Article

Hammill Institute on Disabilities

The Journal of Special Education
2017, Vol. 51(1) 38–49

DOI: 10.1177/0022466916668164
journalofspecialeducation.sagepub.com
SAGE

Modeling the Time-Varying Nature of Student Exceptionality Classification on Achievement Growth

Joseph F. T. Nese, PhD[1], Joseph J. Stevens, PhD[1], Ann C. Schulte, PhD[2], Gerald Tindal, PhD[1], and Stephen N. Elliott, PhD[2]

Abstract

Our purpose was to examine different approaches to modeling the time-varying nature of exceptionality classification. Using longitudinal data from one state's mathematics achievement test for 28,829 students in Grades 3 to 8, we describe the reclassification rate within special education and between general and special education, and compare four alternative growth models for students with and without disabilities with different specifications of disability classification as time-variant (TVC) or time-invariant (TIC) covariates. Although model fit statistics were inconsistent in endorsing a single model, we found that the TIC results were generally preferable to the TVC; however, the choice of model specification may rest on the purpose of the researcher and goals of representing the influence of covariates on growth.

Keywords

achievement growth, time-varying covariates, exceptionality classifications, structural equation modeling

The Every Student Succeeds Act (ESSA) was signed into law in 2015 and requires each state to (a) implement a set of high-quality student academic assessments in mathematics, reading or language arts, and science, and (b) develop a statewide accountability system that includes long-term goals for improved academic achievement, as measured by proficiency on those student academic assessments and high school graduation rates. The accountability systems must include a valid indicator of student growth and also include performance reports for all students and disaggregated by economically disadvantaged students, race/ethnicity, English learners, and students with disabilities (SWD). How ESSA is implemented remains to be seen, but the performance of SWD remains an important part of the law's requirements for state accountability systems.

On average, SWD receiving special education services have lower achievement scores than their peers in general education (Morgan, Farkas, & Wu, 2011; Schulte, Stevens, Elliott, Tindal, & Nese, 2016; Shin, Davison, Long, Chan, & Heistad, 2013; Wei, Blackorby, & Schiller, 2011; Wei, Lenz, & Blackorby, 2012). Despite more recent attention to academic growth in accountability systems (e.g., ESSA), only recently have researchers begun to explore the achievement growth of SWD (e.g., Schulte & Stevens, 2015; Schulte et al., 2016; Stevens, Schulte, Elliott, Nese, & Tindal, 2015; Wei et al., 2011; Wei et al., 2012). For example, an unpublished literature synthesis examined student achievement growth research from 2007 to 2012 in 47 education and psychology journals (Stevens, Nese, & Tindal, 2013). Only 17% of the 277 reviewed studies evaluated academic growth for special education students, most often as a dichotomous categorization (receiving special education services or not). Only 3% of the reviewed studies evaluated special education results for specific exceptionality categories, and then almost exclusively for only one or two categories (usually students with learning disabilities and/or students with speech–language impairments).

However, it is important to understand the heterogeneity of academic performance within special education given that it can differ substantially across specific exceptionality categories (Schulte et al., 2016; Stevens et al., 2015; Wei et al., 2012). In addition, many students move in and out of SWD classification over time (Schulte & Stevens, 2015). These changes in student classification create complexities in correctly representing student status, but also present opportunities to model the representation of SWD status as either a time-invariant covariate (TIC) or a time-variant

[1]University of Oregon, Eugene, USA
[2]Arizona State University, Tempe, USA

Corresponding Author:
Joseph F. T. Nese, University of Oregon, 175 Education, Eugene, OR 97403, USA.
E-mail: jnese@uoregon.edu

covariate (TVC; see Curran, Lee, Howard, Lane, & MacCallum, 2012, for a review). TICs as predictors of growth provide direct tests of between-person differences in growth, whereas TVCs account for within-person changes in status over time and directly predict the respective repeated-measures outcome controlling for the influence of the latent growth trajectory factors (Curran et al., 2012).

Variability in Exceptionality Classification Across Students and Time

Changes in special education classification and the movement of students into and out of special education are well documented (Carlson & Parshall, 1996; Saven, Anderson, Nese, Farley, & Tindal, 2016; Schulte & Stevens, 2015; Walker et al., 1988; Ysseldyke & Bielinski, 2002). Walker et al. (1988) conducted a 2-year follow-up of 1,184 elementary-school SWD and found that of those who remained in the school district, 17% were no longer receiving special education services, and another 12% changed disability classification. Carlson and Parshall (1996) reported that 7% of SWD exit special education services each year with 4% of those students returning to special education within 3 years. Herring, McGrath, and Buckley (2007) also reported that the percentage of SWD grows from kindergarten (4%) through Grade 5 (12%). Ysseldyke and Bielinski (2002) reported that 9% to 13% of students exited special education and 8% to 17% entered special education across Grades 4 to 7, with those proportions generally decreasing across grades and with higher proportions entering than exiting. They compared two methods of defining special education membership: identification as an SWD in the first year of testing (similar to Schulte et al., 2016; Stevens et al., 2015; Wei et al., 2011; Wei et al., 2012) or current SWD status. In general, Ysseldyke and Bielinski found larger achievement differences between SWD and students without disabilities (SWOD) using the current year classification compared with the first-year classification. These findings were confirmed and extended by Thurlow, Wu, Lazarus, and Ysseldyke (2016), who reported that students who move from SWD to SWOD classification tend to be higher performing than those who remain in special education, and those who move from SWOD to SWD tend to be lower performing than those who remain in general education.

Similarly, Schulte and Stevens (2015) examined mathematics growth using three longitudinal methods for identifying the dichotomous disability subgroup: identification at Wave 1 (i.e., Grade 3), identification at any time during Grades 3 to 7 (“ever”), and continuous identification across all 5 years (“always”). Students always classified in special education were the lowest achieving group and experienced the least growth across grades, whereas students in special education at Grade 3 or those students ever classified in special education had higher and similar growth trajectories. A substantial gap between the achievement of SWD and SWOD was observed at each grade level, and the gap widened more across grades when disability subgroup membership was defined by the current year and allowed to vary by year.

Students also are sometimes reclassified from one exceptionality category to another. For example, Marder (2009) found that across a 3-year time span (2001–2004), only 61% of students identified with a disability in 2001 had the same primary disability category 3 years later. Of the remaining students, 24% had a different disability category, and 15% were no longer classified as having a disability. And as Puranik, Petscher, Al Otaiba, Catts, and Lonigan (2008) reported, the trajectories may differ for students of the same initial disability category based on whether their classification changes at a later point in time. For example, students with a persistent speech–language disability performed lower than students who exited services, and students initially identified with a speech–language disability who were later reclassified as having a learning disability performed the lowest of all.

Modeling Multiple Exceptionality Categories in Growth Models

Although exceptionality classification is expected to vary as a function of time, most often researchers use initial classification at the first measurement occasion as the only representation of SWD status over time, a time-invariant representation (e.g., Schulte et al., 2016; Stevens et al., 2015; Wei et al., 2011; Wei et al., 2012). Four recent studies have examined the achievement growth of SWD by specific exceptionality categories using a TIC representation, and only one study used a TVC.

Wei and colleagues (2011; 2012) estimated reading and math growth trajectories for a nationally representative sample of 3,421 SWD students aged 7 to 17 representing 10 federal disability categories using the Special Education Elementary Longitudinal Study (SEELS) database. They found a deceleration of achievement growth across time. In general, in comparison with students with learning disabilities, students with speech or visual impairments performed highest but, along with students with autism, improved at a lesser rate than students with learning disabilities; students with multiple disabilities or intellectual disabilities performed lowest. Although average achievement differed by exceptionality category, in general, students with different exceptionality classifications had similar growth curves for both reading and mathematics. Including specific SWD group predictors explained approximately 23% and 28% of the between-student variance in letter–word identification and passage comprehension, respectively, and only approximately 2% and 6% of the slope trajectory variance (i.e., linear and quadratic variance) in these domains, respectively (unconditional growth model results were not reported for math by Wei et al., 2012).

Schulte et al. (2016) and Stevens et al. (2015) estimated year-end reading and math growth trajectories in two state-wide samples of more than 90,000 SWOD and SWD students in seven exceptionality categories across Grades 3 to 7 with results showing faster growth in the early grades and some deceleration of both reading and math achievement across time. Schulte et al. reported small differences in reading growth rates across subgroups, with most exceptionalities growing more rapidly than SWOD, and gifted students growing more slowly; however, a relatively stable pattern of achievement gaps remained across grades. Stevens et al. reported that among the SWD subgroups, students with speech–language impairments were the highest performing in math, and students with intellectual disabilities were the lowest performing; they further reported stable, sizable math achievement gaps between SWOD and students in specific exceptionality categories.

Finally, a recent study by Shin et al. (2013) estimated reading and math achievement growth trajectories for 2,517 students across Grades 4 to 7. Dichotomous special education status was modeled using one possible conceptualization of SWD status as a TVC using structural equation modeling (SEM). The authors reported that SWD consistently had lower academic achievement levels, and the achievement gap between SWD and SWOD was stable over time.

The purpose of this study was to compare four different model specifications of exceptionality classification as a TIC or TVC predictor of achievement growth for SWD and examine how results and interpretations vary by model specification. The research questions were as follows for student academic growth from Grades 3 to 8:

Research Question 1: What are the reclassification rates between general and special education and across special education exceptionality categories?
Research Question 2: How do different specifications of exceptionality classification as TIC or TVC affect the estimated growth trajectories for SWD?
Research Question 3: What exceptionality model specification best fits the data: (a) TIC exceptionality classification defined at Wave 1, (b) TIC exceptionality classification defined across all years, (c) TVC exceptionality classification predicting within-student variance, or (d) TVC exceptionality classification with random slopes predicting within-student variance.

Method

Sample

The original sample for this study included the cohort of Grade 3 students in 2007–2008 who matriculated through Grade 8 in 2012–2013, and whose scores on the *Oregon Assessment of Knowledge and Skills* (OAKS) mathematics and reading tests were included in school accountability calculations in any of those years ($N = 39,216$). We excluded students who took the state alternate assessment and those off grade sequence (e.g., retained or skipped a grade). We included only students with a valid test score and complete data on SWD predictors in all 6 years. The final sample consisted of 28,829 students (74% of the original sample).

We recoded the state classification of students' disability in each year into six categorical indicator variables to reflect the following disability categories: Specific Learning Disability (SLD), Communication Disorder (CD), Emotional Disturbance (ED), Other Health Impairments (OHI), Autism Spectrum Disorder (ASD), and Other disabilities (i.e., Intellectual Disability, Hearing Impairment, Visual Impairment, Deaf–Blindness, Orthopedic Impairment, and Traumatic Brain Injury). The "Other" disability category was created by combining disability classifications that consisted of less than 0.12% of the sample across years. Note that this state listed only a primary disability classification. This sample included 49.3% female, 1.7% American Indian/Alaskan Native, 4.3% Asian/Pacific Islander, 2.2% Black/African American, 19.7% Hispanic, 67.6% White, and 4.5% Multiracial (see Table 1 for time-variant characteristics including free or reduced lunch [FRL] recipients, limited English proficiency [LEP] classification, exceptionality classification, the rate of entering/exiting special education services, and the rate of disability reclassification within special education).

Measures

For all analyses, the outcome measure was the student developmental scale score on the standardized OAKS (Oregon Department of Education [ODE], 2012a) mathematics test. The OAKS is a summative, computer-adaptive assessment based on the Oregon content standards (ODE, 2008). The test specifications varied by grade and subject and were intended to measure the core content standards in the state curriculum (ODE, 2012a). The tests were administered under standardized conditions (ODE, 2012b). OAKS raw scores were converted to scale scores using one parameter item response theory (IRT). The resulting Rasch Unit (RIT) scale scores were based on the number of items answered correctly while taking item difficulty into account (ODE, 2010) with the test vertically linked and centered on a scale score of 200 in Grade 3 (see Table 2 for the descriptive statistics [*M* and *SD*] of the OAKS mathematics test scores across grades and student subgroups).

Information on the development, operational procedures, and technical features of the state assessment system examined here are publicly available in annual reports updated to describe student performance and document

Table 1. Student Descriptive Characteristics and Reclassification Percentages Between General and Special Education and Within Special Education (N = 28,829).

Characteristic	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Ever[a]	Always[a]
FRL recipient	44.0	45.6	48.2	47.5	48.0	47.3	58.6	33.5
LEP	12.9	12.1	10.3	7.0	4.0	2.2	13.5	2.0
SWOD	89.5	88.2	88.3	89.1	89.7	90.2	95.1	82.8
SWD	10.5	11.8	11.7	10.9	10.3	9.8	17.2	4.9
CD	5.4	4.9	3.6	2.2	1.4	0.9	7.4	0.5
ED	0.2	0.2	0.3	0.3	0.3	0.3	0.5	0.1
OHI	0.7	0.9	1.2	1.4	1.6	1.7	2.2	0.4
ASD	0.6	0.7	0.8	0.8	0.9	0.8	1.0	0.5
SLD	3.4	4.8	5.6	5.9	5.9	5.8	7.9	2.0
Other disability	0.1	0.1	0.1	0.1	0.1	0.1	0.3	0.1
Enter	—	2.8	1.8	1.3	0.7	0.6	—	—
Exit	—	1.5	1.9	2.2	1.3	1.1	—	—
Disability reclassification	—	0.5	0.5	0.5	0.4	0.3	—	—

Note. FRL = free reduced lunch status; LEP = limited English proficiency; SWOD = students without a disability; SWD = students with a disability; CD = communication disorder; ED = emotional disturbance; OHI = other health impairments; ASD = autism spectrum disorder; SLD = specific learning disability; and other disabilities include the following: intellectual disability, hearing impairment, visual impairment, deaf-blindness, orthopedic impairment, traumatic brain injury. Enter/Exit = percent entering/exiting special education services during that year. Disability reclassification = percent of disability reclassification within special education.
[a]The percent of students who "ever" or "always" received that service or who were classified with that disability.

Table 2. Mean (SD) of State Mathematics Test Scores Across Grades and Student Subgroups.

	Grade					
Characteristic	3	4	5	6	7	8
FRL recipient	208.34	216.10	222.55	226.00	232.55	234.55
	(9.17)	(9.10)	(8.39)	(9.45)	(8.58)	(9.59)
LEP	204.58	211.83	218.35	219.17	224.86	224.71
	(9.22)	(8.33)	(7.46)	(8.31)	(7.45)	(7.80)
SWOD	212.28	220.23	226.33	230.50	236.69	239.11
	(9.34)	(9.83)	(9.04)	(9.63)	(9.19)	(10.14)
SWD	207.09	214.06	219.24	220.91	227.25	228.10
	(10.53)	(10.68)	(9.43)	(10.51)	(8.99)	(9.88)
CD	209.21	17.54	223.18	225.13	230.22	230.71
	(10.65)	(11.31)	(10.08)	(11.98)	(10.96)	(11.60)
ED	206.65	213.74	218.77	223.32	228.89	231.84
	(11.19)	(9.80)	(10.16)	(10.11)	(9.02)	(10.84)
OHI	205.21	212.25	218.02	220.95	226.96	227.67
	(10.48)	(9.83)	(8.70)	(10.40)	(8.75)	(10.11)
ASD	207.15	215.00	220.97	224.97	232.35	234.22
	(10.11)	(11.15)	(10.69)	(12.07)	(10.42)	(11.50)
SLD	204.17	210.82	216.81	218.62	225.86	226.77
	(9.57)	(8.69)	(7.83)	(8.89)	(7.70)	(8.57)
Other disability	205.97	211.56	217.54	220.98	225.20	226.22
	(10.82)	(13.85)	(11.19)	(11.94)	(11.27)	(12.60)

Note. FRL = free reduced lunch status; LEP = Limited English proficiency; SWOD = students without a disability; SWD = students with disabilities; CD = communication disorder; ED = emotional disturbance; OHI = other health impairments; ASD = autism spectrum disorder; SLD = specific learning disability; and other disabilities = intellectual disability, hearing impairment, visual impairment, deaf-blindness, orthopedic impairment, traumatic brain injury.

changes to the system (ODE, 2012a). The state also reports on test development, standard setting, validity evidence, test administration procedures, and interpretive guides (ODE, 2012a). Reliability evidence for the OAKS mathematics tests includes test information curves that suggest that across grades and subgroups (i.e., ethnicity, LEP, and SWD), 90% of the sample had a standard error (SE) of measurement of approximately 3 RIT scale score points (ODE, 2007). The SE increased for students in the tails of the distribution, particularly for those in the 99th percentile. Concurrent validity evidence across grades is provided by high correlations of OAKS mathematics scores with scores on the California Achievement Test (.74−.80), and the Iowa Tests of Basic Skills (.76−.85; ODE, 2007).

Analyses

Four alternative models for representing mathematics growth for SWD were compared, each with different specifications of disability classification as a TIC or TVC predictor. All models were specified as latent growth curve models using Mplus Version 7.31 (Muthén & Muthén, 1998–2014) and maximum likelihood estimation with robust SEs. It is important to note that an SEM framework was applied for all growth models (multivariate outcome), and that this approach to model specification is different from a multilevel (univariate outcome) framework (e.g., hierarchical linear modeling [HLM]; Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011; lme4, Bates, Maechler, Bolker, & Walker, 2014). Although both frameworks are suitable and appropriate for these models, they parameterize the covariance structure among the TVCs and the random intercepts differently. For a detailed, informative comparison of

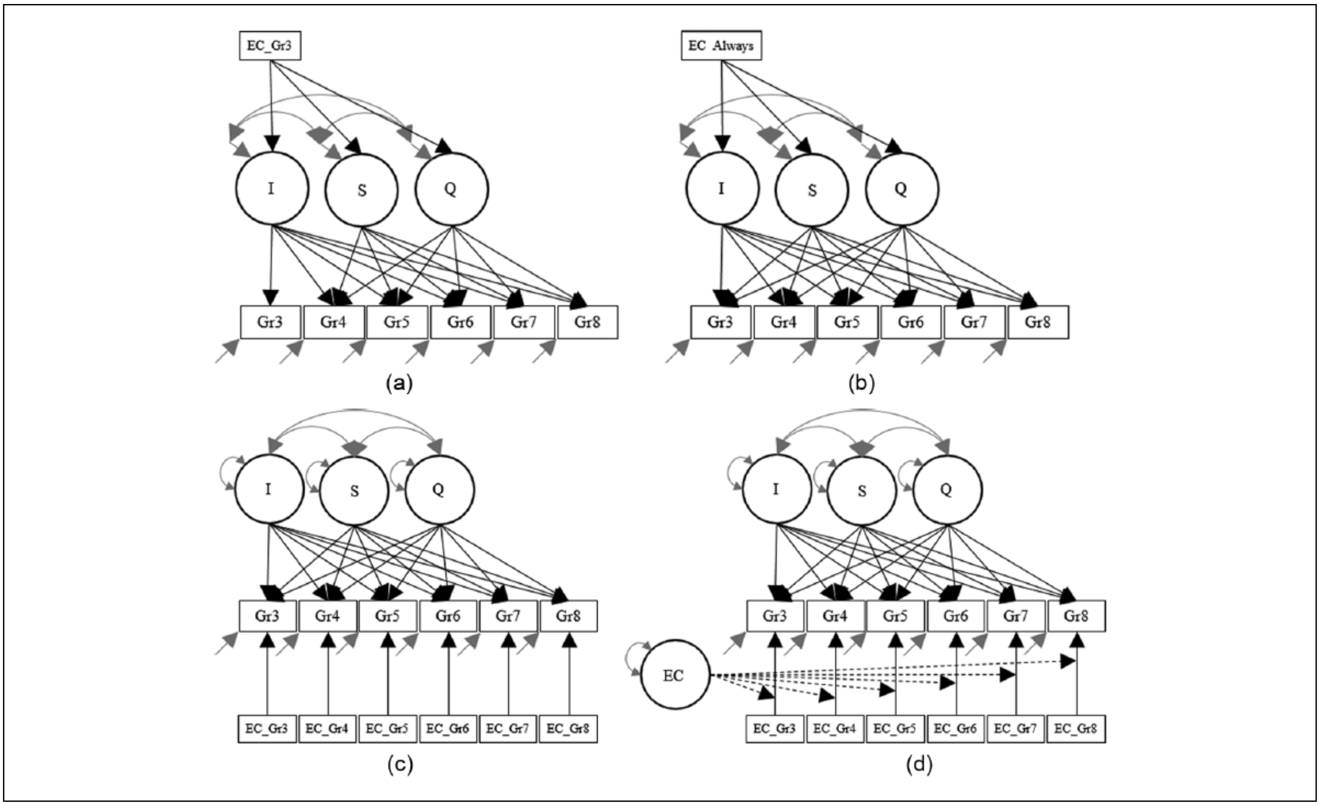

Figure 1. Structural equation models representing each of the four SWD specifications: (a) TIC–Wave 1 model: Grade 3 SWD predictors as TIC; (b) TIC–always model: Grade 3 through 8 SWD predictors as TIC; (c) Intra-individual TVC-fixed effects model: SWD predictors as TVC; and (d) Intra-individual TVC–random slopes model: SWD predictors as TVC with random slopes.
Note. Note that EC represents each of the six exceptionality codes used in the analyses—communication disorder, emotional disturbance, other health impairments, autism spectrum disorder, specific learning disability, and other disabilities (i.e., intellectual disability, hearing impairment, visual impairment, deaf-blindness, orthopedic impairment, and traumatic brain injury). SWD = students with disabilities; TIC = time-invariant covariate; TVC = time-variant covariate; EC = exceptionality code.

SEM and multilevel growth models with TVC and TIC, see Curran et al. (2012).

TIC–Wave 1 model. The TIC–Wave 1 model considered the SWD exceptionality categories as time-invariant predictors, estimating the between-student effects of Grade 3 exceptionality category on latent math growth trajectory parameters (intercept, linear, and quadratic; Figure 1a). This model is comparable with the specification of exceptionality categories in most previous academic growth modeling research (e.g., Schulte et al., 2016; Stevens et al., 2015; Wei et al., 2011; Wei et al., 2012). In this model, the TIC exceptionality categories at Grade 3 had direct effects on the latent trajectory parameters and indirect effects on the mathematics outcome at all occasions.

TIC–always model. The TIC–always model considered SWD exceptionality categories as time-invariant predictors, estimating the between-student effects of being classified in a particular SWD classification in all Grades 3 through 8 on latent math growth trajectory parameters (intercept, linear, and quadratic; Figure 1b). This model is somewhat comparable with the specification of exceptionality as participation in special education (dichotomous) in all Grades from 3 to 7 (always in special education) in Schulte and Stevens (2015); however, in this model, "always" indicates classification across all grades in each specific exceptionality category rather than special education overall. In this model, the TIC "always" exceptionality variables had direct effects on the latent trajectory parameters and indirect effects on the mathematics outcome at all occasions.

Intra-individual TVC–fixed effects model. The intra-individual TVC–fixed effects model considered the separate SWD exceptionality categories as time-varying covariates, regressing the mathematics scores at each grade on the corresponding grade-level status of each student in the specific exceptionality categories (Figure 1c). In this model, the SWD predictors account for the within-student variance of math scores, controlling for the between-student growth effects. As described by Curran et al. (2012), this model captures the within-student relation between the

SWD exceptionality categories and the math score at Time *t*, but does not provide information about the potential between-student relation of math proficiency and SWD classification.

Intra-individual TVC–random slopes model. This model is the same as the previous model with one change in specification: The intra-individual TVC–random slopes model specified the SWD exceptionality categories as time-varying predictors with random rather than fixed slope parameters (Muthén & Muthén, 1998–2014). The intra-individual TVC–random slopes model allows the values of the TVC coefficients to vary randomly between students such that a random effect for each SWD exceptionality category is estimated for each student. In other words, each exceptionality category was modeled as a latent variable with random regressions on the mathematics outcomes at each grade (Figure 1d).

Results

Table 1 displays the reclassification percentages between SWOD and SWD and for specific exceptionality categories for the analytic sample of students from Grade 3 to Grade 8. The percentage of SWOD increased from Grade 4 (88%) through Grade 8 (90%); however, the increase was small, from 0.1% to 0.9% between consecutive years. Approximately 83% of students were "always" in general education throughout the study duration. Between 10% (n = 2,867) and 12% (n = 3,448) of the sample were classified as an SWD in any one year from Grades 3 to 8. Only a small proportion of the total sample remained in the same SWD exceptionality category across Grades 3 to 8: CD = 0.5%, ED = 0.1%, OHI = 0.4%, ASD = 0.5%, SLD = 2.0%, and Other = 0.1% (SWOD = 82.8%); however, this is a function of the small number of students in any of the SWD categories. The percentages of students who were always in a category as a percentage of those who were ever in that category (always / ever) were quite varied: CD (6%), ED (15%), OHI (19%), SLD (25%), Other (27%), and ASD (46%). The percentage of students with CD decreased from Grade 3 (5.4%) through Grade 8 (0.9%), and the percentage of students with SLD increased from Grade 3 (3.4%) through Grade 8 (5.8%). In addition, the percentage of students classified as OHI steadily increased across grades (from 0.7% to 1.7%), whereas the percentage of students classified with ASD, ED, or with Other disabilities remained relatively stable across time.

The percentage of students who entered special education services in each year decreased from Grade 4 (2.8%) through Grade 8 (0.6%; see row *Enter* in Table 1). The percentage of students exiting special education services each year increased from Grade 4 (1.5%) to Grade 5 (1.9%) and Grade 6 (2.2%), followed by a decrease in Grade 7 (1.3%) and Grade 8 (1.1%; see row *Exit* in Table 1). On average, only 0.3% to 0.5% of the sample reclassified disability category from one year to the next, but reclassifications became slightly less common in Grades 7 and 8 (see row "Disability Reclassification" in Table 1).

Growth Models

In our growth analyses, we first applied a baseline, unconditional quadratic growth model. The quadratic form was chosen based on observed student and group mean mathematics score trajectories, as well as better information criteria (model fit) in comparison with a linear growth model. The unconditional growth model had a mean Grade 3 (intercept) math score of 211.88 (SE = 0.06), an initial linear slope of 7.65 (SE = 0.03), and an initial quadratic deceleration of −0.48 (SE = 0.01). Across models, all estimated trajectory means (intercept, linear, and quadratic) and *SE*s for the SWOD reference group were similar to those of the unconditional model.[1]

Figures 2 and 3 show the estimated growth trajectories for the specific exceptionality categories over time for the alternative models. Note that Figure 2 and Figure 3 represent the same estimated trajectories organized differently. Each panel of Figure 2 represents one of the four alternative growth model specifications and is designed to facilitate comparisons of the separate exceptionality group growth trajectories within each panel and also includes the SWOD trajectory. However, each panel of Figure 3 shows only one of the six exceptionality groups and is designed to facilitate comparisons of the four alternative growth models for that exceptionality group. An important caveat in interpreting these figures is that the mean trajectories do not necessarily represent the same student subgroups across models. For example, the SLD trajectory for the TIC–Wave 1 model represents the average intercept and growth for students who were classified as SLD in Grade 3, irrespective of later classification. The SLD trajectory for the TIC–always model represents the average intercept and growth of students who were classified as SLD in every Grade from 3 to 8. Thus, the direct comparison of estimated means across models represents differences in the definition of student exceptionality subgroups, and therefore, somewhat different groups of students may be represented in the trajectories for the different models. These different model trajectories are further described below.

TIC–Wave 1 Model

The relation of each of the Grade 3 SWD exceptionality category predictors with Grade 3 math scores (intercept) was negative and statistically significant in comparison with the SWOD reference group (see the upper left panel of Figure 2). Students with a SLD had the largest difference from the intercept ($\beta = -8.26$), followed by OHI ($\beta = -7.07$),

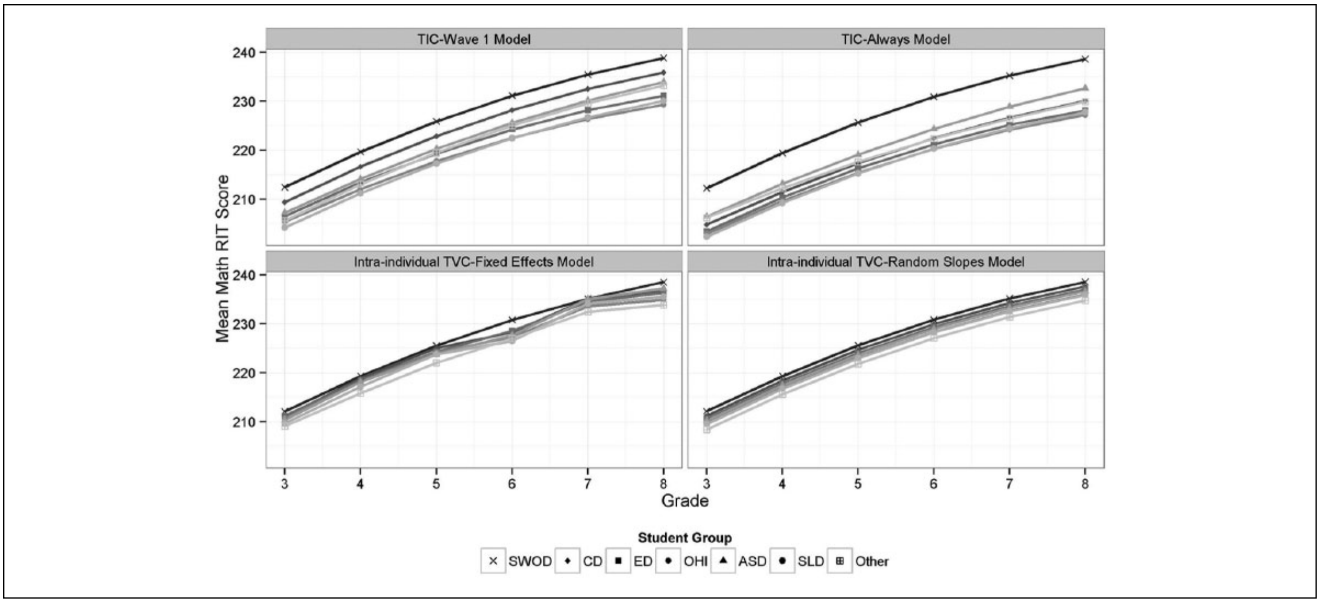

Figure 2. Estimated mean mathematics achievement trajectories for the four alternative models.
Note. SWOD = students without disabilities; CD = communication disorder; ED = emotional disturbance; OHI = other health impairments; ASD = autism spectrum disorder; SLD = specific learning disability; and Other = intellectual disability, hearing impairment, visual impairment, deaf-blindness, orthopedic impairment, and traumatic brain injury; TIC = time-invariant covariate; TVC = time-variant covariate.

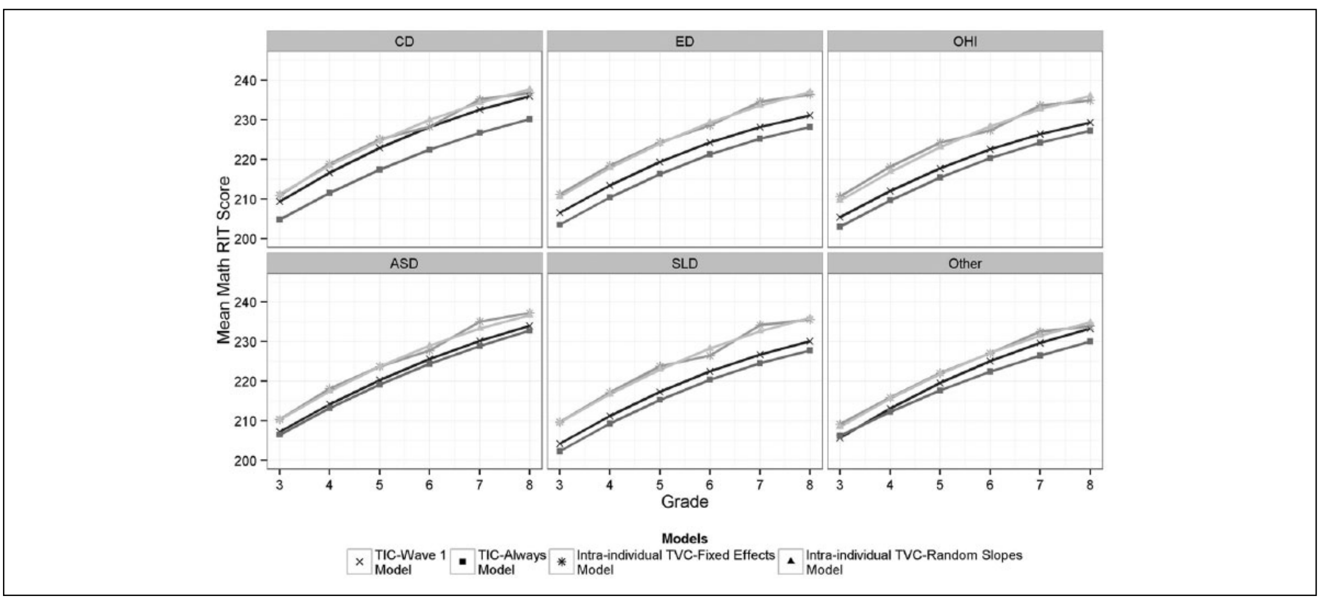

Figure 3. Estimated mean mathematics achievement trajectories across models for students in each disability category.
Note. CD = communication disorder; ED = emotional disturbance; OHI = other health impairments; ASD = autism spectrum disorder; SLD = specific learning disability; and Other = intellectual disability, hearing impairment, visual impairment, deaf-blindness, orthopedic impairment, and traumatic brain injury; TIC = time-invariant covariate; TVC = time-variant covariate.

Other ($\beta = -6.85$), ED ($\beta = -5.95$), ASD ($\beta = -5.24$), and CD ($\beta = -3.09$). No Grade 3 SWD exceptionality category TIC was a statistically significant predictor of either the linear or quadratic growth parameters. Thus, initial math scores were lower on average for students in each SWD exceptionality category than their SWOD peers; however, we found no evidence that students in any of the SWD exceptionality categories grew at statistically different rates from SWOD. Examination of the variance components for the model parameters showed that there were statistically significant between-student differences in their intercept, linear, and quadratic parameters.

TIC–Always Model

The relation of each of the "always" SWD exceptionality category predictors with Grade 3 math scores (intercept) was negative, statistically significant in comparison with the SWOD reference group, and with the exception of the Other classification larger than the coefficients for the TIC–Wave 1 model (see the upper right panel of Figure 2). Students with a SLD had the largest difference from the intercept ($\beta = -9.93$), followed by students classified as OHI ($\beta = -9.28$), ED ($\beta = -8.78$), CD ($\beta = -7.43$), Other ($\beta = -6.03$), and ASD ($\beta = -5.77$). Similar to the TIC–Wave 1 model, no SWD exceptionality category TIC was a statistically significant predictor of either the linear or quadratic growth parameters so that initial math scores for students in each exceptionality category were lower on average than their SWOD peers, but growth trajectories were not statistically different from SWOD. As with the TIC–Wave 1 model, examination of the variance components for the model parameters showed that there were statistically significant between-student differences in their intercept, linear, and quadratic parameters.

Intra-Individual TVC–Fixed Effects Model

All intercept coefficients of the SWD TVC were negative in direction, but ranged substantially in magnitude from −0.02 to −4.64 in comparison with SWOD (see the lower left panel of Figure 2). In other words, SWD exceptionality category classification, relative to grade-level SWOD peers, was generally associated with lower math scores. Students in the SLD group were statistically significantly lower than SWOD at each grade (−0.98 to −4.33), but the performance of students in the ED group never differed significantly from the SWOD group at any grade. The results for the remaining exceptionality subgroups yielded no consistent patterns of differences across grades in comparison with the SWOD group: The CD group differed significantly at Grades 3, 6, and 8; the OHI group differed at Grades 5 through 8; the ASD group differed at Grades 5 and 6; and the Other group differed at Grades 4, 5, 6, and 8. Thus, controlling for the between-student growth effects, classification as a student with a SLD accounted for significant within-student variance of math scores at each wave, whereas the relation of the other SWD exceptionality groups was inconsistent across waves.

Intra-Individual TVC–Random Slopes Model

The means of the latent SWD exceptionality category TVC coefficients were negative and statistically significant for all groups in comparison with the SWOD group (CD = −0.91; OHI = −2.51; ASD = −1.90; SLD = −2.61; Other = −3.75; see the lower right panel of Figure 2) with the exception of students with ED (−1.55, SE = 0.48). The means for each exceptionality category represented the average relation of the exceptionality category membership with math performance over time, indicating that (except for students with ED) the SWD exceptionality groups had statistically significant relations with math achievement over time, controlling for average latent growth parameters. Examination of the random variance components showed that only the OHI and SLD groups had statistically significant variation among students; the remaining TVC exceptionality slope parameters did not randomly vary within the other student exceptionality groups.

Model Comparison and Fit

The model fit statistics are presented in the upper portion of Table 3.[2] Model fit was evaluated using rules of thumb suggested by Hu and Bentler (1999). All models but the intra-individual TVC–fixed effects model showed adequate fit on the Tucker–Lewis Index (TLI; ≥ .95). Comparing the root mean squared error of approximation (RMSEA) and the standardized root mean squared residual (SRMR), only the intra-individual TVC–fixed effects model showed adequate fit (RMSEA ≤ .06, SRMR ≤ .08). Thus, despite the added complexity, the intra-individual TVC–fixed effects model that included SWD covariates at each grade yielded the best values on two of the three fit indices.

Comparing fit across models using Akaike information criterion (AIC), the Bayesian information criterion (BIC), and the adjusted Bayesian information criterion (ABIC), the intra-individual TVC–fixed effects and the TIC–Wave 1 models showed the best fit; the former had the lowest AIC values, while the latter had the lowest BIC and ABIC values. Note that the BIC and ABIC both apply a penalty for every parameter estimated based on sample size (with the ABIC penalty less severe). Thus, the intra-individual model TVC–fixed effects model was subject to penalties for complexity, and the TIC–Wave 1 model by comparison was favored in terms of parsimony. The bottom portion of Table 3 shows the R^2 for the endogenous variables of the respective models. Compared with the unconditional quadratic growth model, the TIC–Wave 1, TIC–always, and intra-individual TVC–fixed effects models did not account for any additional variance in any of the repeated achievement measures. The TIC–Wave 1 and TIC–always models explained equal variance in the intercept, linear, and quadratic latent growth factors.

Discussion

In response to the first research question, we documented the movement of students between general and special education as well as reclassification from one disability category to another (e.g., Carlson & Parshall, 1996; Schulte & Stevens, 2015; Thurlow et al., 2016; Walker et al., 1988;

Table 3. Model Fit Indices, Model Information Criteria, and R^2 for Endogenous Variables.

	Unconditional model	TIC–Wave 1 model	TIC–always model	TVC–fixed effects model	TVC–random slopes model
Fit indices					
Chi-square	4,203.27	4,091.18	3,727.51	6,893.90	—
df	12	30	30	192	—
AIC	11,24,766.39	11,23,447.38	11,23,598.57	11,23,372.50	11,24,084.83
BIC	11,24,890.43	11,23,720.26	11,23,871.45	11,23,794.22	11,24,308.09
ABIC	11,24,842.76	11,23,615.38	11,23,766.58	11,23,632.14	11,24,222.29
RMSEA 90% CI	[.11, .11]	[.07, .07]	[.06, .07]	[.03, .04]	—
TLI	.96	.95	.95	.94	—
SRMR	.33	.18	.18	.07	—
Endogenous variables (R^2)					
Intercept	—	.04	.04	—	—
Linear slope	—	.00	.00	—	—
Quadratic slope	—	.00	.00	—	—
Math Grade 3	.76	.76	.76	.76	—
Math Grade 4	.76	.76	.76	.75	—
Math Grade 5	.80	.80	.80	.80	—
Math Grade 6	.79	.79	.79	.79	—
Math Grade 7	.83	.83	.83	.82	—
Math Grade 8	.85	.85	.85	.84	—

Note. TIC = time-invariant covariate; TVC = time-variant covariate; AIC = Akaike information criterion; BIC = Bayesian information criterion; ABIC = adjusted Bayesian information criterion; RMSEA = root mean squared error of approximation; CI = confidence interval; TLI = Tucker–Lewis Index; SRMR = standardized root mean squared residual. For the intra-individual TVC–random slopes model, the variance of the repeated measures varied with disability classification, which precludes the calculation of standardized coefficients and chi-square and related fit statistics.

Ysseldyke & Bielinski, 2002). Similar to the findings by Schulte and Stevens (2015), approximately 10% (n = 2,830) to 12% (n = 3,409) of the sample cohort received special education services in any one grade from 3 through 8, with 11% (n = 3,032) receiving services in Grade 3, about 17% (n = 4,957) receiving special education services in at least 1 year, and approximately 5% (n = 1,404) receiving special education services in every year in Grades 3 to 8. Furthermore, a very small proportion of students remained in the same disability category across Grades 3 to 8, ranging from 0.07% for ED to about 2% for SLD (see column "Always" of Table 1).

Approximately 7% (n = 2,035) of the sample entered special education services at some point during Grades 4 to 8, which is lower than the 8% to 17% range reported by Ysseldyke and Bielinski (2002). Approximately 8% (n = 2,241) of the sample exited special education, which is similar to the 7% reported by Carlson and Parshall (1996) and at the lower end of the 9% to 13% range reported by Ysseldyke and Bielinski (2002). Except at Grade 4, a higher percentage of students exited special education services than entered, opposite to the pattern reported by Ysseldyke and Bielinski. The percentage of students with CD decreased from Grade 3 through 8, and the percentage of students with SLD increased from Grade 3 through 8, similar to the findings by Herring et al. (2007). The percentage of students classified with OHI increased across grades, and the percentage of students classified with ED, ASD, or Other remained relatively stable across grades. Thus, our sample demonstrated that there is considerable movement of students in, within, and out of special education classifications. This finding is similar to that reported by others and facilitates the generalization of our subsequent findings.

In response to the second research question, each of the model specifications for disability classification resulted in very similar average growth trajectories for SWOD but substantially different estimated growth trajectories for SWD (Figure 2). The differences across models were related to specifications that imply different conceptualizations of growth and its relation to the covariates, and therefore, the different models provided different interpretations of intra- and inter-individual differences in growth. Specifically, for the intra-individual TVC–fixed effects and random slopes models, the SWD exceptionality TVCs accounted for variance in each of the achievement outcomes controlling for the latent growth trajectory parameters (i.e., within-student variance, controlling for latent math growth). Whereas there was a single fixed effect for each TVC at each occasion in the intra-individual TVC–fixed effects model, the intra-individual TVC–random slopes model allowed the values of the TVC coefficients to vary randomly between students such that each student's TVC value was estimated. Because only the OHI and SLD groups had statistically significant variances in the intra-individual TVC–random slopes model, this model generally did not fit the data as well as the more parsimonious fixed effects model. Although SWD had generally similar growth trajectories in this sample, in other models with different TVCs that represent larger variability in intra-individual differences, the TVC–random slopes model might be more useful.

Given previous research documenting the movement of students within special education and between general and special education, the intra-individual TVC–fixed effects and random slopes models with SWD represented as TVC may represent the most intuitive specifications for characterizing participation in special education over time. In general, the estimated mean growth trajectories for the SWD categories were similar between these models (Figure 3); however, the estimated mean growth trajectories of the intra-individual TVC–fixed effects model yielded knots in the curve (Figure 2, lower left panel) with less achievement gain evident in the transition from elementary into middle school (Grades 5 to 6) for all SWD groups except students with Other disabilities. Less achievement gain was evident for all SWD groups from Grades 7 to 8 (the average SWOD trajectory appears smoothed with no knots). These knots were not evident in the intra-individual TVC–random slopes model as Figures 2 and 3 represented the average (across students) relation of SWD classification on the mathematics outcomes applied uniformly across time.

The TIC–Wave 1 model estimated growth trajectories for students with a specific disability in the initial year (Grade 3), whereas the TIC–always model estimated growth trajectories for SWD who remained in the same disability category at each of the six grades. Both TIC models explained the same amount of variance in the intercept, linear, and quadratic growth factors (Table 3). Both models specified SWD exceptionality categories as predictors of the latent growth trajectory factors. The TIC–Wave 1 produced higher estimated Grade 3 intercepts and higher estimated Grade 8 means than the TIC–always model (Figure 3) because the trajectories of the latter represented students in the same disability classification across time. As reported by recent research (Schulte & Stevens, 2015), lower estimated growth trajectories were expected for students always receiving special education services. The trajectories for both models were sometimes similar (ASD or SLD) and sometimes different (CD and ED).

The TIC–Wave 1 model regressed the latent growth trajectory factors on the Grade 3 SWD covariates, the same approach taken by most other researchers investigating student achievement growth (e.g., Schulte et al., 2016; Stevens et al., 2015; Wei et al., 2011; Wei et al., 2012). The results from the TIC–Wave 1 model were generally consistent with findings by Wei et al. (2012), who found intercepts for SWD groups to be statistically significantly lower than the SWOD groups. In addition, the results for the TIC–Wave 1 model did not yield any statistically significant SWD coefficients (compared with SWOD) of either linear or quadratic slope factors, which is generally similar to the findings by Wei et al. (2012) for math growth (except for a significant relation between Autism and linear change and the relation of Speech impairments with quadratic change for the Calculation outcome). However, Wei et al. (2012) modeled subdomain growth in mathematics, whereas we modeled growth based on the state's composite measure of mathematics.

The findings of the TIC–Wave 1 model were less consistent with findings by Stevens et al. (2015), despite the use of statewide math achievement tests in both studies. Stevens et al. reported statistically significant SWD relations with math at Grade 3 (intercept), similar to findings reported here. However, Stevens et al. also reported statistically significant relations of all student exceptionality groups except ED, Hearing Impairment, and Speech–language impairment, with linear change, and significant relations with quadratic change for students with ASD, Hearing Impairment, OHI, SLD, and Speech–language impairment. These studies differed, however, in the SWD exceptionality categories used, the student demographic covariates considered, the sample selection procedures, and sample sizes.

The TVC–fixed effects and random slopes models average estimated growth trajectories were spaced more closely than those estimated by the TIC–Wave 1 and TIC–always models. The observed mean of the sample at Grade 3 was best represented by the estimated mean intercepts of the student groups in the TIC–Wave 1 model; both TVC–fixed effects and TVC–random slopes models tended to overestimate the Grade 3 achievement scores for the SWD groups. Perhaps the best reasons for these estimated effects were that (a) estimated growth is removed, and (b) although there were disability reclassifications across years, there was not sufficient changing of groups over time for there to be variability across the repeated measures. Because the number of reclassified students was quite small compared with the total sample, the SWD classification covariates acted as TICs (i.e., invariant over time) but were regressed on the repeated measures (at "level 1" in a univariate framework like HLM).

In response to the third research question about model fit, our model fit indices (i.e., comparative fit index [CFI], TLI, SRMR, RMSEA) and information criteria (AIC, BIC, ABIC) were not consistent in suggesting a single best-fitting model. Depending on which index of model fit was used, either the intra-individual TVC–fixed effects model or the TIC–Wave 1 model best fit the data. Despite the added complexity, the intra-individual TVC–fixed effects model had superior values of AIC, RMSEA, and SRMR, and the TIC–Wave 1 model had superior values of TLI, BIC, and ABIC. Perhaps more importantly, the TIC–Wave 1 model yielded trajectories that were more representative of the data (Table 2) and the typical theoretical purpose of similar research—to describe and compare the academic growth trajectories of student subgroups, often for purposes of accountability. That is, the results of the TIC–Wave 1 model yielded a more meaningful, straightforward interpretation related to similar research and results that are likely more accessible for applied researchers and consumers. Furthermore, the intra-individual TVC–fixed effects model did not account for variance in the mathematics outcomes beyond either of the TIC models, indicating that they were not consequential predictors of mathematics scores beyond the growth trajectories. As discussed previously, due to the small proportion of reclassified students in our sample, the SWD exceptionality category covariates were generally more useful in accounting for variance between students than variance within students. Thus, despite the intuitive nature of modeling SWD as TVC, the application of SWD as TIC in this study was supported more by both theory and statistics.

Limitations

Our reported results may not generalize to different states (particularly those with different populations of SWD) and testing systems, different student cohorts, or analytic samples using different inclusion or exclusion rules. Specifically, several decisions were made related to sample selection that should be considered when reviewing results, such as excluding retained and accelerated students, and those who had incomplete data on SWD predictors across all 6 years (the latter of which reduced the sample by about 25%). Although students excluded from the analytic sample (e.g., due to attrition such as mobility) are likely missing at random (MAR) with respect to test scores, there may be non-random missingness associated with student demographics (e.g., SWD and economic disadvantage) that would result in some under-representation in the analytic sample and may influence the estimated model parameters (Zvoch & Stevens, 2005). In addition, because each model generally yielded different SWD coefficients, the comparison across models was partially based on the estimated mathematics outcome means at the particular measurement occasion. This basis of comparison is broad and perhaps overly simplistic. Moreover, some of our statistically non-significant results may be due to small sample size for some of the SWD exceptionality categories. Finally, to more directly demonstrate the modeling approaches for time-varying exceptionality classification, we chose to be parsimonious and not model the variance in achievement between schools or districts. Although this approach may be preferred, it would have required a cross-classified, multilevel model and distracted from our purpose.

Implications for the Improvement of Practice

We found that, although the average values of growth model parameters were quite similar across models, the TIC regression coefficients for student exceptionality categories were different from those of the TVC models (fixed effects and random slopes). In general, any of the models we studied might be preferred given different purposes, types of covariates, or data with different characteristics. Before choosing a particular model specification, researchers should consider how much variability exists on the covariate across measurement occasions. With low variability (i.e., little switching of covariate categories over time), it may be preferable and more parsimonious to include TVC such as SWD, English language learners, or FRL recipients as TIC. Future statistical research could explore potential guidelines on minimum or optimal predictor variance across time to inform modeling as a TVC or TIC. In our case, we found that the small proportion of reclassified students over time resulted in a situation in which models were probably better specified as accounting for variance between students rather than accounting for variance within students. When the purpose is to control for relations or multicollinearity between the TVC and growth model parameters, models specified as TVC may be the better choice. Applied researchers using TVCs should take care to understand the assumptions of the longitudinal modeling framework (univariate or multivariate; Curran et al., 2012), the distinction between within- and between-student effects, and the challenges in communicating results in accurate and accessible terms to the target audience and stakeholders. Accountability systems must include performance reports for student subgroups, including economically disadvantaged students, English learners, and SWD, all of which represent covariates that can and do vary across time. As these systems are now more focused on student growth, researchers and analysts must be mindful about the time-varying nature of these covariates and the longitudinal modeling decisions, and the implications of those decisions on results.

Authors' Note

The opinions expressed are those of the author and do not necessarily represent views of the Institute or the U.S. Department of Education.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by a Cooperative Service Agreement from the Institute of Education Sciences (IES) establishing the National Center on Assessment and Accountability for Special Education —NCAASE (Award Number R324C110004) at the University of Oregon.

Notes

1. Additional details of all analyses are available upon request from the corresponding author. A p value of .001 was used as the a priori decision rule for all null-hypothesis significance tests.
2. Note that there are no fit statistics for the intra-individual time-variant covariate (TVC)–random slopes model as chi-square and related fit statistics cannot be calculated because the variance of the repeated mathematics measures varies with the values of students with disabilities (SWD) predictors.

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7. Retrieved from http://CRAN.R-project.org/package=lme4

Carlson, E., & Parshall, L. (1996). Academic, social, and behavioral adjustment for students declassified from special education. *Exceptional Children*, *63*, 89–100.

Curran, P. J., Lee, T., Howard, A. L., Lane, S., & MacCallum, R. (2012). Disaggregating within-person and between-person effects in multilevel and structural equation growth models. In J. R. Harring & G. R. Hanccok (Eds.), *Advances in longitudinal methods in the social and behavioral sciences* (pp. 217–253). Charlotte, NC: Information Age.

Every Student Succeeds Act (ESSA). 114–95 § 1177 (2015). Retrieved from http://www.ed.gov/ESSA

Herring, W. L., McGrath, D. J., & Buckley, J. A. (2007). *Demographic and school characteristics of students receiving special education in the elementary grades*. Washington, DC: National Center for Education Statistics.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.

Marder, C. (2009). *Facts from SEELS: Perspectives on students' disability classifications*. Menlo Park, CA: SRI International. Retrieved from http://www.seels.net/info_reports/DisabilityClassif1.9.09.pdf

Morgan, P. L., Farkas, G., & Wu, Q. (2011). Kindergarten children's growth trajectories in reading and mathematics: Who falls increasingly behind? *Journal of Learning Disabilities*, *44*, 472–488.

Oregon Department of Education. (2007). *Technical report: Oregon's statewide assessment system test development* (Vol. 4: Reliability and Validity). Salem, OR: Author.

Oregon Department of Education. (2008). *Technical report: Oregon's statewide assessment system test development* (Vol. 3: Standard Setting). Salem, OR: Author.

Oregon Department of Education. (2010). *Technical report: Oregon's statewide assessment system test development* (Vol. 1: Annual Report). Salem, OR: Author.

Oregon Department of Education. (2012a). *Technical report: Oregon's statewide assessment system test development* (Vols. 1–10). Salem, OR: Author.

Oregon Department of Education. (2012b). *Technical report: Oregon's statewide assessment system test development* (Vol. 5: Test Administration). Salem, OR: Author.

Puranik, C. S., Petscher, Y., Al Otaiba, S., Catts, H. W., & Lonigan, C. J. (2008). Development of oral reading fluency in children with speech or language impairments: A growth curve analysis. *Journal of Learning Disabilities*, *41*, 545–560. doi:10.1177/0022219408317858

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & du Toit, M. (2011). *HLM 7: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.

Saven, J., Anderson, D., Nese, J. F. T., Farley, D., & Tindal, G. (2016). Patterns of statewide test participation for students with significant cognitive disabilities. *The Journal of Special Education*, *49*, 209–220.

Schulte, A. C., & Stevens, J. J. (2015). Once, sometimes, or always in special education: Mathematics growth and achievement gaps. *Exceptional Children*, *81*, 370–387.

Schulte, A. C., Stevens, J. J., Elliott, S. N., Tindal, J., & Nese, J. F. T. (2016). Achievement gaps for students with disabilities: Stable, widening, or narrowing on a state-wide reading comprehension test. *Journal of Educational Psychology*. Advance online publication. doi:10.1037/edu0000107

Shin, T., Davison, M. L., Long, J. D., Chan, C. K., & Heistad, D. (2013). Exploring gains in reading and mathematics achievement among regular and exceptional students using growth curve modeling. *Learning and Individual Differences*, *23*, 92–100.

Stevens, J. J., Nese, J. F. T., & Tindal, G. (2013). *Using longitudinal models to track student achievement: A literature synthesis*. Unpublished raw data.

Stevens, J. J., Schulte, A. C., Elliott, S. N., Nese, J. F. T., & Tindal, G. (2015). Growth and gaps in mathematics achievement of students with and without disabilities on a statewide achievement test. *Journal of School Psychology*, *53*, 45–62.

Thurlow, M. L., Wu, Y.-C., Lazarus, S. S., & Ysseldyke, J. E. (2016). Special education–non-special education achievement gap in math: Effects of reporting methods, analytical techniques, and reclassification. *Exceptionality*, *24*, 32–44.

Walker, D., Singer, J., Palfrey, J., Orza, M., Wenger, M., & Butler, J. (1988). Who leaves and who stays in special education: A 2-year follow-up study. *Exceptional Children*, *54*, 393–402.

Wei, X., Blackorby, J., & Schiller, E. (2011). Growth in reading achievement of students with disabilities, ages 7 to 17. *Exceptional Children*, *78*, 89–106.

Wei, X., Lenz, P. B., & Blackorby, J. (2012). Math growth trajectories of students with disabilities: Disability category, gender, racial, and socioeconomic status differences from ages 7 to 17. *Remedial and Special Education*, *34*, 154–165.

Ysseldyke, J., & Bielinski, J. (2002). Effect of different methods of reporting and reclassification on trends in test scores for students with disabilities. *Exceptional Children*, *68*, 189–200.

Zvoch, K., & Stevens, J. J. (2005). Sample exclusion and student attrition effects in the longitudinal study of middle school mathematics performance. *Educational Assessment*, *10*, 105–123.