and Special Education. *Education Sciences*. *10*(9), 238. https://doi.org/10.3390/educsci10090238

Frey, J.R.(2019). Assessment for special education: Diagnosis and placement. *The ANNALS of the American Academy of Political and Social Science*. *683*(1),149-161. doi:10.1177/0002716219841352).

Hibel, J., Farkas, G., & Morgan, P. L. (2010). Who Is Placed into Special Education? *Sociology of Education*, *83*(4), 312–332.

Imbens G.W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics*, *86,* 4–29.

Keele, L. J. (2015) Rbounds: An R package for sensitivity analysis with matched data. R. package version 2.1. Retrieved from https://cran.r-project.org/web/packages/rbounds/rbounds.pdf.

Landrum T.J., Tankersley, M., Kauffman, J.M., (2003). What Is Special About Special Education for Students with Emotional or Behavioral Disorders? *The Journal of Special Education. 37*(3),148-156. doi:10.1177/00224669030370030401 https://cran.r-project.org/web/packages/rbounds/rbounds.pdf.

Lombardi, A., Doren, B., Gau, J. M., & Lindstrom, L. E. (2013). The influence of instructional settings in reading and math on postsecondary participation. Journal of Disability Policy Studies, 24(3), 170-180.

Lombardi, A. R., Rifenbark, G. G., Poppen, M., Reardon, K., Mazzotti, V. L., Morningstar, M. E., & Raley, S. K. (2021). Development and Validation of the Secondary Transition Fidelity Assessment. Assessment for Effective Intervention, 15345084211014942.

Mazzotti, V.L., Rowe, D.A., Kwiatek, S. et al. (2021). Secondary Transition Predictors of Postschool Success: An Update to the Research Base. *Career Development and Transition for Exceptional Individuals. 44*(1):47-64. doi:10.1177/2165143420959793

Morningstar, M. E., Lombardi, A., & Test, D. (2018). Including college and career readiness within a multitiered systems of support framework. AERA Open, 4(1), 2332858418761880.

National Center on Education Statistics (2019). Retrieved from https://nces.ed.gov/pubs2018/2018070.pdf

Rojewski, J. W., Lee, I. H., & Gregg, N. (2015). Causal effects of inclusion on postsecondary education outcomes of individuals with high-incidence disabilities. Journal of Disability Policy Studies, 25(4), 210-219.

Rosenbaum, P.R. (2002). *Observational studies*. (2nd ed.). New York, NY: Springer

Sekhon, J. (2018). Multivariate and propensity score matching software with automated balance optimization: The matching package for R. *Journal of Statistical Software*, *42*.

Stuart E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *25*(1), 1–21. doi:10.1214/09-STS313

The Stata Journal, (2007). Sensitivity analysis for average treatment effects. *7*(1),71–83.

Thoemmes, F., & Kim, E.S., (2011). A Systematic Review of Propensity Score Methods in the Social Sciences. *Multivariate Behavioral Research, 46(1),* 90-118

U.S. Department of Education, (2020). State performance plan and APR forms. Retrieved from https://www2.ed.gov/fund/data/report/idea/sppapr.html

U.S. Department of Education, Office of Special Education and Rehabilitative Services, Office of Special Education Programs, *42nd Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act, 2020*, Washington, D.C. 2021.

Wehmeyer, M., Shogren, K.A. & Kurth, J. (2021). The state of inclusion with students with intellectual and developmental disabilities in the United States. *Journal of Policy and Practice in Intellectual Disabilities. 18*(1)

Westreich, D., Cole, S.R. (2010). Invited commentary: Positivity in practice. *American Journal of Epidemiology*, *171*(6), 674–677, https://doi.org/10.1093/aje/kwp436

Wells, Fox & Cordova-Cobo, (2016). How racially diverse schools and classrooms can benefit all students? The Century Foundation. Retrieved from https://tcf.org/assets/downloads/HowRaciallyDiverse_AmyStuartWells.pdf

Williamson, P., Hoppey, D., McLeskey, J., Bergmann, E., & Moore, H. (2019). Trends in LRE placement rates over the past 25 years. *The Journal of Special Education, 53*(4*), 236-244.*

Table 1

Findings of propensity score matching analyses.

		10th grade Math ISTEP	10th grade ELA ISTEP	Graduation (Any)	Graduation (General)	Graduation (Core 40)	Graduation (Waiver)
8th Grade Matching	Math	-18.42*** *Γ = 2.2* (N = 3,596)		.002 *Γ = 1.1* (N = 3,968)	-.05 *Γ = 1.1* (N = 382)	-0.21*** *Γ = 1.1* (N = 248)	0.13*** *Γ = 1.1* (N = 654)
	ELA		-24.32*** *Γ = 2.7* (N = 3,932)	-.04* *Γ = 1.1* (N = 3,938)	-0.11 *Γ = 1.1* (N = 376)	-.22*** *Γ = 1.1* (N = 248)	0.22*** *Γ = 1.1* (N = 638)
10th Grade Matching	Math			-.03 *Γ = 1.1* (N = 1,090)	-0.19** *Γ = 1.1* (N = 396)	-0.23*** *Γ = 1.1* (N = 276)	0.19*** *Γ = 1.1* (N = 688)
	ELA			-.06 *Γ = 1.1* (N = 1,086)	-0.11 *Γ = 1.1* (N = 396)	-0.22*** *Γ = 1.1* (N = 276)	0.17*** *Γ = 1.1* (N = 688)

Note: *$p < .05$. **$p < .01$. ***$p < .001$

Γ statistic refers to rbounds

ISTEP: Indiana Statewide Testing for Educational Progress