Evaluating the "Least Restrictive Environment" Mandate: The Effects of Mainstreaming on Cognitive and Noncognitive Development

Andrew McGee

Department of Economics

Simon Fraser University

First draft March 2009; revised October 2010

JEL Codes: I21, J14

Keywords: disabilities, mainstreaming

Corresponding author: Andrew McGee, Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, B.C. V5A 1S6.

Email: andrew_mcgee@sfu.ca Phone: (778) 782-4777. Fax: (778) 782-5944.

Abstract: I estimate the marginal effects of mainstreaming, the practice of integrating disabled students into general education classes, on the development of cognitive and noncognitive skills for students in six disability categories controlling for unobserved heterogeneity in ability and past inputs into the production processes for skills. I find that a ten percentage point increase in the percent of the day spent in general education is associated with predicted score increases ranging from 11 to 22 percent of annual math score gains and 5 to 18 percent of annual verbal score gains for students with different disabilities, as well as significant improvements in social skills for nearly all categories of disabled students. By contrast, mainstreaming has little or no effect on disabled students' self-images.

INTRODUCTION

Passage of the Individuals with Disabilities Education Act (IDEA) in 1975 required public school districts to provide a free and appropriate education for all students with disabilities in the "least restrictive" environment possible, leading to the widespread adoption of mainstreaming, the practice of educating students with disabilities in general education classes alongside students without disabilities during specific time periods based on their skills.[1] In 2004, 78% of students between the ages of 6 and 21 receiving services under IDEA spent more than 40% of their day in general education classes, and 52% spent 80% or more of their day in general education classes (U.S. Department of Education, 2009).

Despite its prevalence, whether mainstreaming is in students' best interests remains the subject of debate. Proponents argue that students with disabilities learn more in general education from the more rigorous coursework and that special education classes slow the development of some students by teaching to the ability level of the most disabled in the class. Beyond academic performance, proponents argue that mainstreaming improves disabled students' social skills through positive peer modeling and leads to greater acceptance of the disabled by the non-disabled. Opponents of mainstreaming—including many parents of disabled children—argue that disabled students benefit from the smaller class sizes, specially trained teachers, and customized curricula characteristic of special education classes while avoiding academic and social frustration in general education classes that may harm their social skills and

---

[1] Originally designated the Education for All Handicapped Children Act (EAHCA), IDEA (Public Law 94-142) was renamed and amended in 1997 and 2004. By 2004, 9% of the U.S. population between the ages of 16 and 21 received services under IDEA (Department of Education, 2009). See Bursztyn (2006) for a history of legislation concerning special education.

self-images. Implementing the “least restrictive environment” mandate has proven challenging for school districts, whose policies and practices have been challenged from both sides of the mainstreaming debate.[2] Recognizing the persistent differences in opinions regarding mainstreaming and the difficulties associated with enforcing IDEA’s mandate, the legal profession has also revisited the merits of the “integration presumption.”[3]

The disagreements among educators, parents, and legal scholars reflect a persistent lack of consensus within the education literature concerning mainstreaming’s effects.[4] The widely

---

[2] School districts have been sued by parents for failing to sufficiently integrate their child into general education as in *Grim v. Rhinebeck Central School District, 2d. Cir. 2003,* and *Jennifer D. v. New York City Department of Education, 06 Civ. 15489* and by parents feeling that school districts use the “least restrictive environment” mandate as cover to provide fewer special services to disabled children as in Greece, NY (Hechinger , 2007).

[3] Arguing that the evidence does not “justify a presumption for a fully inclusive educational environment for all children with disabilities,” Colker (2006) suggests that modifying the “integration presumption” will “better serve the substantive goal of according an adequate and appropriate education to the full range of children who have disabilities while still protecting disabled children from inhumane, disability-only educational warehouses.” Arguing against a major departure from the “integration presumption,” Weber (2007) counters that “educational research supports integration that is done properly”, although he does allow that the integration presumption “should not be applied in a simple-minded way to say that general education is always best under all circumstances.”

[4] For example, Hunt (2000) notes that “Virtually all studies reviewed… document the benefits of inclusive educational programs and practices for students with and without disabilities and their

different findings result from common data limitations in this area of research. First, most studies involve non-random assignment to mainstreaming intensity and fail to deal with the concomitant selection issues. Second, most studies analyze the effects of particular programs for integrating disabled students into general education classrooms, meaning that their findings cannot be generalized (e.g., Zigmond et al., 1995; Marston, 1996; Manset & Semmel, 1997). Furthermore, studies avoiding the first pitfall rarely avoid the second. Experimental studies involving random assignment to different intensities of mainstreaming have shown positive effects of mainstreaming on cognitive development (e.g., Calhoun & Elliott, 1977; Waldron & Mcleskey, 1998), but these estimated effects reflect the effects of particular programs and pedagogical approaches to integrating disabled students into general education.[5] Third, the studies often

---

families." By contrast, Carlberg and Kravale (1980) in their literature review note that "the most vocal advocates of mainstreaming have built their arguments on a philosophical rather than empirical foundation" and that "the present trend towards mainstreaming by regular class placement may not be appropriate for certain children. Special class placement was not uniformly detrimental, but appears to show differential effects related to category of exceptionality."

[5] In the experiment described by Calhoun and Elliott (1977), students' placements (special or general education classes) were randomly assigned, but all teachers in the general education classes were also certified special education teachers. Likewise, Waldron and Mcleskey (1998) report the results of a program for integrating disabled students into general education in three schools developed and supported by a local university. Adoption of the program in three other schools in the district was delayed to provide a control group. See Hocutt (1996) for a summary and criticisms of existing studies of the effects of mainstreaming.

employ small samples and focus on the effects of mainstreaming on a particular disability, whereas IDEA's mandate applies to *all* categories of disabled students. Marston (1987), for instance, examines the effects of inclusion in general education classrooms on students with learning disabilities by following the same students from week to week using value-added specifications to identify the effects of changes in treatments (general education classes or special education classes), but the study follows only 11 students. Dealing explicitly with selection issues, on the other hand, Hanushek et al. (2002) find that mainstreaming has little or no effect on the development of math skills among special education students in Texas public schools, but they have very limited information concerning mainstreaming and the students' school programs and examine the effects of mainstreaming on only one domain.[6]

I estimate value-added skill production functions to identify the marginal effects of the percent of the day spent in general education—mainstreaming intensity—on math and verbal test scores and on self-images and social skills. Because the physiological and neurological impairments associated with different disabilities have different effects on the cognitive processes associated with learning and on the way in which disabled students interact with their peers and the world around them, I estimate different skill production technologies for students in six disability categories: learning disabilities, emotional disturbances, speech, orthopedic, and other health impairments, severe disabilities, autism, and hearing and visual impairments.

I estimate the production functions controlling for school and family inputs as well as mainstreaming intensity using information concerning students, their disabilities, families, and

[6] Hanushek et al. control for selection and individual heterogeneity using an array of fixed effects, but they only know whether a student spent all day in general education. They do not report results for verbal skills, though they indicate that the results are similar.

school programs collected by the Special Education Elementary Longitudinal Study (SEELS), a national sample which followed children with disabilities between the ages of 6 and 12 for four years. Even in this rich data set, however, not all past and contemporaneous inputs are observed. Because students with higher skill levels spend more time in general education, mainstreaming intensity will be correlated with past inputs and endowments of skill. Likewise, mainstreaming will be correlated with contemporaneous inputs if school and family input decisions are made in tandem with the mainstreaming decision. To account for mainstreaming's correlation with unobserved inputs and endowments, I estimate value-added production functions including lagged skill measures to proxy for unobserved past inputs and endowments, family income and other background characteristics to proxy for unobserved contemporaneous inputs, and school district fixed effects to account for unobserved school district inputs.

The paper is the first to use the production function approach to identify the relationship between mainstreaming and noncognitive skills and only the second, following Hanushek et al. (2002), to use the approach to identify the relationship between mainstreaming and cognitive skills. The value-added production functions and the related robustness checks deal with the selection issues involved in studies of mainstreaming, and the use of a large longitudinal study with students from different schools makes the findings somewhat more general than studies focusing on particular mainstreaming programs. [7] Additionally, while the effect of the percent of the day spent in general education may seem less important than the effects of particular aspects of a student's school program (e.g., where and with what support they receive math instruction), the percent of the day spent in general education features prominently in litigation involving

[7] Blackorby et al. (2007) use SEELS data to examine outcomes for students with different levels of participation in general education, but they do not control for heterogeneity in ability.

IDEA.[8] Finally, the paper exploits the data on mainstreaming intensity to identify non-linear effects of mainstreaming unremarked upon by previous studies.

I find that spending more time in general education has large predicted benefits for cognitive development. A ten percentage point increase in the percent of the day spent in general education—a mere 48 minutes in an eight-hour school day—is associated with an increase in math skill of as much as 11.2 percent of the average year-to-year gain in math skill for students with sensory impairments and up to 22.3 percent of the average year-to-year gain for students with severe disabilities. A ten percentage point increase in the percent of the day spent in general education is associated with a predicted increase in verbal skill of 5.0 percent of the average year-to-year gain for students with emotional disturbances and up to 17.9 percent of the average year-to-year gain for students with learning disabilities. Mainstreaming has similarly positive and statistically significant effects on the development of social skills for all but students with learning disabilities. By contrast, I find no statistically or economically significant relationship between mainstreaming and disabled students' self-images.

The finding that mainstreaming is not negatively related to development in any dimension, however, comes with a caveat. In specifications allowing the relationship between mainstreaming and skill development to be nonlinear, I find that for some skills and for some disability categories mainstreaming is negatively related to development at either very high or

[8] For instance, in *P. v. Newington Board of Education, 512 F. Supp. 2d 89, 2007 U.S Dist.*, lawyers for a disabled student argued that the school district failed to meet the "least restrictive environment" standard because the student's school program called for him to spend 74% of his school day in general education—less than the 80% of the day they argued that the law required. The courts did not agree and granted summary judgment to the school district.

very low intensities. In these instances, special education classrooms have an important role to play and ought not be too quickly eliminated in any push to meet the "least restrictive environment" mandate.

On the whole, however, mainstreaming appears to be associated with significant benefits the cognitive and social development of most disabled youth while having no deleterious effects on their self-images. On average, the cognitive benefits from exposure to rigorous coursework in general education classrooms appear to outweigh the loss of personalized attention in special education classes, and interaction with non-disabled peers appears to aid the social development of the disabled. Furthermore, any social ostracism and academic frustration experienced by disabled students in general education do not significantly harm disabled students' in ways measured in the data. For most disabled students, educating them in the "least restrictive environment" possible is in their best interests. The differences among disability categories in the marginal products of mainstreaming and the evidence of some non-linear mainstreaming effects, however, highlight the problems with a "one-size-fits-all" approach and the need for mainstreaming policies that recognize differences in skill production technologies and the consequences these differences have for development.

SKILL PRODUCTION FUNCTIONS

I estimate production functions for math skill, verbal skill, social skills, and self-images for each of six categories of disabled children while dealing with the problems of unobserved inputs and the selection of more able students into general education. Consider a production function for a student age *a* who entered school *k* periods ago

$$m_t = \sum_{j=1}^{k} MS_{t-j}\alpha_{t-j} + \sum_{j=1}^{k} S_{t-j}\Theta_{t-j} + \sum_{j=1}^{a} F_{t-j}\beta_{t-j} + \sum_{j=1}^{a} M_{t-j}^{-1}\Gamma_{t-j} + \mu + \varepsilon_t \qquad (1)$$

where m_t denotes the measure of skill in period t, MS_{t-j} is a measure of mainstreaming intensity in period t-j, S_{t-j} is a vector of school inputs in period t-j, F_{t-j} is a vector of family investments in period t-j, M_{t-j}^{-1} is a vector of skill measures other than m in period t-j, μ is the student's innate endowment of skill m, and ε_t represents shocks to production in period t.

This production function assumes that skill is a linear function of the full history of school and family inputs as well as innate endowments of skill and shocks to production and allows for the possibility of complementarities in production between skills. For example, youth with better social skills may find classrooms easier environments to navigate and be more able to focus on learning (Segal, 2008). While Cunha and Heckman (2008) do not find that cognitive skill fosters the development of noncognitive skill, brighter students may develop more favorable self-images as a result of positive experiences in school, and more intelligent youth may be better able to discern and conform to social norms.

The goal of the analysis is to estimate the marginal products of mainstreaming intensity, α_{t-j}. Unfortunately, no data set contains complete histories of mainstreaming intensity, school inputs, family inputs, acquired skills and endowments. Because more able students are more intensively mainstreamed, mainstreaming intensity is correlated with endowments of skill as well as school and family inputs in earlier periods. To proxy for unobserved past school and family inputs and endowments, I adopt the so-called value-added approach by including a lagged measure of skill, m_{t-1}, as a regressor along with inputs applied between skill measures, which I refer to as contemporaneous inputs and denote with the subscript t-1.

Some contemporaneous family inputs such as the time spent by parents assisting disabled children with general education coursework are determined in conjunction with mainstreaming intensity. The inclusion of a lagged skill measure does not mitigate the correlation between contemporaneous mainstreaming intensity and these unobserved contemporaneous inputs. To proxy for unobserved contemporaneous inputs, I include as regressors characteristics of students and their families, C_{t-1}, and family income, Y_{t-1}, in a so-called *hybrid* value-added production function (Todd & Wolpin, 2003).

Some contemporaneous school inputs correlated with mainstreaming intensity such as the attention paid by teachers to disabled students in general education and the availability of assistance for general education teachers with disabled students in their classrooms are also unobserved. If school districts with fewer resources or placing less emphasis on educating the disabled mainstream students more intensively as a cost-saving measure, then the estimated mainstreaming effect without accounting for unobserved school district factors will be smaller than the causal effect of mainstreaming. Alternatively, if school districts with effective curriculum and more effective assistance for disabled students mainstream students more intensively, then the estimated mainstreaming effect without accounting for unobserved school district factors will be greater than the causal effect of mainstreaming. To account for unobserved school inputs, I exploit the clustering of students in school districts by including school district fixed effects, which I denote by D_l.

As a final departure from (1), I include in each production function a time trend Δt_t to account for the fact that the data contain observations in which m_t and m_{t-1} are measured one, two, and three years apart. Because skills are not measured every year, the marginal effects cannot be interpreted as the effects of annual flows of school and family inputs into the skill production processes. Instead, the marginal effects of mainstreaming intensity and other inputs reflect deviations from disability-specific developmental trends caused by different levels of inputs.[9] I estimate production functions of the form

$$m_t = \alpha MS_{t-1} + S_{t-1}\Theta + F_{t-1}\beta + M_{t-1}^{-1}\Gamma + \delta m_{t-1} + \rho Y_{t-1} + C_{t-1}\Psi + \sum_{l=1} D_l + \lambda \Delta t_t + \nu_t \qquad (2)$$

where v_t is the error term. The production functions are estimated using ordinary least squares, and the standard errors of the estimates are corrected for the non-independence of observations at the school district level.

Identification of the causal effect of mainstreaming intensity on skill development requires that the lagged skill measure is a sufficient statistic for unobserved ability and prior inputs and that family income and background characteristics are effective proxies for unobserved contemporaneous inputs.[10] While the lagged measures of other skills are explicitly

---

[9] Interpreting these mainstreaming effects would be problematic if the effects of mainstreaming vary depending on the length of exposure to a given level of mainstreaming intensity. To examine this issue, I estimated the production functions with interactions between dummies for the elapsed time between measures and mainstreaming intensity and tested for equality of the coefficients of these interactions. The data overwhelming fail to reject the equality of the coefficients, suggesting that the effect of mainstreaming intensity is constant over different periods of exposure.

[10] As discussed in Boardman and Murnane (1979) and Todd and Wolpin (2003), the lagged skill measure will be a sufficient statistic if the effects of lagged inputs depend only on the age at which they are applied as assumed implicitly in (1) or if these effects diminish at a constant, geometric rate over time. For a hybrid value-added production function to identify the causal effect of a contemporaneous input like mainstreaming intensity when some contemporaneous inputs are unobserved, it must be the case that the demand functions for inputs to the production

included to allow for complementarities between skills in production, including them undoubtedly also reduces the correlation between contemporaneous mainstreaming intensity and unobserved past inputs and ability. Thus even if the conditions for the lagged skill measure itself to be a sufficient statistic for past inputs are not satisfied, the vector of five lagged skill measures should be an adequate sufficient statistic for past inputs and endowments.

The hybrid value-added production functions control for individual heterogeneity in skill levels. If, however, school districts make mainstreaming decisions based on learning trajectories or skill growth rates unobserved by the econometrician, then the effects of mainstreaming identified will not be causal effects. I conduct two robustness checks to establish the importance of selection based on growth rates. First, I estimate the value-added production functions with two lags of the dependent variable included as regressors. If schools make placement decisions based on skill levels and growth rates, then including two lags is effectively the "reduced-form" of a specification including both the skill level and growth rate in the previous period.[11] Second, I instrument for mainstreaming intensity by creating a measure of a school district's propensity

---

processes are linear functions of prices, characteristics that affect familial preferences regarding development, and income, all of which are included as proxies for unobserved contemporaneous inputs (Todd & Wolpin, 2007).

[11] Alternatively, one could estimate the value-added production functions with child fixed effects. The time trends in the production functions, however, mean that individual heterogeneity would reflect deviations from the average rate of skill development. Eliminating this heterogeneity using child fixed effects leaves little variation to identify the marginal effects of inputs, and with only two observations per student in a value-added specification with student fixed effects the estimates are very imprecise.

to mainstream students in a given disability group. To create the instrument for disability group *i*, I first regress mainstreaming intensity on disability indicators, school characteristics, and school district fixed effects for all students with disabilities other than *i*. I use the estimated school district fixed effect as the instrument for mainstreaming intensity for students with disability *i* and estimate the production functions using two-stage least squares.[12] This instrument is clearly orthogonal to unobserved heterogeneity in skill levels and growth rates, but it will not be a valid instrument if, as seems likely, the unobserved school district factors that affect placement decisions for all disabled students (e.g., the availability of teacher aides) also affect the production of skills for any given disability. By contrast, the production functions in (2) with school district fixed effects account for these unobserved school district factors but not individual heterogeneity in skill growth rates.

Finally, I examine whether the relationship between mainstreaming intensity and skill development is linear. I estimate (2) including a quadratic mainstreaming intensity term to assess whether the marginal effect of mainstreaming intensity is constant at all intensities.[13] For

---

[12] Other instruments for mainstreaming might make use of variation in regional policies affecting mainstreaming practices as, for instance, Cullen (2003) uses changes in Texas' special education funding rules to explain variations in disability classification rates. Unfortunately SEELS contains no information regarding the location of the student's residence.

[13] By contrast, Hanushek et al. (2002) have only a binary indicator for whether the student spends all day in general education, meaning that they implicitly constrain the effects of mainstreaming to be linear. Experimenting with a variety of higher order polynomials, spline functions, and step functions indicated that all non-linearities in the effects of mainstreaming can be captured adequately by incorporating a quadratic term.

instance, mere exposure to more rigorous coursework or non-disabled students for short periods may have large effects on cognitive and social development. After this initial exposure, however, more intensively mainstreaming a student may have little effect. Indeed, mainstreaming may have deleterious effects on development at high intensities: being in general education all day necessarily means students are not receiving individualized special education instruction outside of the general education setting. Alternatively, it may be the case that students need to spend all or most of the day immersed in the more rigorous curriculum or surrounded by non-disabled students before mainstreaming has any effect on their development.

DATA

Special Education Elementary Longitudinal Study

The Special Education Elementary Longitudinal Study (SEELS) was funded by the Department of Education as part of a national assessment of the 1997 reauthorization of the Individuals with Disabilities Education Act (IDEA). The study documented the educational, social, vocational, and personal development over a four-year period of a national sample of students receiving special education services. Students between the ages of 6 and 12 and in at least first grade on December 1, 1999, were sampled from the special education rosters of participating school districts and special schools.[14] Roughly equal numbers of students were sampled from each of the twelve federal disability categories: specific learning disabilities,

[14] A stratified, random sample of 1,124 local education agencies (LEAs) and 77 state-supported special schools were invited to provide rosters of students receiving special education services in the designated age range in their district. Eligible LEAs were stratified by region, student enrollment, and the proportion of the student population living below the federal definition of poverty. A total of 245 LEAs and 32 special schools agreed to participate and provided rosters.

speech impairments, hearing impairments, visual impairments, orthopedic impairments, other health impairments, emotional disturbances, mental retardation, autism, traumatic brain injury, deaf-blindness, and multiple disabilities.[15] Data were collected in three waves in the 2000-2001, 2001-2002, and 2003-2004 school years from parent/guardian interviews, school program surveys, school characteristic surveys, language arts instructor surveys, and direct assessments of the students' math and verbal skills, self-concepts, and attitudes about school.[16] While 11,512 students were selected for participation, the sample is much smaller in practice because not all students could be contacted and not all parents consented to participation.[17]

Most longitudinal studies include few disabled children. Furthermore, many students in special education are routinely exempted from achievement tests, meaning that data concerning their cognitive development are scarce. SEELS is uniquely well-suited for examining the effects of mainstreaming intensity insofar as it provides data on the cognitive and noncognitive development of a large number of students with disabilities—including low-incidence disabilities—and extensive information regarding the school programs of and services received by these disabled students. Moreover, the children in the SEELS sample are relatively young. If

[15] Because they are low-incidence disabilities, all students with traumatic brain injuries or deaf-blindness in participating districts and special schools were selected for the sample.

[16] School characteristics surveys were not collected in the third wave.

[17] For example, only 9,747 parents were interviewed and 3,912 direct assessments conducted in Wave 1. The low completion rate for direct assessments in the first wave reflected a lack of available assessors. SEELS required that the assessor be unknown to the student, which eliminated most teachers in the students' schools. Relaxation of this requirement—combined with higher pay for assessors—resulted in higher completion rates in subsequent waves.

there are critical developmental periods early in life, then mainstreaming may be more likely to affect younger students with disabilities than older students with disabilities.[18] The major disadvantage of SEELS is that it provides no information regarding the location of the student's residence.[19]

Disabilities and Sample Selection

A central feature of my analysis is that I allow the production technologies for skills to vary by disability category. In some cases, different disabilities may result in similar production technologies, making it unnecessary to estimate separate production functions for all 12 federal disability categories. Grouping learning disabled students, emotionally disturbed students, and autistic students with each other or with other disabilities obscures important differences in the effects of mainstreaming. The remaining nine disabilities are grouped into three categories: mentally retarded students are grouped with students with traumatic brain injuries, multiple

[18] Hanushek et al. (2002) document extensive movement into and out of special education among elementary and middle school students. As such, older students remaining in special education may be those with more limiting disabilities for whom mainstreaming as well as any other treatment may be less beneficial.

[19] Additionally, SEELS does not include students who have not received special education services. As a result, SEELS data cannot be used to examine the effects of mainstreaming on the cognitive and noncognitive development of the non-disabled or the attitudes of the non-disabled towards the non-disabled nor to compare the development of disabled students to that of their non-disabled peers. Earlier work has explored the effect of mainstreaming on the cognitive development of the non-disabled (Salend & Duhaney, 1999; Hanushek et al., 2002) and on the attitudes of the non-disabled toward the disabled (Maras & Brown, 1996; Wong, 2008).

disabilities, and deaf-blindness to form the “severe” category; students with speech impairments are grouped with students with orthopedic impairments and other health impairments to form the “speech, orthopedic, and other health impairment” category; and students with hearing and visual impairments are grouped to form the “sensory impairment” category.

I create samples for each disability category for each of four “skills”—math skill, verbal skill, social skill, and self-image. To estimate the hybrid value-added production function described by (2) for skill j for each disability, I require measures of skill j in two waves. For the math skill, verbal skill, and self-image measures, this requires that the student complete the direct assessment in two waves. The social skills measure comes from the parent/guardian interview, meaning that a parent/guardian interview must be completed in two waves. The observations m_t and m_{t-1} may come from consecutive waves or from waves 3 and 1. In addition, I require that information regarding the student’s mainstreaming intensity, school program, and family be available in the same wave as the second skill measure m_t. This amounts to requiring that a school program survey and parent/guardian interview be completed in the same wave as the second skill observation.

Students who are “declassified” in the sense that they no longer have an individualized education plan (IEP) or are no longer receiving special education services and who are spending all day in general education may be unrepresentative of other disabled students. [20] For this

[20] IDEA mandates that students with disabilities have IEPs describing the student’s educational objectives for the current academic year and the instructional resources to be provided by the school district. IEPs are developed by schools in consultation with parents.

reason, I exclude person-year observations for students who indicate that they have been "declassified."[21]

Individuals with valid measures for the dependent variable in all three waves supply two person-year observations each; individuals only observed in two waves contribute a single person-year observation. The numbers of person-year observations in the resulting twenty-four samples are listed in table 4 along with the number of school districts represented in each sample. A total of 250 school districts are represented in the data.[22]

Variables

The measures of math and verbal skill and self-image are drawn from the direct assessment.[23] The direct assessment consists of the Woodcock Johnson III rapid letter naming and segmenting, oral reading fluency, letter-word identification, passage comprehension,

---

[21] The results detailed in the next section are not much affected by the exclusion of these observations. The percent of person-year observations provided students who have been declassified in the learning disorder, emotionally disturbed, speech, orthopedic and other health, severely disabled, autistic, and sensory impairment groups are 7.2%, 7.0%, 13.8% 1.0%, 1.6%, and 4.9%, respectively.

[22] The samples used to estimate the effects of mainstreaming on social skills are much larger than the other samples because the other samples were limited by the low completion rate of the direct assessment in the first wave.

[23] Some very low ability students completed alternate assessments; I do not include any such students in my sample. Students completed the direct assessment if they were able to complete the first item on the Woodcock Johnson III letter-word identification test and the alternate assessment otherwise.

mathematics calculation, and applied problems tests as well as Wick's School Attitude Measure, the Student Self-Concept Scale, and Asher's Loneliness in Children measure. The measures of math and verbal skill are the raw scores on the applied problems and passage comprehension tests. The raw scores are centered around 500, the score expected for an average child at ten years and zero months. The self-image measure from the Student Self Concept Scale reflects whether the student has a positive opinion about him/herself. Scores range from 5 to 20, with higher scores indicating a more positive self-image. The measure of social skills is an index based on parental responses in the parent/guardian interview to questions regarding the number of positive social behaviors the student demonstrates, with scores ranging from 7 to 21 and higher scores indicating better behavior.[24]

Table 1 provides the mean scores for the dependent variables by disability category. The mean math and verbal scores for the severely disabled students are more than a standard deviation below the corresponding means for students with learning disabilities, students with emotional disturbances and students with speech, orthopedic, and other health impairments, with mean scores for students with autism and students with sensory impairments falling between these extremes. Perhaps surprisingly, severely disabled students as a group have the highest average self-images with a mean self-image score of 13.4, compared to 12.8 for autistic students,

[24] Parents indicate whether the student joins groups without being told, makes friends easily, ends disagreements calmly, seems confident in social situations, avoids situations that are likely to result in trouble, starts conversations rather than waiting for others to start, receives criticism well, and controls his/her temper when arguing with other children. When parents failed to provide this information, answers to the same questions from the student's language arts teacher were used to supplement the parental responses.

who as group have the lowest average self-images. Autistic students also have the lowest average social skills, while students with learning disabilities and sensory impairments have the highest average social skills.

Table 2 shows the predicted year-to-year gain for each skill for each disability category from regressions of the change between measures of skill ($m_t - m_{t-1}$) on the elapsed time between measures (Δt_t). I use these predicted year-to-year changes to interpret the magnitudes of the predicted marginal effects of mainstreaming intensity. Two observations regarding these time trends are in order. First, self-images, in general, are improving with age. This is particularly interesting given that in the next section I find no evidence of significant effects of mainstreaming on self-image in spite of the fact that self-images appear to be malleable. second, the self-images and social skills of emotionally disturbed students improve significantly over time, with year-to-year gains several times the size of those of other disabled students.

The primary input of interest is the percent of the day a student spends in general education classrooms. Drawn from the school program survey, the mainstreaming intensity variable is constructed by dividing the number of minutes per week spent in general education by the total number of minutes per week at school.[25] As is evident from table 3 which details the correlation between mainstreaming intensity and the dependent variables, placement in general education is clearly correlated with a student's cognitive abilities and, to a lesser extent, their social skills, underscoring the importance of selection in the placement decision. Table 4 presents the distribution of students in each disability category by the percentage of the day spent

[25] When schools did not provide total minutes per week in school or the minutes per week spent in general education, I measure mainstreaming intensity by dividing the number of classes per day spent in a general education setting by the number of classes per day.

in general education. The table indicates that there is substantial within-group variation in mainstreaming intensity, but some groups clearly spend more time in general education than others. Students with severe disabilities spend the least time in general education, while students with speech, orthopedic, or other health impairments and learning disabilities spend the most time in general education. Students with sensory impairments exhibit the most interesting pattern: more than a quarter of such students spend no time in general education, while nearly fifty percent of these students spend more than three-quarters of their day in general education.

As measures of other school inputs, I include indicators for whether the student's school provided the student with a tutor, personal aide, or peer tutor, behavioral intervention (e.g., seeing a counselor or psychologist, receiving behavior modification training, or social skills instruction), and slower-paced instruction. As measures of family inputs, I include a measure of the weekly frequency with which parents read to the student and a parental interaction index summing the number of times per week parents talked to their child about school, helped with homework, volunteered at the child's school, and attended a general school meeting, school or class event, or parent-teacher conference other than individualized education plan (IEP) meetings. As proxies for familial preferences regarding student achievement affecting the demand for omitted inputs, I include the student's gender, race, mother's highest grade completed, the number of other children in the household, an indicator for the presence of other disabled children, an indicator for the presence of both parents, and household income. Finally, I include in the production function for skill j lagged measures of skills other than j, M_{t-1}^{-1}. To preserve larger sample sizes, I impute missing elements of M_{t-1}^{-1}.[26] The summary statistics for the

[26] When scores on the applied problems and passage comprehension tests are missing, I impute the missing score using the predicted score from regressions of applied problems scores on

covariates used to estimate the production functions for verbal skill are presented in table 5 by disability category.[27]

FINDINGS

I estimate skill production functions while dealing with the problems of unobserved inputs and the selection of more able students into general education to identify the relationship between mainstreaming intensity and development for students in six disability categories.[28] Table 6 presents the estimated coefficients of mainstreaming intensity from the four value-added production functions for each disability category.[29] Mainstreaming is clearly positively related to the development of math skill for most disabled students, with the largest estimated effect for severely disabled students. A ten percentage point increase in the percent of the day spent in general education is associated with a predicted increase in math skill equal to 22.3 percent

---

calculation test scores and passage comprehension scores on letter-word identification test scores. Similarly, I impute missing self-image scores using predicted scores from a regression of self-image on academic self-concept. Remaining missing values for these skills as well as the social skills measure are imputed using disability group means by age.

[27] The summary statistics for the samples used to estimate the production functions for other skills are similar to those reported in table 5.

[28] The importance of unobserved inputs and selection is apparent in specification (1) in appendix table 1, which reports the results of regressions of skill measures on only mainstreaming intensity and elapsed time between measures.

[29] Appendix tables 2, 3 and 4 present all estimated coefficients from the production functions for math skill, social skills and self-images for each disability category. The coefficients from the production functions for verbal skill are similar to those in the math skill production functions.

(=1.12/5.032) of the average year-to-year gain in math skill for severely disabled students, 18.6 percent (=1.37/7.355) for autistic students, 16.8 percent (=0.86/5.122) for emotionally disturbed students, and 15.7 percent (=0.64/4.078) for students with learning disabilities. The smallest predicted mainstreaming effects are for students with sensory impairments and students with speech, orthopedic, and other health impairments; for these groups, ten percentage point increases in mainstreaming intensity are associated with predicted increases in math skill equal to 11.2 percent and 13.3 percent of the average year-to-year gains, respectively. Likewise, ten percentage point increases in mainstreaming intensity are associated with predicted increases in verbal skill ranging between 13 percent and 18 percent of average year-to-year gains for all disabled students except for those with emotional disturbances, for whom a ten percentage point increase in mainstreaming intensity is associated with a predicted increase in verbal skill of only 5.0 percent of the average year-to-year gains, an effect statistically indistinguishable from zero. Table 6 also documents the statistically significant positive effects on social skills of mainstreaming, effects especially notable given that the social skills measure does not exhibit a significant time trend for most disabled students.

Given that a ten percentage point increase in time spent in general education amounts to a relatively small amount of time, these estimated relationships between mainstreaming and cognitive and social skills are considerable. This is not the case for the relationship between mainstreaming and self-image. As table 6 shows, the relationship between mainstreaming intensity and the development of self-image is never statistically significant.

My preferred specification (2) does not deal with unobserved heterogeneity in skill growth rates, but it does account for unobserved school inputs. Appendix table 1 reports the estimates from production functions without school district fixed effects. The estimated effects for cognitive and social skills are almost universally smaller than those in table 6. School districts that are "worse" in some unobserved dimensions appear to mainstream their students more intensively than other school districts.

After accounting for unobserved school inputs, the estimates in table 6 identify the causal effects of mainstreaming only if (1) the vector of lagged skill measures adequately controls for individual heterogeneity in ability levels and (2) schools make placement decisions based on skill levels and not skill growth rates. I conduct robustness checks to assess whether these conditions are satisfied. First, I include two lags of the dependent variable in the production function, effectively controlling for both skill levels and growth rates. As noted previously, the sample sizes for these specifications are substantially smaller and the standard errors are three to five times those in table 6, but the results, which can be found in the appendix, are mostly consistent with those in table 6. I reject equality of the estimated mainstreaming effects in table 6 and the effects from these specifications at the 10 percent level in 13 out of 30 regressions, but in almost half of these cases the estimated effects with two lags are actually larger than those reported in table 6.

Second, I instrument for mainstreaming intensity using the district's propensity to mainstream students in a given disability category. The results from this approach, also presented in the appendix, are again largely consistent in sign and magnitude with those presented in table 6 though the standard errors are two to four times larger.[30] Tests of the null hypothesis of equality of the mainstreaming effects in table 6 and from this instrumental variables approach

[30] Weak instrument tests suggest that the created "propensity to mainstream" variable is a fairly weak instrument, another reason why specification (2) is my preferred specification.

reject the null in only four out of 30 cases at the 10% level.[31] Comparing the estimates in table 6 with those from the specifications with two lags and the IV approach suggests that unobserved individual heterogeneity in skill growth rates does not substantially affect the estimated mainstreaming effects.[32]

To test whether the marginal product of mainstreaming is constant at all intensities, I include a quadratic mainstreaming intensity term in specifications otherwise identical to (2). In most cases, the mainstreaming effect is linear: I fail to reject the null hypothesis that the quadratic term can be excluded ($\alpha_{MS^2} = 0$) in 24 of the 30 production functions estimated. I interpret a negative marginal effect of mainstreaming to mean that instruction in special education settings is superior to instruction in general education settings. Those cases where I fail to reject a non-linear mainstreaming effect indicate that for some disabled students in some

---

[31] The failure to reject equality is in some instances driven more by the imprecision of the IV estimates than the similarity of the estimates themselves, as is the case for math skills. Most encouraging is the similarity between the estimated effects on social skills. The social skills samples are larger than the other samples, and hence the estimates tend to be more precise in all of the specifications.

[32] These robustness checks suggest that the value-added approach may overstate the benefits to mainstreaming for the development of math skill for learning disabled and emotionally disturbed students. The estimated mainstreaming effects on math skill for students with learning disabilities and students with emotional disturbances are -0.076 and -0.108, respectively, in the specification with two lags of the dependent variable and -0.115 and -0.134, respectively, using instrumental variables, though only the latter estimate is statistically significant at the 10 percent level.

domains there are significant benefits to time spent in special education settings. For instance, the math skills of students with sensory impairments are negatively affected by increased mainstreaming until they are spending at least 33 percent of their day in general education. In other words, low ability children with sensory impairments do not benefit from mere exposure to general education—a finding that probably explains the observation that more than a quarter of students with sensory impairments spend none of their day in general education. These students probably require the sort of intensive, individualized attention and communication aides that cannot be delivered in general education if they are to learn. By contrast, I find that spending more time in general education improves the verbal skills of severely disabled students—but only when they are spending less than 71 percent of their day in general education. This indicates that severely disabled benefit from exposure to general education—up to a point. On average, however, these students need to spend some portion of their day—even if only a small portion—pulled out of general education for individualized instruction. Thus these findings indicate that—even were it possible with substantial classroom support—integrating disabled students into general education for the whole day would not necessarily be desirable.

CONCLUSION

I estimate hybrid value-added production functions to identify the marginal effects of mainstreaming intensity on the development of math skill, verbal skill, locus of control, self-image, and social skills. I find that mainstreaming intensity has positive and in some cases very large predicted effects on the development of math, verbal and social skills for nearly all disabled youth. By contrast, mainstreaming has little effect on the loci of control and self-images of disabled youth.

Taken as a whole, these findings support the argument that mainstreaming benefits disabled youth cognitively by exposing them to more rigorous coursework taught by teachers certified in particular subjects areas. Given the data, it cannot be determined whether disabled students in general education fall behind their non-disabled peers, but it is clear that disabled students spending more time in general education learn more than peers spending less time in general education—even after accounting for the selection of more able students into general education. Furthermore, the positive effects on social skills and negligible effects on loci of control and self-images undermine the argument that the social isolation and frustration experienced by disabled students in general education classes harm their noncognitive development.

Given these findings, IDEA's mandate to educate disabled students in the "least restrictive environment" possible appears to be, for the most part, in students' best interests. The findings, however, come with two caveats. First, one can only identify the effects of mainstreaming on domains measured in the data. Thus if the measures of cognitive and noncognitive skills do not include important skills or traits affected by mainstreaming, then these findings may overstate the benefits to mainstreaming. Second, the effect of mainstreaming varies with the level of mainstreaming for some skills and some disability categories. The evidence of mainstreaming effects that vary with intensity level is inconsistent with the basic arguments supporting the "integration presumption" and highlights the need for further research on this issue. Thus while the findings indicate that the "least restrictive environment" mandate serves disabled students well on average, a nuanced approach to integrating disabled students factoring in the student's disability, skill level, and developmental goals is certainly advisable.

REFERENCES

Blackorby, J., Schiller, E., Knokey, A., & Wagner, M. (2007). Relationships Between the School Programs of Students with Disabilities and Their Longitudinal Outcomes. In What Makes a Difference? Influences on outcomes for students with disabilities. Menlo Park, CA: SRI International.

Boardman, A.E. & Murnane, R.J. (1979). Using Panel Data to Improve Estimates of the Determinants of Educational Achievement. Sociology of Education, 52, 113-121.

Bursztyn, Alberto M. (2007). The Praeger Handbook of Special Education. Westport, CT: Praeger Publishers.

Calhoun, G. & Elliott, R.N. (1977). Do contacts make a difference? The effects of mainstreaming on student attitudes toward people with disabilities. Exceptional Children, March 1977, 379-380.

Carlberg, C. & Kavale, K. (1980). The Efficacy of Special Versus Regular Class Placement for Exceptional Children: A Meta-Analysis. Journal of Special Education, 14(3), 295-309.

Colker, R. (2006). The Disability Integration Presumption: Thirty Years Later. University of Pennsylvania Law Review, 154(789).

Cullen, J.B. (2003). The impact of fiscal incentives on student disability rates. Journal of Public Economics, 87, 1557-89.

Cunha, F. & Heckman, J. (2008). Formulating, Identifying, and Estimating the Technology of Cognitive and Noncognitive Skill Formation. Journal of Human Resources, 43, 738-782.

Hanushek, E.A., Kain, J.F., & Rivkin, S.G. (2002). Inferring Program Effects for Special Populations: Does Special Education Raise Achievement for Students with Disabilities? Review of Economics and Statistics, 84, 584-599.

Hocutt, A. (1996). Effectiveness of Special Education: Is Placement the Critical Factor? The Future of Children, 6, 77-102.

Hunt, P. (2000). 'Community' is what I think everyone is talking about. Remedial and Special Education, 21(5), 305-317.

Manset, G. & Semmel, M.I. (1997). Are Inclusive Programs for Students with Mild Disabilities Effective? A Comparative Review of Model Programs. Journal of Special Education, 31(2), 155-180.

Maras, P. & Brown, R. (1996). Effects of contact on children’s attitudes toward disability: A longitudinal study. Journal of Applied Social Psychology, 26, 2113-2134.

Marston, D. (1987). The effectiveness of special education: a time series analysis of reading performance in regular and special education settings. Journal of Special Education, 21(4), 13-26.

Marston, D. (1996). A Comparison of Inclusion Only, Pull-Out Only, and Combined Service Models for Students with Mild Disabilities. Journal of Special Education, 30(2), 121-132.

Salend, S. J., & Duhaney, L. G. (1999). The impact of inclusion on students with and without disabilities and their educators. Remedial and Special Education, 20(2), 114-127.

Segal, C. (2008). Classroom Behavior. Journal of Human Resources, 43, 783-814.

Todd, P.E. & Wolpin, K.I. (2003). On the Specification and Estimation of the Production Function for Cognitive Achievement. Economic Journal, 113, F3-F33.

Todd, P.E. & Wolpin, K.I. (2007). The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps. Journal of Human Capital, 1, 91-36.

U.S. Department of Education. 2009. 28th Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act, 2006, vol. 1. Washington, D.C.: U.S. Government Printing Office.

Waldron, N.L. & McLeskey, J. (1998). The Effects of an Inclusive School Program on Students with Mild and Severe Learning Disabilities. Exceptional Children, 64(3), 395-405.

Weber, M.C. (2008). A Nuanced Approach to the Disability Integration Presumption. University of Pennsylvania Law Review Pennumbra 174, 156.

Wong, D.K. (2008). Do contacts make a difference? The effects of mainstreaming on student attitudes toward people with disabilities. Research in Developmental Disabilities, 29, 70-82.

Zigmond, N., Jenkins, J., Fuchs, L.S., Deno, S., Fuchs, D., Baker, J.N., Jenkins, L., & Couthino, M. (1995). Special Education in Restructured Schools: Findings from Three Multi-Year Studies. Phi Delta Kappan, 76(7), 531-540.

Table 1: Mean scores and standard deviations for dependent variables

	Skill Measure			
Disability	Math	Verbal	Self-image	Social skills
Learning disabled	505.89	493.62	13.40	16.09
	(23.69)	(18.44)	(1.77)	(2.68)
Emotionally disturbed	506.05	496.66	13.00	14.21
	(26.04)	(18.08)	(1.94)	(2.75)
Speech/Orthopedic/Other Health	501.24	495.06	13.25	15.74
	(31.06)	(20.27)	(1.89)	(2.89)
Severely disabled	469.43	473.33	13.48	14.62
	(36.88)	(26.34)	(1.89)	(3.10)
Autistic	478.40	480.83	12.80	13.37
	(43.19)	(25.64)	(2.20)	(2.83)
Sensory impairment	498.34	489.17	13.34	15.94
	(33.65)	(24.10)	(1.80)	(2.95)

Note: Standard deviations in parentheses.

Table 2: Predicted year-to-year changes in skills

	Skill Measure			
Disability	Math	Verbal	Self-image	Social skills
Learning disabled	4.078***	3.520***	0.075*	0.010
	(0.405)	(0.354)	(0.045)	(0.046)
Emotionally disturbed	5.122***	3.795***	0.103*	0.114*
	(0.469)	(0.429)	(0.061)	(0.060)
Speech/Orthopedic/Other Health	5.693***	4.231***	0.052*	0.015
	(0.272)	(0.226)	(0.029)	(0.030)
Severely disabled	5.032***	4.180***	0.099**	-0.060
	(0.402)	(0.345)	(0.046)	(0.039)
Autistic	7.355***	4.940***	-0.012	-0.087**
	(0.708)	(0.451)	(0.070)	(0.049)
Sensory impairment	6.413***	4.277***	0.009	-0.012
	(0.405)	(0.298)	(0.038)	(0.038)

Note: Each cell lists the coefficient on elapsed time between skill measures from a regression of the changes in skills between measures on the elapsed time between skill measures. Standard errors in parentheses.

*** — significant at 1%
** — significant at 5%
* — significant at 10%

Table 3: Correlations between mainstreaming intensity and skill measures

Disability	Math	Verbal	Self-image	Social skills
Learning disabled	0.33	0.33	0.02	0.15
Emotionally disturbed	0.26	0.26	0.00	0.15
Speech/Orthopedic/Other Health	0.41	0.36	0.00	0.19
Severely disabled	0.43	0.38	-0.06	0.17
Autistic	0.46	0.44	-0.09	0.23
Sensory impairment	0.42	0.53	-0.02	0.16

Table 4: Mainstreaming intensity patterns by disability

	Percentage of disability-specific verbal samples spending a given percent of the school day in general education					
Disability	0%	1-25%	26-50%	51-75%	76-99%	100%
Learning disabled	3.0	8.1	16.3	24.9	26.2	21.4
Emotionally disturbed	12.5	14.8	17.3	18.0	22.3	15.3
Speech/Orthopedic/Other Health	4.8	7.6	10.4	17.3	26.3	33.6
Severely disabled	15.4	24.4	23.8	17.1	12.5	6.7
Autistic	11.7	17.4	20.0	17.2	18.4	15.4
Sensory impairment	26.6	6.7	9.6	11.3	22.4	23.4

Table 5: Summary statistics for verbal skill samples

	Learning disabled	ED	Speech/ Orthop./ Other	Severely disabled	Autistic	Sensory
Percentage of day mainstreamed	68.37 (28.85)	55.11 (35.04)	73.39 (31.10)	41.23 (32.06)	51.19 (35.36)	55.22 (41.08)
Frequency of reading to child	2.32 (0.99)	2.17 (0.97)	2.39 (1.02)	2.68 (1.02)	2.69 (1.00)	2.41 (0.98)
Family interaction index	10.64 (2.49)	9.87 (2.70)	11.00 (2.47)	10.40 (2.60)	10.82 (2.70)	10.91 (2.40)
1 if tutor/aide provided	0.20	0.81	0.41	0.78	0.90	0.48
1 if behavior management services provided	0.09	0.56	0.12	0.17	0.31	0.08
1 if slow-paced instruction provided	0.33	0.37	0.29	0.60	0.40	0.33
1 if female	0.29	0.24	0.32	0.42	0.15	0.46
1 if black	0.13	0.21	0.13	0.26	0.12	0.18
1 if Hispanic	0.11	0.07	0.07	0.07	0.10	0.12
Household income ($)	40,326 (23,412)	33,935 (21,697)	46,401 (24,434)	36,022 (23,931)	52,286 (23,310)	42,838 (24,986)
Number of children in household	3.31 (1.22)	3.20 (1.29)	3.21 (1.08)	3.43 (1.36)	3.09 (1.07)	3.24 (1.03)
1 if two-parent household	0.70	0.55	0.73	0.60	0.77	0.70
Mother's highest grade completed	12.87 (1.85)	12.87 (1.76)	13.47 (1.88)	12.68 (1.84)	14.12 (1.68)	13.17 (1.92)
1 if other disabled children in household	0.36	0.31	0.26	0.29	0.22	0.24
Math score at *t-1*	498.12 (25.69)	496.60 (27.47)	491.40 (81.64)	460.90 (35.62)	464.88 (43.00)	487.07 (35.35)
Verbal score at *t-1*	487.06 (19.82)	490.02 (20.96)	487.83 (22.60)	465.86 (28.42)	472.79 (27.51)	481.44 (25.58)
Locus of control at *t-1*	9.95 (0.66)	10.07 (0.67)	10.08 (0.72)	9.81 (0.74)	9.99 (0.74)	10.13 (0.67)
Self-image at *t-1*	13.27 (1.78)	12.93 (1.98)	13.16 (1.83)	13.33 (1.73)	12.83 (1.94)	13.33 (1.71)
Social skills at *t-1*	16.02 (2.62)	14.01 (2.76)	15.74 (2.77)	15.09 (2.67)	14.11 (2.63)	16.09 (2.62)
Years between measures	1.70 (0.57)	1.66 (0.57)	1.67 (0.56)	1.66 (0.58)	1.66 (0.55)	1.71 (0.58)

Note: Standard deviations in parentheses.

Table 6: Estimated mainstreaming coefficients

			Skill Measure	
Disability	Math	Verbal	Self-image	Social skills
Learning disabled	0.064*	0.063**	0.00001	0.005
	(0.035)	(0.028)	(0.004)	(0.004)
Sample size	[602, 178]	[610, 179]	[579, 175]	[1,187, 208]
Emotionally disturbed	0.086***	0.019	0.0005	0.014***
	(0.032)	(0.023)	(0.004)	(0.003)
Sample size	[400, 131]	[411, 132]	[381, 129]	[847, 162]
Speech/Orthopedic/ Other Health	0.076***	0.056***	0.0006	0.003*
	(0.018)	(0.014)	(0.002)	(0.002)
Sample size	[1,510, 203]	[1,526, 203]	[1,416, 202]	[2,740, 218]
Severely disabled	0.112***	0.063**	-0.003	0.008***
	(0.035)	(0.026)	(0.003)	(0.002)
Sample size	[863, 167]	[873, 166]	[717, 153]	[2,313, 220]
Autistic	0.137***	0.075**	-0.002	0.006**
	(0.051)	(0.032)	(0.005)	0.003
Sample size	[506, 101]	[529, 102]	[390, 93]	[1,297, 136]
Sensory impairment	0.072**	0.069***	-0.0004	0.007***
	(0.033)	(0.023)	(0.003)	(0.002)
Sample size	[910, 169]	[963, 170]	[877, 164]	[2,103, 203]

Note: In addition to mainstreaming intensity, all regressions include as controls lagged skill measures, school inputs, family inputs, personal and family characteristics, time trends and school district fixed effects. The standard errors in parentheses are robust to the clustering of observations at the school district level. The number of person-year observations used to estimate the production function represented in each cell is given in brackets followed by the number of school districts represented in the sample.

*** — significant at 1%
** — significant at 5%
* — significant at 10%

Appendix Table 1: Estimated mainstreaming coefficients from alternative specifications

	Math			Verbal			Self-image			Social skills		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
LD	2.833	0.005	-0.076	2.714	0.0002	0.007	0.073	-0.003	0.015	0.103	0.003	**0.010**
	(0.127)	(0.030)	(0.183)	(0.124)	(0.022)	(0.090)	(0.004)	(0.003)	(0.021)	(0.004)	(0.003)	(0.009)
ED	1.960	-0.008	-0.108	1.870	-0.020	0.057	0.046	-0.000	-0.002	0.065	0.009	0.013
	(0.225)	(0.021)	(0.145)	(0.214)	(0.020)	(0.090)	(0.006)	(0.002)	(0.024)	(0.004)	(0.002)	(0.010)
SOO	2.648	0.072	**0.057**	2.512	0.011	**0.076**	0.063	-0.001	**-0.001**	0.084	0.003	**0.006**
	(0.103)	(0.012)	(0.037)	(0.105)	(0.014)	(0.032)	(0.003)	(0.002)	(0.006)	(0.003)	(0.002)	(0.003)
SD	2.354	0.071	**-0.071**	2.209	0.026	**0.012**	0.056	-0.004	0.001	0.072	0.007	**0.011**
	(0.119)	(0.023)	(0.083)	(0.124)	(0.021)	(0.063)	(0.004)	(0.002)	(0.008)	(0.003)	(0.002)	(0.005)
AUT	2.362	0.091	0.145	2.095	0.019	0.088	0.050	-0.002	-0.021	0.059	0.007	**-0.003**
	(0.148)	(0.043)	(0.162)	(0.146)	(0.022)	(0.104)	(0.006)	(0.004)	(0.020)	(0.003)	(0.002)	(0.005)
SEN	1.587	0.008	0.073	1.492	0.046	**0.053**	0.036	-0.004	**-0.002**	0.057	0.003	**0.009**
	(0.193)	(0.027)	(0.094)	(0.178)	(0.021)	(0.056)	(0.005)	(0.002)	(0.008)	(0.005)	(0.002)	(0.005)

Note: Specification (1) includes no controls other than mainstreaming intensity and the elapsed time between skill measures. Specification (2) corresponds to that reported in table 6 with the school district fixed effects omitted. Specification (3) corresponds to that reported in table 6 with the addition of a second lag of the dependent variable (m_{t-2}) included as a regressor. The standard errors in parentheses are robust to the clustering of observations at the school district level. Estimates in bold for specification (3) are those for which I reject equality of the coefficients from specification (3) with those reported in table 6 at the 10% level.

Appendix Table 2: Estimated coefficients from production functions for math skill

	Learning disabled	ED	Speech/ Orthop./ Other	Severely disabled	Autistic	Sensory
Percentage of day	0.064	0.086	0.076	0.112	0.137	0.072
mainstreamed	(0.035)	(0.032)	(0.018)	(0.035)	(0.051)	(0.033)
Frequency of reading to	0.630	-0.191	-0.446	-0.079	0.261	-0.282
child	(0.978)	(1.193)	(0.576)	(0.917)	(1.544)	(0.825)
Family interaction index	-0.412	-0.113	-0.008	0.467	0.530	-0.014
	(0.351)	(0.397)	(0.236)	(0.306)	(0.669)	(0.323)
1 if tutor/aide provided	-10.466	-9.730	-0.344	-4.545	6.652	2.894
	(12.031)	(17.401)	(1.740)	(2.820)	(9.624)	(3.138)
1 if behavior management	5.856	3.161	0.612	-2.470	0.508	2.165
services provided	(3.133)	(2.230)	(1.614)	(2.433)	(3.175)	(3.313)
1 if slow-paced instruction	-2.682	-3.847	-1.277	-2.241	-7.903	-4.378
provided	(1.747)	(2.344)	(1.188)	(1.746)	(2.962)	(2.009)
1 if male	2.955	3.490	5.065	3.641	2.580	2.485
	(2.025)	(2.616)	(1.052)	(1.789)	(3.385)	(1.505)
1 if black	-4.708	-2.148	-1.215	-7.441	-7.105	0.011
	(4.112)	(3.272)	(1.826)	(2.735)	(5.279)	(2.163)
1 if Hispanic	-1.613	4.162	0.449	8.798	-6.004	0.360
	(3.167)	(4.684)	(2.363)	(4.156)	(6.080)	(3.208)
Household income ($ in	0.053	0.031	0.005	-0.036	-0.067	0.054
thousands)	(0.048)	(0.059)	(0.026)	(0.053)	(0.080)	(0.045)
Number of children in	-0.811	1.227	0.055	-0.934	1.105	0.256
household	(0.678)	(1.087)	(0.467)	(0.651)	(1.096)	(0.716)
1 if two-parent household	3.219	-1.838	-1.852	1.851	-5.441	2.863
	(1.888)	(2.332)	(1.277)	(1.856)	(3.754)	(1.760)
Mother's highest grade	0.101	-0.550	0.007	-0.888	1.069	-0.187
completed	(0.534)	(0.670)	(0.294)	(0.524)	(0.943)	(0.455)
1 if other disabled children	0.837	-2.218	0.734	1.880	-0.124	-2.862
in household	(1.871)	(2.499)	(1.179)	(1.854)	(2.950)	(1.835)
Math score at *t-1*	0.430	0.588	0.580	0.640	0.557	0.580
	(0.045)	(0.054)	(0.032)	(0.051)	(0.059)	(0.045)
Verbal score at *t-1*	0.186	0.094	0.202	0.200	0.176	0.182
	(0.058)	(0.065)	(0.036)	(0.051)	(0.082)	(0.046)
Locus of control at *t-1*	-1.794	-0.285	3.228	-0.215	0.380	1.409
	(1.209)	(1.665)	(0.682)	(0.947)	(1.712)	(1.136)
Self-image at *t-1*	0.506	-0.498	-0.832	-0.555	0.132	-0.674
	(0.543)	(0.592)	(0.265)	(0.425)	(0.732)	(0.452)
Social skills at *t-1*	0.221	0.305	0.248	-0.226	0.393	0.120
	(0.304)	(0.397)	(0.192)	(0.296)	(0.586)	(0.270)
Years between measures	1.527	2.457	5.507	4.679	8.765	5.025
	(1.309)	(1.746)	(0.820)	(1.248)	(2.644)	(1.151)
R-squared	0.492	0.620	0.675	0.680	0.579	0.648

Note: Regressions include school district fixed effects. Robust standard errors in parentheses.

Appendix Table 3: Estimated coefficients from production functions for social skills

	Learning disabled	ED	Speech/ Orthop./ Other	Severely disabled	Autistic	Sensory
Percentage of day	0.005	0.014	0.003	0.008	0.006	0.007
mainstreamed	(0.004)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)
Frequency of reading to child	0.058	0.165	-0.013	-0.033	-0.083	0.009
	(0.103)	(0.141)	(0.056)	(0.075)	(0.082)	(0.076)
Family interaction index	0.080	0.009	0.079	0.153	0.094	0.116
	(0.040)	(0.049)	(0.022)	(0.026)	(0.031)	(0.032)
1 if tutor/aide provided	0.298	0.674	0.454	-0.222	0.722	-0.546
	(0.595)	(0.970)	(0.184)	(0.220)	(0.755)	(0.272)
1 if behavior management	-0.579	-0.339	-0.993	-0.442	-0.661	-0.977
services provided	(0.320)	(0.222)	(0.165)	(0.162)	(0.169)	(0.234)
1 if slow-paced instruction	0.122	-0.134	-0.039	0.303	-0.114	0.243
provided	(0.183)	(0.233)	(0.123)	(0.132)	(0.165)	(0.158)
1 if male	-0.300	-0.052	0.009	-0.184	-0.049	-0.043
	(0.191)	(0.270)	(0.102)	(0.129)	(0.194)	(0.128)
1 if black	0.281	0.345	-0.055	0.526	-0.025	0.110
	(0.347)	(0.282)	(0.171)	(0.191)	(0.263)	(0.207)
1 if Hispanic	-0.771	-0.104	0.058	0.202	-0.085	-0.339
	(0.372)	(0.480)	(0.222)	(0.278)	(0.341)	(0.248)
Household income ($ in	0.006	0.002	0.000008	-0.002	0.007	0.005
thousands)	(0.005)	(0.006)	(0.003)	(0.004)	(0.004)	(0.004)
Number of children in	-0.099	0.021	-0.013	-0.086	-0.046	-0.076
household	(0.089)	(0.100)	(0.048)	(0.055)	(0.081)	(0.068)
1 if two-parent household	0.162	0.261	0.012	-0.090	0.039	-0.000
	(0.226)	(0.260)	(0.130)	(0.159)	(0.220)	(0.171)
Mother's highest grade	-0.029	-0.160	0.003	-0.022	0.070	0.006
completed	(0.058)	(0.089)	(0.030)	(0.044)	(0.050)	(0.043)
1 if other disabled children in	0.202	0.057	0.053	0.115	-0.208	0.086
household	(0.216)	(0.270)	(0.115)	(0.162)	(0.192)	(0.167)
Math score at *t-1*	0.004	0.015	0.010	-0.001	0.001	0.010
	(0.005)	(0.006)	(0.003)	(0.003)	(0.003)	(0.003)
Verbal score at *t-1*	0.008	-0.025	-0.007	0.001	-0.008	-0.007
	(0.006)	(0.009)	(0.003)	(0.004)	(0.004)	(0.004)
Locus of control at *t-1*	-0.124	0.192	0.003	-0.074	0.103	0.205
	(0.160)	(0.216)	(0.090)	(0.112)	(0.141)	(0.128)
Self-image at *t-1*	0.015	0.012	0.074	-0.011	0.102	0.056
	(0.061)	(0.064)	(0.032)	(0.047)	(0.050)	(0.050)
Social skills at *t-1*	0.342	0.302	0.469	0.375	0.395	0.355
	(0.033)	(0.041)	(0.019)	(0.023)	(0.031)	(0.025)
Years between measures	-0.046	-0.101	-0.041	0.038	0.055	-0.013
	(0.129)	(0.162)	(0.083)	(0.106)	(0.132)	(0.108)
R-squared	0.186	0.177	0.314	0.200	0.247	0.207

Note: Regressions include school district fixed effects. Robust standard errors in parentheses.

Appendix Table 4: Estimated mainstreaming coefficients from IV specifications

		Skill Measure			
Disability		Math	Verbal	Self-image	Social skills
Learning disabled		-0.115	0.078	-0.020	0.006
		(0.113)	(0.095)	(0.012)	(0.011)
	t-statistic	1.51	-0.14	1.62	-0.05
Emotionally disturbed		-0.134*	-0.046	0.007	0.005
		(0.070)	(0.058)	(0.008)	(0.008)
	t-statistic	2.85***	1.05	-0.79	1.14
Speech/Orthopedic/Other Health		0.110	0.084	-0.016*	0.006
		(0.075)	(0.061)	(0.008)	(0.006)
	t-statistic	-0.44	-0.44	1.91*	-0.50
Severely disabled		0.103	0.110**	0.003	0.015***
		(0.065)	(0.053)	(0.006)	(0.005)
	t-statistic	0.12	-0.79	-0.86	-1.18
Autistic		0.155	0.039	0.008	0.014*
		(0.128)	(0.072)	(0.010)	(0.007)
	t-statistic	0.01	0.46	-0.91	-0.99
Sensory impairment		-0.029	0.073***	-0.006	-0.004
		(0.031)	(0.023)	(0.003)	(0.003)
	t-statistic	2.24**	-0.14	1.37	3.09***

Note: All regressions include as controls lagged skill measures, school inputs, family inputs, and personal and family characteristics. The regressions are estimated via two-stage least squares using the district's propensity to mainstream students with a given disability described in the text as the instrument for mainstreaming intensity. The standard errors in parentheses are robust to the clustering of observations at the school district level. The t-statistic tests the equality of the reported coefficients with those in table 6.

*** — significant at 1%
** — significant at 5%
* — significant at 10%

Appendix Table 5: Estimated mainstreaming and mainstreaming-squared coefficients

Disability			Math	Verbal	Self-image	Social skills
Learning disabled	MS		-0.0133	0.0253	-0.0015	**0.0268**
			(0.1172)	(0.0969)	(0.0127)	**(0.0121)**
	MS^2		0.0007	0.0003	0.00001	**-0.0002**
			(0.0010)	(0.0008)	(0.0001)	**(0.0001)**
		p-value	0.4844	0.6782	0.8992	**0.0613**
Emotionally disturbed	MS		0.0061	0.0313	-0.0169	**-0.0075**
			(0.1164)	(0.0837)	(0.0139)	**(0.0117)**
	MS^2		0.000793	-0.0001	0.0002	**0.0002**
			(0.0011)	(0.0008)	(0.0001)	**(0.0001)**
		p-value	0.4847	0.8847	0.1943	**0.0564**
Speech/Orth./Other	MS		0.0307	**-0.0310**	0.0031	-0.0016
			(0.0711)	**(0.0510)**	(0.0083)	(0.0064)
	MS^2		0.0004	**0.0008**	-0.00002	0.00004
			(0.0006)	**(0.0004)**	(0.00007)	(0.0001)
		p-value	0.4972	**0.0696**	0.7442	0.4544
Severely disabled	MS		0.1442	**0.1847**	-0.0143	0.0122
			(0.0980)	**(0.0792)**	(0.0087)	(0.0070)
	MS^2		-0.0003	**-0.0013**	0.0001	-0.00005
			(0.0009)	**(0.0007)**	(0.0001)	(0.00007)
		p-value	0.7211	**0.0821**	0.1968	0.5247
Autistic	MS		0.2384	0.1153	0.0044	0.0135
			(0.1829)	(0.0968)	(0.0182)	(0.0086)
	MS^2		-0.0010	-0.0004	-0.00006	-0.0001
			(0.0016)	(0.0009)	(0.0002)	(0.0001)
		p-value	0.5529	0.6521	0.6903	0.3531
Sensory	MS		**-0.1269**	-0.0227	0.0047	0.0162
			(0.1062)	(0.0785)	(0.0106)	(0.0085)
	MS^2		**0.0019**	0.0009	0.00005	0.0001
			(0.0010)	(0.0007)	(0.0001)	(0.0001)
		p-value	**0.0517**	0.2032	0.6129	0.2632

Note: Reported above are the coefficients of mainstreaming and mainstreaming squared from specifications identical to those reported in table 6 except for the addition of the mainstreaming squared term. The standard errors in parentheses are robust to the clustering of observations at the school district level. The p-values are from Wald tests of the hypothesis that $\alpha_{MS^2} = 0$. Cells in which the null $\alpha_{MS^2} = 0$ is rejected at the 10% level are in bold.