# The Impact of Inclusion on Student Academic Outcomes

Sandi M. Cole, Hardy R. Murphy, Michael B. Frisby, Teresa A. Grossi, Hannah R. Bolte

This is the author's manuscript of the article published in final edited form as:

Cole, S. M., Murphy, H. R., Frisby, M. B., Grossi, T. A., & Bolte, H. R. (2020). The Relationship of Special Education Placement and Student Academic Outcomes. The Journal of Special Education, 0022466920925033. https://doi.org/10.1177/0022466920925033

**Abstract**

This study investigates the academic outcomes (Indiana State Test of Educational Progress or ISTEP ELA and Math scores), of a cohort of special education students in the state of Indiana placed in high, mixed and low inclusion settings. Students are followed from the third grade in 2013 through the eighth grade in 2018.

The methodology included comparative analyses of academic outcomes conducted between treatment and control group outcomes for students designated as high inclusion, mixed inclusion, and low inclusion. Propensity score matching was used in the creation of treatment and control groups to improve the balance of primary disability type and performance distributions.

Results of this study show that students with disabilities who spend 80% or more of their time in a general education inclusive classroom do significantly better in both reading and math assessment than their peers who spend more time in separate special education classrooms.

*Keywords:* Inclusion, Access into the General Education curriculum, academic outcomes, impact

## The Impact of Inclusion on Student Academic Outcomes

The Every Student Succeeds Act (ESSA, 2015), with its focus upon college and career readiness, continued the interest upon the factors influencing outcomes for public school students. Together with the Individual with Disabilities Education Act (IDEA) and its focus on students with disabilities, these two legislative landmarks encompass the range of challenges and opportunities in public education. However, students with disabilities experience the impact of these respective educational legislative initiatives differently than their general education peers through the creation of what Burrello, Lashley, and Beatty (2001) describe as a separate subsystem of services for those not fitting the dominant system.

As with all students, successful post-school opportunities for students with disabilities are more likely with high quality educational experiences prior to graduation. Further, what students experience in early elementary and middle grades is important to their preparation for high school and beyond (Cole, Waldron, & Majd, 2004; Conn-Powers, Crawford, Dixon & Hall, 2017; Grossi & Cole, 2008; McLeskey, Rosenberg, & Westling, 2013). Given the funding invested into special education, the number of students receiving services and evidence that these students lag behind their peers in terms of K–12 outcomes, research into strategies that make a difference in these outcomes for students with disabilities is timely (Goodman, Bucholz, Hazelkorn & Duffy, 2014; Wagner, Cameto, Knokey & Shaver, 2010;). Strategies that dismantle the practices that create the separate system described by Burello, et.al. (2001) are of particular interest.

## Literature Review

In 2016, more than 60 percent of all students with disabilities spent 80 percent or more of their day in a general education classroom (Gilmour, 2018). Arguably, there has not been an issue more contentious than placement; inclusion of students with disabilities is the most controversial issue related to the education of students with disabilities (Bateman, Tankersley, Wills-Floyd & Alves, 2015; Farrell, 2010; Kauffman & Badar, 2014; Slee, 2011).

To some in the field of special education, inclusion symbolizes the dismantling of the hard-fought right for children with disabilities to receive special education services and all that is good about existing services (Farrell, 2010; Kauffman & Hallahan, 2005; Kauffman & Badar, 2014; Slee, 2011). However, proponents of inclusion see it as a philosophy that challenges 'ableism' and a value related to the civil rights of individuals with disabilities (Gerrard, 2006, Hehir, 2007, Runswick-Cole, 201;Villa & Thousand, 1995, 2005). Another, equally passionate and somewhat cynical view is that inclusion is meaningless and only a catchword used to give some legitimacy to whatever program or intervention people want to defend (Kauffman, 1999).

**Inconclusive Evidence**

Case law decisions characterized by a lack of consistency in judicial reasoning from which no precedent can be established (Lupini & Zirkel, 2003; Newcomer & Zirkel, 1999). The number of cases in which the courts ruled for a more inclusive placement is about equal to the number of rulings for less inclusive placements. Even though inclusion became more common in the past decade, it is not because of a robust evidence base that supports its effectiveness (Gilmour, 2018). It is likely that parental advocacy is the reason. Parents win slightly more often than districts when seeking inclusive placement, and when seeking a less inclusive placement, districts overwhelming win by 8:1 (Bateman, et al., 2015).

Zigmond (2015) argues that research evidence on the relative efficacy of one special education placement over another is scarce, methodologically flawed, and inconclusive. She notes that the research does not support the superiority of any one service delivery model over another and writes that the setting is less important than what is going on in the setting. Kalambouka, Farrell, Dyson, and Kaplan (2007) and Lindsay (2007) investigated the effects of inclusion of children with mild special educational needs, but neither drew definitive conclusions on the effects of including this group of students. Kavale and Forness (2000) state that the role and validity of the research on inclusion has been contentious with respect to the research methodology and recognize the need for more empirical evidence on inclusion.

A major limitation of research to date is the failure to account for selection bias. Students with higher academic ability or with fewer behavioral concerns are more likely to be placed in inclusive settings and the findings often reflects this bias (Gilmour, 2018). Because LRE assignment is not randomized; i.e., placement for special education services is a deliberative and intentional process; a random treatment design cannot be used to investigate the research questions and test hypotheses associated with high and low inclusion. The inclusion research literature cites this structural bias in the placement process as an explanation for the dearth of research into the relationship between inclusion and academic outcomes. Rojoewski, Lee, and Greg (2015) postulate: “Lack of experimental control can explain the absence of comparative studies on inclusion. While random assignment is the best way to eliminate experimental bias, this is neither feasible nor desirable with educational interventions due to ethical logistic, and legal barriers” (p. 211*)*.

**Merits of Inclusion**

Some studies have demonstrated the merits of including students with disabilities, finding that there are academic benefits for students with disabilities who are included in the general education classroom. Cole, Waldron and Majd (2004) studied the effects of inclusive school settings for students in six school corporations. A review of group means and the percentage of students making comparable or greater than average academic progress when compared to students without disabilities indicates a pattern in favor of inclusive settings, namely for students with learning disabilities and mild mental disabilities.

Recent research has employed designs to address some of the limitations of previous research in order to reach more definitive conclusions. Theobald, Goldhaber, Gratz & Holden (2017) used an analytic approach that controlled for a variety of student characteristics, including baseline measures of performance. They found that students with disabilities who spend 80% or more of the school day in general education classrooms had fewer absences, higher academic performance, higher rates of grade progression, on-time graduation, and college attendance and employment.

Lombardi, Gau, Doren, & Lindstrom, (2013) examined instructional settings in language arts and math as predictors of participation in post-secondary education at two- and four-year institutions for youth with disabilities. Results indicated students who received math and language arts instruction in a general education classroom were more likely than students in more segregated settings to participate in a two-year postsecondary education program.

More recently, examining high school inclusion (80% or more credits taken in general education classes) as a predictor of post-school education outcomes for students with learning disabilities and emotional disturbance, Rojewski, Lee, & Gregg (2015) found a significant positive relationship between students in inclusive high school placements and participation in postsecondary education 2-years after high school.

**Support for Special Education Classrooms**

Some research evidence contradicts findings supporting inclusion. Rogers and Thiery (2003) investigated whether students with learning disabilities performed better in an inclusive setting or in a setting in which they had reading lessons in separate classes. The results showed a negative effect of inclusion: four out of five students showed a decrease in performance after switching their reading lessons to the inclusive setting. Manset and Semmel (1997) compared eight inclusion models for elementary students with high-incidence disabilities, primarily Learning Disabilities, reported in the research literature between 1984 and 1994. They note that inclusive programs can be effective for some, although not all, students with high-incidence disabilities. Waldron and McLeskey (1998) agreed with this conclusion. In their research, students with severe LD made comparable progress in reading and math in pull-out settings, although students with mild LD were more likely to make gains commensurate with nondisabled peers when educated in inclusive environments than when receiving special education services in a resource room.

Other research is consistent with these findings and reports disappointing or unsatisfactory academic and social achievement results from inclusion models. Using a comprehensive data base to look at the relationship between inclusion and academic achievement at the school level, Dyson, Farrell, Polat and Hutcheson (2004) found no evidence of a relationship between inclusion and attainment at the school level and found a very small and negative statistical relationship between the level of inclusion and the attainment of its students.

Kauffman & Hallahan (1997) argue that the dilemma of providing the differentiation needed for inclusion to be successful is insurmountable since all instruction cannot be offered with the same effectiveness in the same place and at the same time. Kauffman (1993) also states that it is necessary to keep “place” in perspective as setting has limited impact on the outcomes for students with disabilities.

## Research Design

**Methods and Motivating Factors**

This study used propensity score matching to create comparison groups, i.e., treatment and control, to address the effects of structural bias identified as a limitation in special education research. Propensity score matching is a quasi-experimental design that allows for the closer approximation of causal relationships in the absence of randomized control trials. The goal of propensity score matching is to eliminate selection bias into either of two groups by selecting, as best as possible, twins from both groups based on a set of relevant and observed covariates (Becker & Ichino, 2002). By generating two groups which are approximately homogenous on variables pertaining to placement, subsequent discovery of an effect related to placement is therefore less confounded with placement variables, thus lending stronger support to a causal claim. Moreover, propensity score matching is used in similar research on special education students (Morgan, Frisco, Farkas, & Hibel, 2008; Theobald, Goldhaber, Gratz & Holden, 2017)).

**Sample and Analysis Plan**

This study follows a single cohort of Indiana students with disability in order to assess the relationship between high, mixed and low inclusion settings with school outcomes. Students are followed from the third grade in 2013 through the eighth grade in 2018. Data was collected through the use of data share agreements with the Indiana Department of Education (IDOE). Special education students in Indiana receiving a placement code each year dictating their approximate percentage of time spent in the general education classroom were eligible for the study and selected on the following criteria. First, students not receiving a placement code for one or more years were not eligible for analysis. Second, students had to consistently progress through the schooling system each year for grades 3-8. In other words, students could not be held back a year. Third, any student who was in the same grade for two consecutive school years between grades 3-8 was removed from the study. Fourth, eligible students could not be classified as having a language or speech impairment. Students classified as language or speech impaired for any grade grades 3-8 were removed from the study. Fifth, students had to have math scores for all grades 3-8 (ISTEP++), ELA scores for all grades 3-8 (ISTEP++), 3rd grade reading scores (IRead), and 3rd grade attendance records. Students missing any of these in the 3rd grade were omitted from the study. If a student was missing a non-3rd grade ELA score, but had all of their math scores, the student was considered only for the analyses of math scores, and was omitted from the analyses of ELA scores. Students missing non-3rd grade math scores but having all of their ELA scores were not eligible for the math analyses, but were used in the ELA analyses. Finally, two categories representing emotional disability (full-time and other) have been combined in this study.

**Average Treatment Effect on the Treated (ATT)**

Following propensity score matching, ATT estimates were used to evaluate differences between the treatment and control groups. In a randomized design, it can be assumed that no selection bias affects placement into the treatment and control groups. Consequently, the treatment effect should equal the difference in the means of the two groups. In the absence of such a design, the ATT is used to estimate what the difference would have been had the control group exhibited similar (randomized) attributes as the treatment group.

**Treatment & Control Groups**

Four higher level analyses (two sets of two) were conducted, with three sub-analyses within each level. While the three sub-analyses remain fixed, the four higher level analyses vary by definitions of treatment and control groups. These analyses are summarized in Table 1.

We defined high, mixed, and low inclusion as follows. Students receiving placement in general education classrooms for 80% or more of their school day for all grades 3 through 8 were considered *high inclusion*. Students receiving placement in general education classroom for less than 80% of their school day for at least one school year (but not all school years) between grades 3 through 8 were considered *mixed inclusion*. Students who never received placement in a general education classroom for 80% or more of their school day for all grades 3 through 8 were considered *low inclusion*. These three definitions were used to define and vary the treatment and control groups in our two sets of propensity score analyses.

In the first set of analyses, we compared high inclusion students (treatment)—students consistently spending 80% or more of their time in a general education classroom for all grades 3 through 8—to mixed inclusion students (control in analysis 1), and low inclusion students (control in analysis 2). With each treatment and control defined, students were compared in several sub-analyses. First, students were evaluated for math and ELA score differences (ATT) while not controlling for their specific disabilities (*ignoring disability codes* in Table 1). In these analyses, students' disabilities were used as neither a matching nor control variable. Second, students were evaluated for math and ELA score differences (ATT) after matching them on their disabilities (*matching on disability codes* in Table 1). As a result, the treatment and control groups were approximately equally represented by students with each disability. Lastly, students were evaluated for math and ELA score differences (ATT) when looking only at other students with their same disability (*within unique disability codes* in Table 1). In these analyses, the data is filtered to only evaluate students with specific disabilities, thereby providing a within-disability comparison of students in high-, mixed-, and low-inclusion settings. Many specific disabilities are limited in available sample size, and thus the within-disability analysis were limited to emotional disability, specific learning disability, mild cognitive disability, and autism spectrum disorder.

The second set of analyses was motivated by historical and statistical considerations. Historically, the original legislation to provide a free and appropriate education to students with disabilities established special education as a "treatment" to support the learning of this group of students. Yet, many studies have considered inclusion as the 'treatment." In fact, the preliminary analysis for this study assigned inclusion as the treatment group and the special education classroom as the control. This discussion and the statistical considerations below led us to a second analysis.

Numerically, we find that the number of high-inclusion students is often greater than the mixed-inclusion students by a factor of 2 to 1, and greater than the low inclusion students by as high as a factor of 34 to 1. Since the matching process resamples from the control group to generate a one-to-one match for the treatment group, this may result in a resampling of a much smaller pool of students whenever the ratio is greatly skewed in favor of the treatment group, which may in turn distort the quality of the matched sample. As such, the second set of analyses reverses the treatment and control groups in the first set of analyses, thereby making the smaller group the treatment group, and resampling from the larger control group. The control group in the second set of analyses consists strictly of high-inclusion students, while the treatment group varies between mixed- and low-inclusion students. All sub analyses remain the same as those in the first set of analyses (see Table 1 for details).

Place Table 1 about Here

**Matching and Outcome Variables**

**Inclusion Category**. This variable is defined above in the *Treatment & Control Groups* section, and consists of three categories: high-, mixed-, and low-inclusion. Depending on the analysis, inclusion category served as the basis for the definition of being in the treatment or control group. As such, it is not a matching variable, but is the outcome variable of the matching process.

**IRead Scores**. At the end of a student's third grade year, the state of Indiana administers a standardized reading test, IRead, to assess the student's reading ability. Scores for the IRead are continuous, approximately normally distributed, and range from 300 to 650. Students' third grade IRead scores were used as a matching variable.

**Primary Disability**. Primary disability is a matching variable in the sub-analysis *Matching on Disability Code*, and was a variable used to look within disabilities in the *Within Unique Disability Codes* sub-analysis.

**ISTEP+ Scores**. All students with disabilities in Indiana who took the state assessment (ISTEP+) were included in the matching process. Scores for the ISTEP+ are continuous and approximately normally distributed, with ranges varying by grade level. ISTEP+ scores in grades 4 through 8 were used as outcome measures. For analyses of ELA and math scores, third grade scores were used as a matching variable.

**Attendance**. Student third grade attendance records, in days, was used as a matching variable such that students should be paired with students with similar attendance records.

**Matching Details and Robustness Checks**

This research matched on three key variables: students' third grade attendance records, students' third grade reading ability as determined by reading assessment scores (IRead), and students' state assessment scores in either the ELA or mathematics (ISTEP+) content domain corresponding to the outcome of interest. In all cases, one-to-one matching on Mahalanobis distance was used with a matching caliper of 0.1, thereby requiring all paired individuals to be within 0.1 standard deviations of one another to be considered a match. Rosenbaum bounds (Rosenbaum, 2002) were used to assess the sensitivity of ATT estimates to unobserved variable bias. These were estimated via the *rbounds* package (Keele, 2015) in R version 3.5.0. Additionally, Kolmogorov-Smirnov (KS) tests were performed using 500 bootstrapped samples to assess the homogeneity of matched distributions. Matching, ATT estimates, and KS tests were performed using the *matching* package in R (Sekhon, 2018). Unique ATT estimates are compared for each content area for each grade ($4^{th}$ – $8^{th}$), and for every sub-analysis within each higher level analysis (see Table 1).

## Results

In broad strokes, our analyses reveal overwhelming support for special education students in high-inclusion settings. Approximately 96% of our results were found to favor high-inclusion for special education students—that is, the average ELA and mathematics ATT estimates suggest that special education students in high-inclusion settings performed better than their mixed- and low-inclusion peers. Roughly 65% of all analyses were both favoring high-inclusion and were statistically significant. Out of all analyses, less than 4% disfavored high-inclusion, and less than 1% disfavored high-inclusion while being statistically significant. KS-tests overwhelmingly supported homogeneous distributions of matching covariates across treatment and control groups, while Rosenbaum bounds suggest some sensitivity to unobserved variable bias affecting placement into the treatment and control groups (min = 1.2, max = 2.2). These bounds (Calindo and Kopeinig, 2005; DiPrete, & Gangl, 2004) are within the range of what is commonly reported in education research. Below, we review each set of sub-analyses in greater detail, starting with the analyses in which we did not control for student disability, then turning to the analyses matching on student disability, and lastly looking within unique disability categories.

**Not Controlling For Disability**

Each set of analyses (2) has two variations (2) to account for different definitions of treatment and control, with each one estimating average treatment effects for each unique grade 4-8 (5) for both math and ELA content domains (2). The result is 2 x 2 x 2 x 5 = 40 comparisons. Of the 40 comparisons of students not controlling for student disability, 92.5% of them are statistically significant and favor high-inclusion settings for students with disabilities. ATT estimates suggest that means are in the direction favoring high-inclusion settings. Specifically, ATT estimates for 2018 ELA scores ($\Delta = 24.14$, n.s.), 2014 math scores ($\Delta = 3.66$, n.s.), and 2015 math scores ($\Delta = 8.01$, n.s.) for high-inclusion versus low-inclusion students directionally favored high-inclusion, but were not statistically significant. Results are summarized in Table 2.

Insert Table 2 About Here

**Matching On Student Disability**

Next, we hypothesized that some of the significant differences may be related to a potential imbalance in student disabilities represented in high-, low- and mixed-inclusion settings. To assess this, we added students' primary disability as a matching variable such that the treatment and control groups would be balanced with respect to each unique student disability. Findings suggested that even with a balance of student disabilities, special education students in high-inclusion settings still fared better than their peers in mixed- and low-inclusion settings. Out of all 40 analyses, 95% of them favored high-inclusion settings for special education students and were statistically significant. Of the 5% that were not statistically significant, they continued to directionally favor high-inclusion settings. Specifically, ATT estimates for 2018 ELA scores ($\Delta$ = 29.92, n.s.) and 2015 math scores ($\Delta$ = 16.99, n.s.) for high-inclusion versus low-inclusion students directionally favored high-inclusion, but were not statistically significant. These non-significant results again come from analysis set 1, which may be afflicted by oversampling of a smaller population. Results for this second set of analyses are summarized below in Table 3.

Insert Table 3 About Here

**Disability Category Only**

In light of what appears to be overwhelming support for special education students' placement in high-inclusion settings, we explored whether or not such settings may have differential effects for students with specific disabilities. In other words, we asked whether high-inclusion settings matter for some disabilities, but not for others. Limitations in sample sizes prevent us from looking at every disability in great detail. We have looked at those in which sample size arguably permits, though it is evident that the sample sizes in some comparisons are unacceptably small to yield generalizable results. The primary disabilities we considered are limited to emotional disability, specific learning disability, mild cognitive disability, and autism spectrum disorder.

The direction of the results generally favors high-inclusion settings for students with disabilities, regardless of the specific disabilities they have. However, statistical significance for these results is spotty—varying to some degree by year and by analysis. Moreover, there were some (few) cases where high-inclusion settings were disfavored and statistically significant. For students with emotional disabilities, 90% of results favored high-inclusion settings, but only 20% of results both favored and were statistically significant. Of the 10% that disfavored high-inclusion, none of them were statistically significant. For students with specific learning disabilities, 100% of results favored high-inclusion settings, and 80% of these were statistically significant. Students with mild cognitive disabilities also showed strong results favoring high-inclusion settings. The results showed that 97.5% of analyses directionally favored high-inclusion settings for students with mild cognitive disabilities, while 42.5% were statistically significant. The one analysis disfavoring high-inclusion—ELA 2014 in analysis set 2—was non-significant. Lastly, students with autism spectrum disorder showed the greatest degree of inconsistency, but the inconsistent results are largely confined to comparisons of high- and low-inclusion students in analysis set 1, which is marred with issues of sample size (only four students were consistently in a low-inclusion setting). Nevertheless, results are reported here for completeness, though they should be interpreted cautiously. These results suggest that 90% of analyses favor high-inclusion settings for students with autism, and 60% of them are both favoring and statistically significant. 10% of results disfavor high-inclusion settings, and 5% of the forty analyses both disfavor and are statistically significant. Results for all specific disabilities are summarized below in Tables 4, 5, 6, and 7.

Insert Tables 4, 5, 6, and 7 About Here

## Discussion

Prior to this research, the fundamental question of whether 'place A' or 'place B' is better for students with disabilities could not be answered with a definitive statement. As noted in the literature review, to date, evidence of whether students with disabilities learn more in one placement over another is at best inconclusive. We have addressed the question of relationships between 'place,' 'service delivery,' and outcomes through the use of a method, i.e., propensity matching that incorporates educational outcomes from elementary through middle school and addresses some of the confounding variables resulting from bias in the placement process. Thus, the results shed a more definitive light on relationships between inclusion and academic outcomes.

For students with a disability, placement decisions at IEP case conferences use a variety of data with a range of subjectivity to determine whether services are provided in separate, self-contained classrooms or in general education classrooms. This placement decision-making process can be improved by the findings of this research that clearly identifies relationships between placement and educational outcomes.

Parents and guardians are the best advocates for their child; yet, they often lack the essential, research-based information for making informed decisions (Beaulieu & Welsh, 2014). Parents and guardians will benefit by having information that helps them understand the costs and benefits of service placement decisions for their child. For parents and advocates the results provide additional encouragement for ensuring that their children are educated in inclusive setting with their regular education peers.

The results of this study suggest that employing a decision-making process that avoids a separate placement for students with disabilities is the best option. Importantly, the results present the opportunity for a paradigm shift in the operational relationships among decisions concerning the assessment of disability, placements for the delivery of services and the services themselves. If student placement in inclusive settings yields greater academic outcomes for all but those with the most intense and atypical needs, the challenge in the decision-making process is to ensure the right array of services that can become part of the teaching and learning architecture of the general education classroom. This approach to decision-making enables restructuring of the decision-making process into a framework of belief in the capabilities of teachers and students, and the ability of educational professionals to truly create educational opportunity for all students in classrooms with their general education peers. Assessment takes on a clearer purpose for creating insights for instructional modifications and accommodations in general education classrooms rather than in separate classroom settings.

This approach to decision-making in the placement and service delivery process also becomes the framework for increasing teaching capacity to address the needs of all children. Moreover, the current terminology of college and career readiness, currently dominating the educational dialogue shifts into a human capital model with an infinitely greater capacity for preparing all students for more productive and fulfilling life experiences as productive community members.

The potential value of this research is comprehensive in nature in that it could be of use for legislation, policy, systems design and practices that define the educational and life experiences for students with disabilities. In doing so, it can play a crucial role in resolving the tensions regarding expectations, resources, inconsistent decisions and best practices. Most importantly, providing evidence regarding the relationships between place and outcomes will allow for data-based research to guide and support decision-making for students with disabilities.

The educational policy context is fraught with tension as it brings together high expectations for the achievement of all students and limited resources and capacities. Separate systems are expensive and cumbersome (Burrello et. al, 2001). Inclusive education, from a policy perspective means taking a holistic approach to education reform and changing the way educational systems and policy tackles inclusion (UNESCO, 2009). This research sheds light for policies related to school improvement and school reform. The training of personnel could operate out of a clear theoretical framework concerning placement and service delivery.

Resource allocations and programmatic decisions could be based upon definitive evidence based practices concerning placements and service delivery.

Finally, there are important implications related to teacher and administrator training and licensing programs. Definitive research concerning the best placement location for special education services provides a clearer focus for both pre-service training as well as in-service professional development. Mastropieri and Scruggs (2005) noted that practicing teachers are ill-prepared to work in inclusive environments. There seem to be deep divides separating special and general educations with both practice and philosophy (Sindelar, Adams, & Leko, 2014). If the placement of students with disabilities significantly impacts their academic achievement, then the skills, dispositions and knowledge related to teaching students with disabilities must be considered. For school administrators, appropriate student placement for special education services requires an understanding of the relationship between placement and outcomes.

It is expected that the results of this study can be generalized to other states and that the sample for this study is in fact, representative of other students with disabilities across the United States. All states must comply to the same guidance and regulations outlined by the U.S. Department of Education related to the identification of students with disabilities, the development of an individual education plan that outlines placement on a continuum, identification of primary exceptionality code, and assessment (per ESSA) on the state measure per their state accountability system. Sadly, placement in general education classrooms often depends on where a student lives and what disability label the student holds (Giangreco, Dymond, & Shogren, 2016). Placement data trends show that placement varies by state; research related to the impact of inclusion on student outcomes to enable equity in access across states is important.

Webster defines inclusion as an act of taking in as part of a whole**:** the state of being taken in. When considered for students with disabilities, the word suggests that they were not a "part of the whole" -- they were once excluded and now the system is considering how to include. If being considered a part of the general education classroom results in better academic achievement, then policy and practice should have a greater focus on how to ensure that students with disabilities are not excluded, but rather, the resources necessary to keep them as "a part of the whole" are provided.

## Limitations

Though we believe our results suggest a strong advantage to special education students' placement in the general education classroom with respect to their performance on ELA and mathematics assessments, there are some important limitations to consider.

**Matching Quality**. While most Rosenbaum bounds were within the range (or better) of what is generally seen and accepted in social science and education research (Calindo and Kopeinig, 2005; DiPrete and Gangl, 2004), we should note that this does not imply that the sensitivity to unobserved variable bias is therefore negligible. Rather, these numbers imply that unobserved variables may have a strong impact on many of the ATT estimates found, and that other research showing similar bounds are also implicated. Similarly, in spite of the generally positive results associated with the KS-tests, not all of the 720 KS results were non-significant, suggesting that some matches may be less reliable than others. However, we also note that prior to matching, differences in third grade reading, ELA, and math assessments were on average more disparate than they were after matching, which may suggest that the numbers found in our analyses could actually be more conservative than they otherwise may have been.

**Data Quality / Primary Disability Changes**. Discrepancies in the data provided by the Indiana Department of Education may have unobserved influences on results. For example, it may be that students have different primary disabilities in grades 4 through 8 than they had in grade 3, or that their primary disability category shifts from year to year. Similarly, other effects such as school transfer, teacher, and classroom are not accounted for in these analyses, yet may plausibly influence students' performance on statewide ELA and math assessments. Further research using data with greater granularity is needed to assess the influence of these unobserved factors.

**Sample Sizes and Occasional Inconsistency**. The trend in the analyses consistently favors students who have been included. However, in some few analyses the results are counter to this trend. When this happens it could likely be the result of variations in sample size. This is discussed in the *treatment and control groups* section above, but we reiterate the limitation here with reference to Tables 4-7.

## Future Research

This study researched outcomes of students grades 3-8. An important next step is to study the post-secondary outcomes of students with disabilities who are included and to look at placement in high school and the impact on graduation.

Determining the impact of one setting vs. another could lead to research that is more comprehensive on the factors that influence positive school outcomes for students with disabilities. The results from this study will afford researchers an opportunity to delve deeper into questions related to placement: What are the practices that result in more positive student outcomes? What are the systems and policies that need to be discussed and developed? Should there be differences in licensing and training requirements and experiences?

Unanswered questions regarding the most effective policy, strategies and practices for providing services to students with disabilities define the on-going debate regarding the placement of students with disabilities. Identifying and implementing effective strategies with fidelity is as much of a challenge for addressing the needs of students with disabilities as it is for the many programs and strategies that define the education of students across the spectrum of student characteristics in public education. If one placement shows evidence of more positive student outcomes, it will be important to study which policies, practices and structures ensure effective implementation. Once the results of this and other studies are able to provide clarity regarding the relationships between the placement and service delivery model chosen for students with disabilities and school and post-secondary outcomes including employment, integrating the findings into the procedures of a coherent decision-making process and programs implemented with fidelity will be the next challenge.

Finally, this study did not consider variables such as free and reduced lunch and race/ethnicity. These matching variables could be included in a future study to determine additional information on demographic characteristics of students with disabilities.

## References

Bateman, B., Lloyd, J. W. & Tankersley, M. (2015). *Enduring issues in special education: Personal perspectives.* New York, N.Y.: Routledge

Becker, S.O., & Ichino, A. (2002). Estimation of average treatment effects based on propensity scores. Retrieved from http://sobecker.userweb.mwn.de/psj7.pdf

Burrello, L., Lashley, C. & Beatty, E. (2001). *Educating all students together: How school leaders create unified systems.* Thousand Oaks, CA: Corwin Press

Calindo, M. & Kopeinig, S. (2005, May) *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

Cole, C. M., Waldron, N., & Majd, M. (2004). Academic progress of students across inclusive and traditional settings. *Mental Retardation*, *42*, 136-144.

Conn-Powers, M., Crawford, S.R., Dixon, S. & Hall, M.L. (2017). *Indiana's efforts to provide inclusive special education services for preschoolers with disabilities*. Retrieved from https://www.iidc.indiana.edu/styles/iidc/defiles/ECC/Preschool%20Inclusion%20Report_Final.pdf .

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

Dyson, A., Farrell, P., Polat, F., & Hutcheson, G. (2004). *Inclusion and pupil achievement*. Research Report 578, University of Newcastle. Retrieved from https://webarchive.nationalarchives.gov.uk/20130402121316/https://www.education.gov.uk/publications/eOrderingDownload/RR578.pdf

Every Student Succeeds Act, S.1177, 114th Cong. (2015). Retrieved from http://edworkforce.house.gov/uploadedfiles/ every_student_succeeds_act_-_conference_report.pdf

Farrell, M. (2010). *Debating special education*. London: Routledge.

Gerrard, L. (2006). Inclusive education: An issue of social justice. *Equity and Excellence in Education. 27*(1), 58-67.

Giangreco, M.F., Dymond, S.K., & Shogren, K.A. (2016). Educating students with severe disabilities: Foundational concept and practices. In F. Brown, J. McDonnell, & M.E. Snell (Eds.), *Instruction of students with severe disabilities* (8th ed., pp 1-26). Upper Saddle River, NJ: Pearson Education/Prentice Hall.

Gilmour, A. (2018). Has inclusion gone too far? *Education Next. 18*(4).

Goodman, J., Bucholz, J., Hazelkorn, M., & Duffy, M. (2014). Using graduation rates pf students with disabilities as an indicator of successful inclusive education. *Measuring Inclusive Education: International Perspectives on Inclusive Education. 77*, 279-301.

Grossi, T. & Cole, C. (2008). Postschool outcomes of students with disabilities across inclusive and traditional settings. Unpublished manuscript. Indiana University, Bloomington, IN.

Guo, S. & Fraser, M. W. (2015). *Propensity score analysis: Statistical methods and applications* (2nd Ed.). Thousand Oaks, CA: Sage Publications.

Hehir, T. (2007). Confronting ableism. *Educational Leadership, 64*(5), 8-14.

Individuals with Disabilities Education Act (2004). Retrieved from https://sites.ed.gov/idea/

Joyce-Beaulieu, D. & Welsh, J. (2014). What are the roles of related service personnel in inclusive schools? In J. McLeskey, N. Waldron, F. Spooner, & B. Algozzine. *Handbook of effective inclusive schools: Research and practice.* New York, NY: Routledge

Kalambouka, A., Farrell, P., Dyson, P., & Kaplan, A. (2007). The impact of placing pupils with special educational needs in mainstream schools on the achievement of their peers. *Educational Research, 49,* 365-382.

Kauffman J.M. (1999).Commentary: Today's special education and its messages for tomorrow. *The Journal of Special Education***,** 32(4), 244-254.

Kauffman, J. M. & Badar, J. (2014) Better thinking and clearer communication will help special education. *Exceptionality: A Special Education Journal*, *22*(1), 17–32.

Kauffman, J. M. & Hallahan, D. P. (eds) (2005) *The illusion of full inclusion: A comprehensive critique of a current special education bandwagon* (2nd Ed.), Austin, TX: PRO-ED.

Kavale, K.A. & Forness, S.R. (2000). History, rhetoric, and reality: Analysis of the inclusion debate. *Remedial and Special Education 21*(5), 279-296.

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.

Lindsay, G. (2007). Annual review: Educational psychology and the effectiveness of inclusive education/mainstreaming. *British Journal of Educational Psychology*, *77*, 1-24.

Lombardi, A., Gau, J., Doren, B., & Lindstrom, L. (2013). The influence of instructional setting of common core subjects on postsecondary participation. *Journal of Disability Policy Studies*, *24,* 170-180.

Lupini, W.H., & Zirkel, P.A. (2003). An outcome analysis of education litigation. *Educational Policy, 17*(2), 257-279. doi.org/10.1177/0895904803017002004.

Manset, G., & Semmel, M. (1997). Are inclusive programs for students with mild disabilities effective? A comparative review of model programs. *Journal of Special Education, 31*(2), 155–181.

Mastropieri, M.A. & Scruggs, T.E. (2005). Feasibility and consequences of response to intervention: Examining the issues and scientific evidence as a model for the identification of individuals with learning disabilities. *Journal of Learning Disabilities*, *38*, 525-531.

McLeskey, J, Rosenberg, M., & Westling, D. (2013). *Inclusion: Effective practices for all students* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Morgan, P.L., Frisco, M.L., Farkas, G., & Hibel, J. (2008). A propensity score matching analysis of the effects of special education services. doi.org/10.1177/0022466908323007

Newcomer, J.R., & Zirkel, P.A. (1999). An analysis of judicial outcomes of special education cases. *Exceptional Children,* doi.org/10.1177/001440299906500403

Rogers, D. P., & Thiery, I. M. (2003, November). *Does an inclusive setting affect reading comprehension in students with learning disabilities?* Paper presented at the annual meeting of the Mid-South Educational Research Association.

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*, 210-219.

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

Runswick-Cole, K. (2011). Time to end the bias towards inclusive education? *British Journal of Special Education, 38*(3), 112-119.

Sindelar, P., Adams, A., & Leko, C. (2014). How can teacher education improve effective inclusive schools? In J. McLeskey, N. Waldron, F. Spooner, & B. Algozzine. *Handbook of effective inclusive schools: Research and practice.* New York, NY: Routledge

Slee, R. (2011) *The irregular school: Exclusion, schooling and inclusive education*. London: Routledge.

Theobald, R., Goldhaber, D., Gratz, T., & Holden, K. (2017). *Career and technical education, inclusion, and postsecondary outcomes for students with disabilities*. Working paper. National Center for Analysis of Longitudinal Data in Education Research.

UNESCO, (2009). *Policy guidelines on inclusion in education*. Published by the United Nations Educational, Scientific and Cultural Organization, 7, place de Fontenoy, 75352 Paris 07 SP, France

Villa, R., & Thousand, J. (1995). Inclusion: Alive and well in the Green Mountain State. *Kappan Phi Delta* , *77*(4), 288-297.

Villa, R., & Thousand, J. (2005). *Creating an inclusive school.* 2nd. ed.) Alexandria, VA: Association for Supervisions and Curriculum Development

Waldron, N., & McLesky, J. (1998). The effects of an inclusive school programs on students with mild and severe learning disabilities. *Exceptional Children, 64*(3), 395–505.

Zigmond, N. (2015). Where should students with disabilities receive their education? In Bateman, B., Wills-Lloyd, S., & Tankersley, M. (2015). Enduring issues in special education: Personal perspectives. New York: Routledge

Table 1. Analytical Scope for Inclusion Study

| Analysis Set 1 | | Analysis Set 2 | |
|---|---|---|---|
| Treatment: High Inclusion<br>Control: Mixed Inclusion | Ignoring Disability Codes<br>Matching on Disability Codes<br>Within Unique Disability Codes | Treatment: Mixed Inclusion<br>Control: High Inclusion | Ignoring Disability Codes<br>Matching on Disability Codes<br>Within Unique Disability Codes |
| Treatment: High Inclusion<br>Control: Low Inclusion | Ignoring Disability Codes<br>Matching on Disability Codes<br>Within Unique Disability Codes | Treatment: Low Inclusion<br>Control: High Inclusion | Ignoring Disability Codes<br>Matching on Disability Codes<br>Within Unique Disability Codes |

*Unique disability codes include emotional disability, specific learning disability, mild cognitive disability, and autism spectrum disorder.
**In analyses that ignore or match on disability codes, all disabilities other than speech or language impairments are used.

Table 2. Results for Analyses Not Controlling for Student Disability

| | Analysis Set 1 | | | | Analysis Set 2 | | | |
|---|---|---|---|---|---|---|---|---|
| | | | ATT | p-value | | | ATT | p-value |
| ELA | Treatment: High-Inclusion<br>Control: Mixed-Inclusion | 2014 | 5.31 | 0.005 | Treatment: Mixed-Inclusion<br>Control: High-Inclusion | 2014 | -8.35 | < 0.001 |
| | | 2015 | 8.34 | < 0.001 | | 2015 | -7.44 | < 0.001 |
| | | 2016 | 12.01 | < 0.001 | | 2016 | -11.14 | < 0.001 |
| | | 2017 | 13.57 | < 0.001 | | 2017 | -12.88 | < 0.001 |
| | | 2018 | 26.84 | < 0.001 | | 2018 | -23.16 | < 0.001 |
| | | N | 3408 | | | N | 1874 | |
| Math | Treatment: High-Inclusion<br>Control: Mixed-Inclusion | 2014 | 7.53 | 0.003 | Treatment: Mixed-Inclusion<br>Control: High-Inclusion | 2014 | -7.28 | 0.002 |
| | | 2015 | 5.95 | 0.005 | | 2015 | -6.29 | 0.002 |
| | | 2016 | 8.69 | < 0.001 | | 2016 | -8.16 | < 0.001 |
| | | 2017 | 10.3 | < 0.001 | | 2017 | -7.06 | < 0.001 |
| | | 2018 | 16.96 | < 0.001 | | 2018 | -13.42 | < 0.001 |
| | | N | 3506 | | | N | 2014 | |
| ELA | Treatment: High-Inclusion<br>Control: Low-Inclusion | 2014 | 16.35 | 0.018 | Treatment: Low-Inclusion<br>Control: High-Inclusion | 2014 | -22.92 | < 0.001 |
| | | 2015 | 17.26 | 0.025 | | 2015 | -15.69 | < 0.001 |
| | | 2016 | 19.68 | 0.015 | | 2016 | -20.34 | < 0.001 |
| | | 2017 | 26.24 | 0.009 | | 2017 | -34.77 | < 0.001 |
| | | 2018 | 24.13 | 0.052 | | 2018 | -19.25 | 0.006 |
| | | N | 3388 | | | N | 146 | |
| Math | Treatment: High-Inclusion<br>Control: Low-Inclusion | 2014 | 3.66 | 0.69 | Treatment: Low-Inclusion<br>Control: High-Inclusion | 2014 | -21.79 | < 0.001 |
| | | 2015 | 8.01 | 0.28 | | 2015 | -56.59 | < 0.001 |
| | | 2016 | 22.77 | < 0.001 | | 2016 | -24.98 | < 0.001 |
| | | 2017 | 28.89 | < 0.001 | | 2017 | -23.84 | < 0.001 |
| | | 2018 | 32.69 | 0.003 | | 2018 | -21.91 | 0.008 |
| | | N | 3588 | | | N | 168 | |

*Sample sizes reflect that of the entire match. With one-to-one sampling, sample sizes in either the treatment or control group is half that of the total.

**In analysis set 2, negative values reflect higher scores in the high-inclusion setting (treatment - control).

Table 3. Results for Analyses Matching on Student Disability

| | Analysis Set 1 | | | | Analysis Set 2 | | | |
|---|---|---|---|---|---|---|---|---|
| | | | ATT | p-value | | | ATT | p-value |
| ELA | Treatment: High-Inclusion<br>Control: Mixed-Inclusion | 2014 | 6.19 | 0.005 | Treatment: Mixed-Inclusion<br>Control: High-Inclusion | 2014 | -8.51 | < 0.001 |
| | | 2015 | 7.94 | < 0.001 | | 2015 | -7.89 | < 0.001 |
| | | 2016 | 11.57 | < 0.001 | | 2016 | -9.32 | < 0.001 |
| | | 2017 | 11.4 | < 0.001 | | 2017 | -11.3 | < 0.001 |
| | | 2018 | 20.89 | < 0.001 | | 2018 | -22.27 | < 0.001 |
| | | N | 3390 | | | N | 1882 | |
| Math | Treatment: High-Inclusion<br>Control: Mixed-Inclusion | 2014 | 8.55 | 0.002 | Treatment: Mixed-Inclusion<br>Control: High-Inclusion | 2014 | -9.27 | 0.001 |
| | | 2015 | 7.96 | < 0.001 | | 2015 | -6.59 | 0.005 |
| | | 2016 | 8.34 | < 0.001 | | 2016 | -8.85 | < 0.001 |
| | | 2017 | 8.96 | < 0.001 | | 2017 | -6.42 | 0.003 |
| | | 2018 | 19.54 | < 0.001 | | 2018 | -14.48 | < 0.001 |
| | | N | 3496 | | | N | 2018 | |
| ELA | Treatment: High-Inclusion<br>Control: Low-Inclusion | 2014 | 34.65 | 0.017 | Treatment: Low-Inclusion<br>Control: High-Inclusion | 2014 | -20.46 | < 0.001 |
| | | 2015 | 29 | 0.047 | | 2015 | -11.11 | 0.04 |
| | | 2016 | 53.21 | 0.003 | | 2016 | -12.51 | 0.03 |
| | | 2017 | 45.38 | 0.006 | | 2017 | -24.43 | < 0.001 |
| | | 2018 | 29.92 | 0.17 | | 2018 | -17.88 | 0.02 |
| | | N | 3386 | | | N | 144 | |
| Math | Treatment: High-Inclusion<br>Control: Low-Inclusion | 2014 | 37.31 | 0.035 | Treatment: Low-Inclusion<br>Control: High-Inclusion | 2014 | -23.57 | 0.002 |
| | | 2015 | 16.99 | 0.13 | | 2015 | -26.24 | < 0.001 |
| | | 2016 | 32.5 | 0.005 | | 2016 | -23.55 | < 0.001 |
| | | 2017 | 46.28 | < 0.001 | | 2017 | -25.51 | < 0.001 |
| | | 2018 | 37.94 | 0.005 | | 2018 | -31.7 | < 0.001 |
| | | N | 3490 | | | N | 174 | |

*Sample sizes reflect that of the entire match. With one-to-one sampling, sample sizes in either the treatment or control group is half that of the total.
**In analysis set 2, negative values reflect higher scores in the high-inclusion setting (treatment - control).

Table 4. Results for Analyses of Students with Emotional Disabilities

| | Analysis Set 1 | | | | Analysis Set 2 | | | |
|---|---|---|---|---|---|---|---|---|
| | | | ATT | p-value | | | ATT | p-value |
| ELA | Treatment: High-Inclusion; Control: Mixed-Inclusion | 2014 | 4.46 | 0.53 | Treatment: Mixed-Inclusion; Control: High-Inclusion | 2014 | -9.15 | 0.25 |
| | | 2015 | 8.15 | 0.21 | | 2015 | -7.18 | 0.29 |
| | | 2016 | 1.08 | 0.89 | | 2016 | -10.2 | 0.22 |
| | | 2017 | 11.8 | 0.17 | | 2017 | -16 | 0.052 |
| | | 2018 | 7.29 | 0.59 | | 2018 | -39.26 | 0.02 |
| | | N | 158 | | | N | 184 | |
| Math | Treatment: High-Inclusion; Control: Mixed-Inclusion | 2014 | -1.36 | 0.89 | Treatment: Mixed-Inclusion; Control: High-Inclusion | 2014 | -1.61 | 0.87 |
| | | 2015 | 3.43 | 0.68 | | 2015 | -4.39 | 0.6 |
| | | 2016 | 6.97 | 0.39 | | 2016 | -5.39 | 0.48 |
| | | 2017 | 2.81 | 0.68 | | 2017 | -6.16 | 0.4 |
| | | 2018 | 10.07 | 0.29 | | 2018 | -14.69 | 0.19 |
| | | N | 160 | | | N | 204 | |
| ELA | Treatment: High-Inclusion; Control: Low-Inclusion | 2014 | 17.62 | 0.18 | Treatment: Low-Inclusion; Control: High-Inclusion | 2014 | -19.57 | 0.23 |
| | | 2015 | 3.75 | 0.73 | | 2015 | -33.57 | 0.001 |
| | | 2016 | 7.58 | 0.51 | | 2016 | -12.71 | 0.39 |
| | | 2017 | 12.21 | 0.2 | | 2017 | -35.93 | 0.006 |
| | | 2018 | -15.79 | 0.33 | | 2018 | -36.64 | 0.025 |
| | | N | 104 | | | N | 28 | |
| Math | Treatment: High-Inclusion; Control: Low-Inclusion | 2014 | -15.77 | 0.31 | Treatment: Low-Inclusion; Control: High-Inclusion | 2014 | -16.64 | 0.37 |
| | | 2015 | -7.95 | 0.49 | | 2015 | -20.33 | 0.1 |
| | | 2016 | 16.3 | 0.057 | | 2016 | -30 | 0.023 |
| | | 2017 | 23.53 | 0.014 | | 2017 | -14.67 | 0.22 |
| | | 2018 | 53.86 | 0.002 | | 2018 | -57.11 | 0.01 |
| | | N | 114 | | | N | 18 | |

*Sample sizes reflect that of the entire match. With one-to-one sampling, sample sizes in either the treatment or control group is half that of the total.

**In analysis set 2, negative values reflect higher scores in the high-inclusion setting (treatment - control).

Table 5. Results for Analyses of Students with Specific Learning Disabilities

| | **Analysis Set 1** | | | | **Analysis Set 2** | | | |
|---|---|---|---|---|---|---|---|---|
| | | | ATT | p-value | | | ATT | p-value |
| | | 2014 | 4.32 | 0.08 | | 2014 | -8.94 | < 0.001 |
| | Treatment: High-Inclusion | 2015 | 6.09 | 0.015 | Treatment: Mixed-Inclusion | 2015 | -8.99 | < 0.001 |
| ELA | | 2016 | 8.27 | 0.005 | | 2016 | -10.26 | < 0.001 |
| | Control: Mixed-Inclusion | 2017 | 10 | < 0.001 | Control: High-Inclusion | 2017 | -12.12 | < 0.001 |
| | | 2018 | 17.19 | < 0.001 | | 2018 | -16.66 | < 0.001 |
| | | N | 1834 | | | N | 1022 | |
| | | 2014 | 7.29 | 0.02 | | 2014 | -9.52 | 0.002 |
| | Treatment: High-Inclusion | 2015 | 5.95 | 0.03 | Treatment: Mixed-Inclusion | 2015 | -7.58 | 0.002 |
| Math | | 2016 | 7.23 | 0.008 | | 2016 | -5.04 | 0.04 |
| | Control: Mixed-Inclusion | 2017 | 8.78 | < 0.001 | Control: High-Inclusion | 2017 | -5.33 | 0.017 |
| | | 2018 | 11.53 | 0.009 | | 2018 | -8.33 | 0.056 |
| | | N | 1908 | | | N | 1112 | |
| | | 2014 | 15.52 | 0.068 | | 2014 | -25.08 | < 0.001 |
| | Treatment: High-Inclusion | 2015 | 37.31 | 0.016 | Treatment: Low-Inclusion | 2015 | -9.17 | 0.27 |
| ELA | | 2016 | 29.63 | 0.07 | | 2016 | -20.61 | 0.011 |
| | Control: Low-Inclusion | 2017 | 43.89 | 0.006 | Control: High-Inclusion | 2017 | -25.87 | 0.005 |
| | | 2018 | 60.81 | 0.004 | | 2018 | -26.6 | 0.01 |
| | | N | 1764 | | | N | 52 | |
| | | 2014 | 11.54 | 0.36 | | 2014 | -20.49 | 0.009 |
| | Treatment: High-Inclusion | 2015 | 23.31 | 0.041 | Treatment: Low-Inclusion | 2015 | -23.96 | < 0.001 |
| Math | | 2016 | 15.7 | 0.11 | | 2016 | -12.95 | 0.032 |
| | Control: Low-Inclusion | 2017 | 19.29 | 0.024 | Control: High-Inclusion | 2017 | -13.55 | 0.002 |
| | | 2018 | 21.37 | 0.085 | | 2018 | -14.97 | 0.018 |
| | | N | 1824 | | | N | 66 | |

*Sample sizes reflect that of the entire match. With one-to-one sampling, sample sizes in either the treatment or control group is half that of the total.

**In analysis set 2, negative values reflect higher scores in the high-inclusion setting (treatment - control).

Table 6. Results for Analyses of Students with Mild Cognitive Disabilities

| | Analysis Set 1 | | | | Analysis Set 2 | | | |
|---|---|---|---|---|---|---|---|---|
| | | | ATT | p-value | | | ATT | p-value |
| ELA | Treatment: High-Inclusion | 2014 | 4.61 | 0.64 | Treatment: Mixed-Inclusion | 2014 | 3.72 | 0.64 |
| | | 2015 | 12.39 | 0.22 | | 2015 | -5.21 | 0.52 |
| | | 2016 | 8.78 | 0.52 | | 2016 | -6.34 | 0.54 |
| | Control: Mixed-Inclusion | 2017 | 22.72 | 0.13 | Control: High-Inclusion | 2017 | -17.45 | 0.11 |
| | | 2018 | 36.22 | 0.002 | | 2018 | -12.31 | 0.21 |
| | | N | 36 | | | N | 58 | |
| Math | Treatment: High-Inclusion | 2014 | 1.62 | 0.89 | Treatment: Mixed-Inclusion | 2014 | -2.5 | 0.78 |
| | | 2015 | 11.54 | 0.14 | | 2015 | -22 | 0.008 |
| | | 2016 | 15.23 | 0.12 | | 2016 | -20.24 | 0.018 |
| | Control: Mixed-Inclusion | 2017 | 25.19 | 0.004 | Control: High-Inclusion | 2017 | -30 | < 0.001 |
| | | 2018 | 31.27 | 0.003 | | 2018 | -36.37 | < 0.001 |
| | | N | 52 | | | N | 76 | |
| ELA | Treatment: High-Inclusion | 2014 | 12.38 | 0.29 | Treatment: Low-Inclusion | 2014 | -13.75 | 0.26 |
| | | 2015 | 17 | 0.18 | | 2015 | -13.75 | 0.22 |
| | | 2016 | 18.81 | 0.12 | | 2016 | -16.13 | 0.11 |
| | Control: Low-Inclusion | 2017 | 47.25 | < 0.001 | Control: High-Inclusion | 2017 | -57.58 | < 0.001 |
| | | 2018 | 52 | < 0.001 | | 2018 | -37.13 | 0.003 |
| | | N | 32 | | | N | 16 | |
| Math | Treatment: High-Inclusion | 2014 | 24.11 | 0.096 | Treatment: Low-Inclusion | 2014 | -33.13 | 0.017 |
| | | 2015 | 33.11 | 0.066 | | 2015 | -41.75 | 0.002 |
| | | 2016 | 10.44 | 0.33 | | 2016 | -13.06 | 0.056 |
| | Control: Low-Inclusion | 2017 | 29.33 | 0.008 | Control: High-Inclusion | 2017 | -23.19 | 0.002 |
| | | 2018 | 61.17 | < 0.001 | | 2018 | -48.94 | < 0.001 |
| | | N | 36 | | | N | 16 | |

*Sample sizes reflect that of the entire match. With one-to-one sampling, sample sizes in either the treatment or control group is half that of the total.

**In analysis set 2, negative values reflect higher scores in the high-inclusion setting (treatment - control).

Table 7. Results for Analyses of Students with Autism Spectrum Disorder

| | **Analysis Set 1** | | | | **Analysis Set 2** | | | |
|---|---|---|---|---|---|---|---|---|
| | | | ATT | p-value | | | ATT | p-value |
| ELA | Treatment: High-Inclusion; Control: Mixed-Inclusion | 2014 | 6.03 | 0.24 | Treatment: Mixed-Inclusion; Control: High-Inclusion | 2014 | -8.55 | 0.13 |
| | | 2015 | 14.95 | 0.008 | | 2015 | -16.7 | 0.002 |
| | | 2016 | 21.83 | < 0.001 | | 2016 | -21.52 | < 0.001 |
| | | 2017 | 19.75 | 0.005 | | 2017 | -20.1 | 0.005 |
| | | 2018 | 20.5 | 0.029 | | 2018 | -26.67 | 0.01 |
| | | N | 414 | | | N | 142 | |
| Math | Treatment: High-Inclusion; Control: Mixed-Inclusion | 2014 | 0.11 | 0.99 | Treatment: Mixed-Inclusion; Control: High-Inclusion | 2014 | -16.24 | 0.014 |
| | | 2015 | 6.09 | 0.31 | | 2015 | -18.37 | < 0.001 |
| | | 2016 | 8.77 | 0.16 | | 2016 | -18.32 | 0.001 |
| | | 2017 | 8.69 | 0.17 | | 2017 | -15.15 | 0.006 |
| | | 2018 | 10.74 | 0.32 | | 2018 | -21.9 | 0.015 |
| | | N | 430 | | | N | 164 | |
| ELA | Treatment: High-Inclusion; Control: Low-Inclusion | 2014 | 38.58 | < 0.001 | Treatment: Low-Inclusion; Control: High-Inclusion | 2014 | -25 | 0.07 |
| | | 2015 | 9.53 | 0.07 | | 2015 | -9.25 | 0.6 |
| | | 2016 | 51.37 | < 0.001 | | 2016 | 1 | 0.94 |
| | | 2017 | 6.1 | 0.46 | | 2017 | 14.75 | 0.46 |
| | | 2018 | -29.17 | < 0.001 | | 2018 | 48 | < 0.001 |
| | | N | 118 | | | N | 8 | |
| Math | Treatment: High-Inclusion; Control: Low-Inclusion | 2014 | 63.21 | 0.045 | Treatment: Low-Inclusion; Control: High-Inclusion | 2014 | -12.36 | 0.007 |
| | | 2015 | 19.08 | 0.38 | | 2015 | -34.93 | < 0.001 |
| | | 2016 | 44.69 | 0.03 | | 2016 | -56.57 | < 0.001 |
| | | 2017 | 154.57 | < 0.001 | | 2017 | -70.54 | < 0.001 |
| | | 2018 | 76 | 0.02 | | 2018 | -66.75 | < 0.001 |
| | | N | 388 | | | N | 8 | |

*Sample sizes reflect that of the entire match. With one-to-one sampling, sample sizes in either the treatment or control group is half that of the total.

**In analysis set 2, negative values reflect higher scores in the high-inclusion setting (treatment - control).