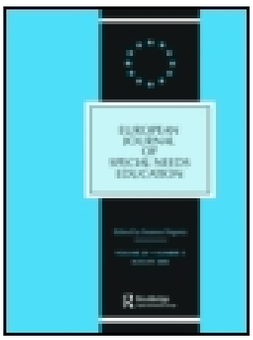

European Journal of Special Needs Education

ISSN: 0885-6257 (Print) 1469-591X (Online) Journal homepage: http://www.tandfonline.com/loi/rejs20

Effects of special education on academic achievement and task motivation: a propensity-score *and* fixed-effects approach

Marianne Nilsen Kvande, Oda Bjørklund, Stian Lydersen, Jay Belsky & Lars Wichstrøm

To cite this article: Marianne Nilsen Kvande, Oda Bjørklund, Stian Lydersen, Jay Belsky & Lars Wichstrøm (2018): Effects of special education on academic achievement and task motivation: a propensity-score *and* fixed-effects approach, European Journal of Special Needs Education, DOI: 10.1080/08856257.2018.1533095

To link to this article: https://doi.org/10.1080/08856257.2018.1533095

Published online: 12 Oct 2018.

EUROPEAN JOURNAL OF SPECIAL NEEDS EDUCATION
https://doi.org/10.1080/08856257.2018.1533095

Routledge
Taylor & Francis Group

ARTICLE

OPEN ACCESS Check for updates

Effects of special education on academic achievement and task motivation: a propensity-score *and* fixed-effects approach

Marianne Nilsen Kvande[a], Oda Bjørklund[b], Stian Lydersen[c], Jay Belsky[d] and Lars Wichstrøm[a,b,e]

[a]Human Development unit, NTNU Social Research, Trondheim, Norway; [b]Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway; [c]Regional Centre for Child and Youth Mental Health and Child Welfare, Norwegian University of Science and Technology, Trondheim, Norway; [d]Department of Human Ecology, University of California, Davis, CA, USA; [e]Department of Child and Adolescent Mental Health, St. Olav's Hospital, Trondheim, Norway

ABSTRACT

As traditional teaching methods may fail to serve children with special needs, special education (SE) services aim to compensate for the shortcomings of conventional schooling. However, despite of numerous studies on the effectiveness of SE services, the influence of potential selection bias remains a real challenge, and only a few studies have applied methodology aiming to surmount these shortcomings. Therefore, by combining two methods (i.e. propensity score and fixed effects regression) to account for potential confounders, we examined the effects of receiving SE services in first and third grades on Norwegian students' academic achievement and task motivation in third and fifth grades (n = 745). Thus, we controlled for a propensity score that was calculated based on observed selection into SE, and combined this with fixed effects regression that has the advantage of ruling out all time-invariant confounders (e.g. genetics). Results revealed that SE in third grade adversely affected math achievements in fifth grade, and SE had no effect on reading and writing achievements or task motivation for reading, writing and math. The efficacy of SE services is called into question, and potential explanations and solutions are explored.

ARTICLE HISTORY
Received 30 July 2018
Accepted 3 October 2018

KEYWORDS
Special education; education; elementary school; propensity scoring; fixed effects regression

Students with poor academic records or those who do not benefit sufficiently from standard teaching practices are subject to varying degrees of modifications and accommodations to their daily schooling. These initiatives and interventions, termed special education (SE), are implemented with the primary goal of helping children develop to their fullest potential, academically and socially. An annual total of $50 billion was spent on SE in the USA for the 1999–2000 academic year (Parrish et al. 2003) and in Norway, where we conducted this study, a sixth of the public education budget is used for SE (Norwegian Directorate of Education and Training 2013; Union of Education 2016). Here, we evaluate the effects of SE on children's academic achievement and task motivation,

CONTACT Marianne Nilsen Kvande marianne.n.kvande@ntnu.no

taking advantage of a large community study followed from first to fifth grade. Importantly, our work extends existing research on the efficacy of SE by more rigorously discounting alternative explanations of any detected SE effects.

Efficacy of special education

SE refers to a wide range of adaptions of everyday schooling including but not limited to alternative teaching methods, curricula, and learning goals; use of special equipment; small group or one-on-one teaching; personalised assistance for attention and memory; and provision of richer explanations of concepts while simplifying curricula. Some students are also enrolled in special classes or schools. Despite extensive research on SE, the efficacy of these well-intended yet costly efforts remains unclear. Not only are prospective designs infrequently used to evaluate SE's impact, the uncertain state of affairs is due to one important, understandable methodological shortcoming. It would be highly unethical – and in many countries unlawful – to deny a subset of children their SE needs for experimental purposes; therefore, investigators must rely on observational studies rather than an optimal randomised, well designed and controlled trial to evaluate SE. Towards reducing the potential impact of confounding factors, researchers typically adjust for a range of covariates in their regression-type models that might select students into SE in the first place (e.g. poverty, gender, intelligence, self-efficacy, school district), which likely affect the alleged outcome of SE.

A more robust way to negate the effects of potential confounders is propensity scoring (Austin 2011; Rosenbaum and Rubin 1983). This approach involves balancing or matching those who receive SE and those who do not by whichever measured factor predicts receipt of SE in that data set, thereby reducing or eliminating selection bias. A handful of studies reviewed below have evaluated the effects of SE using this statistical method (Dempsey and Valentine 2017; Dempsey, Valentine, and Colyvas 2016; Morgan et al. 2010; Sullivan and Field 2013; Lekhal 2017). Independent from the merits of this work, propensity scoring still presupposes that all confounding factors have been discounted, a condition that is unlikely to be met. For example, consider confounding factors that are unlikely to be included in most data sets, like the proclivity of teachers to instigate a process that results in SE assignment that also biases teaching and evaluating student performance; aspects of school climate that affect both the inclination to refer students for SE and students' achievement; or parental characteristics (e.g. parenting style, homelife stability, genetics) that influence both the likelihood of receiving SE and student learning. Failure to account for such unmeasured confounders can inflate, and also potentially deflate, any observed effects of SE. Typically, these hard-to-measure potential confounders are fully (e.g. genetics) or predominantly (e.g. personality, parenting) stable over time (Wichstrøm, Belsky, and Steinsbekk 2017a; Allison 2009; Wichstrøm et al. 2017b).

One statistical approach, fixed effects regression, has the advantage of ruling out all time-invariant confounders – even when they are unknown (Allison 2009; Firebaugh, Warner, and Massoglia 2013; Bollen and Brand 2010), thereby substantially reducing, if not entirely eliminating, the uncertainty of causal conclusions of observational research, including those using propensity scoring. In the current investigation, we break new ground in the study of SE impact – and potentially many other topics – by combining the benefits of fixed effects and propensity scoring, which adjusts for all *measured* time-invariant and time-

variant confounders as well as all *unmeasured* time-invariant confounders. We applied this novel approach to a powerful data set that assessed and documented academic achievement and task motivation in Norwegian students repeatedly for 5 years, from first to fifth grade.

Meta-studies reveal that SE programmes targeting children with specific learning disabilities prove highly effective, with standardised effect sizes in the range of .70–1.00 (Berkeley, Scruggs, and Mastropieri 2010; Scruggs et al. 2010), while programmes for children with behavioural or emotional disorders are generally less promising (Harrison et al. 2013). Because the relevant investigations examine the effect of specific programmes targeting specific difficulties in particular groups of students (e.g. fast-paced instruction among children with attention deficit hyperactivity disorder – ADHD), the results of SE efficacy studies may not generalise widely to eligible students in the regular school system. However, two longitudinal studies have asked whether SE services as conventionally delivered in the school system are effective – yielding evidence that SE increases skill in mathematics (Hanushek, Kain, and Rivkin 2002) and reading (Ehrhardt et al. 2013; Hanushek, Kain, and Rivkin 2002). Notably, using fixed effects methods as applied here, Hanushek, Kain, and Rivkin (2002) found that students with learning difficulties and/or emotional problems, and for both mainstreamed and non-mainstreamed students, SE improved academic achievements throughout elementary school. In the more narrow approach without applying fixed effects or propensity scores, Ehrhardt et al. (2013) reported on improvements in one skill (i.e. reading) for students diagnosed with reading disorder. By contrast to the two studies above, six longitudinal investigations (using propensity scoring) have found that SE either has no effect – or a negative one – on children's academic skills and psychosocial development (Morgan et al. 2010; Sullivan and Field 2013; Dempsey and Valentine 2017; Dempsey, Valentine, and Colyvas 2016; Keslair, Maurin, and McNally 2012; Lekhal 2017). These results call into question the effectiveness of SE services delivered in the American, Australian, and Norwegian school-systems, which motivated the research presented in this report.

Method

Procedure and sample

The Trondheim Early Secure Study (TESS) started in 2007 when the participating children were 4 years old. The work presented herein uses data from the second, third and fourth waves of data collection when the children were 6 (first grade), 8 (third grade) and 10 years old (fifth grade). The cohorts born in 2003 and 2004 and their parents living in Trondheim, Norway were invited to participate. The children were recruited at a community health check-up for 4-year-olds, which is a free service for all Norwegian children. A letter of invitation was sent to all parents (N = 3,456) prior to meeting at the well-child clinic. Of these, 3,358 (97%) met at the clinic. At the checkup, the health nurse informed about the study and written consent was obtained. Parents (n = 176) who lacked proficiency in the Norwegian language were excluded. The health nurses failed to ask 166 parents. A total of 2,475 of the 3,016 eligible parents consented. To increase variability and thus statistical power, children with emotional or behavioural problems were oversampled. Towards this end, parents completed the Strengths and Difficulties Questionnaire (SDQ) (Goodman

1997). The SDQ total difficulties scores were divided into four strata (cut offs: 0 – 4, 5 – 8, 9 – 11, 12 – 40). The higher the score on the SDQ, the more likely the child was to be drawn to participate in the study. The drawing probability increased with the SDQ scores of each of the four strata being 0.37, 0.48, 0.70 and 0.89, respectively. Details concerning the procedure and recruitment are further described in Wichstrøm et al. (2012). As a result of the procedures described above, 1,250 families were randomly drawn to participate, of which 936 (74.9%) were examined for the first wave. Those who dropped out at this point did not vary by SDQ strata ($\chi^2 = 5.70$, $df = 3$, $p = .13$) or gender ($\chi^2 = .23$, $df = 1$, $p = .63$). For the second wave 2 years later, 795 children (50.5% boys) participated in the follow-up assessment. Four and six years later, in the third and fourth waves, 699 and 702 children participated, respectively. In the second, third, and fourth waves, which are included in the present study, 781, 627 and 659 teachers participated, respectively, by providing information on SE. Attrition in waves three and four was not predicted by academic achievements or task motivation at waves two and three. Of students receiving SE at T1, 66% also received SE at T2 and 74% at T3. Of those who received SE at T2, 68% also did so at T3. All students are mainstreamed, and none are enrolled in special schools. SE students at T1 and T2 received educational services such as help from assistants/special teachers, alternative books, small groups and one-on-one teaching, or seeing a speech therapist regularly. Other sample characteristics at T1, the second wave, are provided in Table 1. The project was approved by the Regional Committee for Research Ethics, Mid-Norway.

Measures

Special education

Information on SE was provided by the primary teacher who was asked the following: 'has there been initiated any special services for the student such as remedial teaching, additional assistance, special class/school etc.?' Answers were coded (1) 'no' (2) 'yes.'

Academic performance

Formal grades are not given to Norwegian students before the eighth grade. Therefore, to assess the level of academic performance, the primary teacher rated the students' performance in reading, writing and math skills on a scale that ranges from (1) 'far below the class' mean performance' to (5) 'far above the class' mean performance.'

Task motivation; reading, writing and mathematics

Children's motivation for reading, writing and mathematics was assessed in the first (T1), third (T2) and fifth (T3) grades using the Task Value Scale for Children (Nurmi and Aunola 1999; see also Aunola, Leskinen, and Nurmi 2006; Nurmi and Aunola 2005). For each of the three subjects of reading, writing and mathematics, the children were asked three questions regarding their interest in each subject; 'How much do you like reading/writing/mathematics tasks?'; 'How much do you like doing reading/writing/mathematics tasks in school?'; How much do you like doing reading/writing/mathematics tasks at home?" The children reported their interest in a particular task on a scale ranging from (1) 'I do not like it at all' to (5) 'I like it very much.'

The three questions on each task were then summed for a total score that ranged from 3 to 15. For each time point, Cronbach's alphas where respectively .78, .88 and .88 for reading;

Table 1. Sample characteristics at T1.

	n	%
Gender of child	745	100
Male	363	48.7
Ethnic origin of biological mother	702	100
Norwegian	656	87.9
Ethnic origin of biological father	702	100
Norwegian	656	87.9
Child care when child was 5–6 years	483	100
Official daycare centre	433	89.6
Others	43	8.9
None	6	1.2
Biological parents' marital status	694	100
Married	419	60.4
Cohabitating >6 months	174	25.1
Separated	8	1.2
Divorced	76	11
Widowed	-	-
Cohabitating <6 months	5	0.7
Never lived together	12	1.7
Parental socio-economic status	650	100
Leader	99	15.2
Professional, higher level	226	34.8
Professional, lower level	219	33.7
Skilled workers	100	15.4
Farmers/fishermen	1	0.2
Unskilled workers	5	0.8
Mother's highest level of completed education	659	100
Junior high school (10th grade)	13	2
Senior high school (13th grade)	91	13.8
Some education after senior high school/or vocational (13th grade)	66	10
College degree	276	41.9
University degree	213	32.3
Father's highest level of completed education	655	100
Junior high school (10th grade)	30	4.6
Senior high school (13th grade)	86	13.1
Some education after senior high school/or vocational (13th grade)	123	18.8
College degree	202	30.8
University degree	214	32.7

T1 = first grade. $N = 745$.

.78, .91 and .91 for writing; and .81, .94 and .95 for mathematics. Task-motivation has been prospectively related to math performance (Aunola, Leskinen, and Nurmi 2006) and self-concepts of ability (Nurmi and Aunola 2005).

Potential confounders

The child's gender was coded (1) for a boy and (2) for a girl. Socio-economic status of the parents was coded according to the International Classifications of Occupations (International Labor Office 1990). When there were two parents, the parent with the highest-rated occupation was selected. Level of parental education was assessed ranging from (1) 'not completed junior high school' to (11) 'Ph.D. completed or ongoing', and the mean of parental education level was used.

Test scores (grade 1 and 3) were obtained from the Trondheim local Municipality offices. The Norwegian Directorate for Education and Training (2008) administers mandatory tests in reading (grade 1 to 3) and voluntary tests in math (grade 1 and 3) for all Norwegian students. In the first grade-reading test, the students performed tasks

including writing the letters of the alphabet, reading words and sentences. The scores of all tasks were summed and ranged from 0 to 105. The third grade reading test has four parts dealing with word chains, fiction and non-fiction reading comprehension, and vocabulary. The total score ranged from 0 to 102. The math mapping tests evaluate basic skills such as counting, sorting numbers by size, completing a series of numbers, and performing addition and subtraction. The scores from the numeracy test ranged from 0 to 50 in first grade, and 0 to 85 in third grade.

Based on a priori knowledge of important confounders of selection into SE (Kvande, Belsky, and Wichstrøm 2017; Hibel, Farkas, and Morgan 2010; Mann, McCartney, and Park 2007), we assessed the teacher's level of helplessness by asking the child's primary teacher in first and third grades to respond to the following question, with the answer coded on a five-point scale ranging from (1) 'not at all' to (5) 'very strongly': 'When you teach this student, to what degree do you feel helpless?'. To assess the students' ability to learn, the primary teacher were asked the following, 'Compared to other students of same age, how much is he/she learning?' for which answers range from (1) 'far below the class' mean' to (7) 'far above the class' mean'.

Intelligence was assessed using the two subtests of vocabulary and matrix reasoning of the Wechsler's Abbreviated Scales for Intelligence (WASI) (Wechsler 1999). Following the standard protocol for administration, the children orally defined different words in the vocabulary test, and completed gridded patterns in the matrix reasoning task. The scores of both tests were summed to yield a total score.

Symptoms of ADHD, oppositional defiant disorder (ODD), and conduct disorder (CD) were recorded using the semi-structured diagnostic interview-based Child and Adolescent Psychiatric Assessment (CAPA) (Angold and Jane Costello 2000) developed to assess mental disorders according to the Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) (American Psychiatric Association 2000). The child and parent were interviewed separately. The CAPA contains a structured protocol with mandatory questions and optional follow-up questions. A symptom is considered present if reported by either child or parent. The interviewers (n = 7) had at least a bachelor's degree in the relevant field and were trained by the CAPA team. Blinded raters recoded 15% of the interviews and the resulting intra-rater reliabilities between multiple raters were ICC = .90 for ADHD, ICC = .90 for ODD, and ICC = .85 for CD.

Statistical analysis

The analyses were performed in three steps: (1) *Propensity score modelling*. We calculated the propensity to receive SE in first and third grades by constructing two propensity score logistic regression models with the exposures (SE in first and third grades) as dependent variables and selected potential confounders as covariates. We used the log odds of exposure as propensity score. The following 10 variables served as potential confounders based on prior evidence of their importance (Kvande, Belsky, and Wichstrøm 2017; Hibel, Farkas, and Morgan 2010; Mann, McCartney, and Park 2007): child's gender, symptoms of ADHD, ODD/CD, test scores in reading and math, intelligence, ability to learn; parental socio-economic status and educational level; and teacher's sense of helplessness when teaching the child. In the propensity score modelling, missing values were handled by multiple imputation (MI), with 100 imputed data sets.

The MI model included all confounders used to calculate the propensity score, and all dependent variables (i.e. skills in reading, writing and math; motivation for reading, writing and math) included in the SEM-models in steps two and three. The mean log odds across the 100 data sets were calculated for each respondent.

(2) *Autoregressive models with propensity adjustment*. Due to the number of outcomes (6 x 2 time points) relative to the number of students, we were unable to include all outcomes in one model; we, therefore, analysed the impact of SE in six models, one for each outcome. Then, the dependent variable in question (e.g. math score) was regressed on SE, using structural equation modelling (SEM) controlling for previous level of the outcome, and the propensity score 2 years prior. We chose not to use propensity score matching (but rather control for the log odds of receiving SE) because there is no straightforward method for accomplishing this with weighted data. At each time point, the predictors were allowed to correlate.

(3) *Fixed effects regression with propensity adjustment*. Fixed effects were added to the models in step two, as described in Allison (2009) and Wichstrøm, Belsky, and Steinsbekk (2017a). A latent time-invariant factor was created by loading on the dependent variables in third and fifth grades; thus, the effect of SE was adjusted for all unmeasured time-invariant confounders and all measured confounders captured by the propensity score. To avoid negative degrees of freedom and thus identify the models, we constrained the autoregressive paths, within-time correlations, and the impact of the propensity score on the outcome to be similar over time. To identify models, the regression paths between SE and the propensity score were constrained to be equal across time points. Also, in each of the six models, the latent time-invariant factor was allowed to correlate with the propensity score and exposure to SE. The propensity score was calculated using SPSS version 24.0 (IBM 2016), and SEM models (steps two and three) were calculated using Mplus version 7.4 (Muthèn and Muthèn 1998–2015).

Because we oversampled children with mental health problems, the data were weighted back with a factor corresponding to the number of children in the stratum divided by the number of participating children. We used a robust maximum likelihood estimator which provides robust standard errors and is robust to deviations from normality and missing data were handled according to a full information maximum likelihood procedure.

Results

Descriptives and propensity scores

The number of students receiving SE increased from first to fifth grades (Table 2). SE students performed more poorly and with less motivation than their peers in reading, writing and math. This pattern was found across all grades, except for motivation for writing in third grade and motivation for math in fifth grade. SE students had higher log odds for receiving SE in both first and third grades (Table 3). There was a moderate overlap in log odds between students who received SE and those who did not.

Table 2. Descriptive statistics for study variables among special education and no special education students in 1st, 3rd, and 5th grade.

	M (SD)	M (SD)	95% CI for difference	Estimated mean difference	*t*	*p*-value
1st grade variables	**SE in 1st grade (*n* = 85)**	**No SE in 1st grade (*n* = 696)**				
Reading skills	2.16 (.87)	3.44 (.82)	1.09 to 1.46	1.27	13.34	<.001
Writing skills	2.09 (.78)	3.36 (.77)	1.09 to 1.44	1.26	14.19	<.001
Math skills	2.56 (.87)	3.43 (.71)	.67 to 1.06	.86	8.82	<.001
Reading motivation	10.40 (4.22)	11.43 (3.52)	−.07 to 2.12	1.03	1.87	.066
Writing motivation	10.49 (4.15)	11.72 (3.42)	.13 to 2.32	1.23	2.24	.009
Math motivation	10.45 (4.44)	11.71 (3.61)	.10 to 2.42	1.26	2.17	.034
3rd grade variables	**SE in 3rd grade (*n* = 103)**	**No SE in 3rd grade (*n* = 524)**				
Reading skills	2.38 (1.01)	3.58 (89)	1.00 to 1.41	1.20	11.45	<.001
Writing skills	2.34 (.91)	3.41 (.82)	.88 to 1.25	1.07	11.13	<.001
Math skills	2.66 (1.04)	3.51 (.83)	.63 to 1.09	.86	7.41	<.001
Reading motivation	11.18 (3.87)	12.17 (3.05)	.13 to 1.84	.99	2.29	.024
Writing motivation	10.87 (4.02)	11.52 (3.35)	−.25 to 1.54	.65	1.44	.154
Math motivation	10.98 (4.49)	12.04 (3.46)	.07 to 2.05	1.06	2.12	.036
5th grade variables	**SE in 5th grade (*n* = 115)**	**No SE in 5th grade (*n* = 544)**				
Reading skills	2.17 (.86)	3.57 (.87)	1.20 to 1.58	1.39	14.43	<.001
Writing skills	2.04 (.76)	3.44 (.83)	1.23 to 1.57	1.40	16.19	<.001
Math skills	2.38 (.88)	3.51 (.86)	.94 to 1.32	1.13	11.84	<.001
Reading motivation	10.11 (3.65)	11.92 (2.81)	1.04 to 2.57	1.81	4.68	<.001
Writing motivation	10.23 (3.24)	11.34 (3.40)	.41 to 1.79	1.10	3.14	.001
Math motivation	10.60 (3.98)	11.34 (3.40)	−.10 to 1.59	.74	1.75	.053

Independent samples t-tests were calculated to test for significant differences of the means.

Table 3. Propensity scores: Log odds for receiving special education versus no special education in 1st and 3rd grade.

		Log odds			
	N	Mean	SD	Minimum	Maximum
SE 1st grade	64	−.34	2.09	−4.29	4.73
No SE 1st grade	589	−3.70	1.32	−6.20	1.39
SE 3rd grade	103	−.34	1.61	−4.58	3.87
No SE 3rd grade	524	−2.65	1.27	−5.98	2.18

The calculation of the respective Log odds is based on the following variables: child's gender, symptoms of ADHD, ODD/CD, test scores in reading and numeracy, intelligence, ability to learn; Parental SES and occupational level/type; Teacher's level of helplessness when teaching the child.

Propensity score analysis

The results from the autoregressive regression model controlling for the propensity score showed that SE in first grade predicted *higher* math skills in third grade (Table 4). SE in third grade, however, predicted *lower* skills in reading and writing in fifth grade. SE did not predict task motivation at any of the time points.

Propensity score and latent fixed effects analysis

When we controlled for both observed time-varying and time-invariant confounders (i.e. propensity score) and unobserved time-invariant confounders (i.e. latent fixed effects) in the regression models, results indicated that first grade SE no longer predicted academic skills or task motivation in third grade, while SE in third grade predicted *reduced* math

Table 4. Estimated effects of special education on academic performance and task motivation adjusted for the propensity for special education – without and with adjustment for all time-invariant confounders.

	1st grade → 3rd grade			3rd grade → 5th grade		
	B	95% CI	*p*-value	B	95% CI	*p*-value
M1: Adjusted for Log odds for SE						
a) SE → Reading skills	.09	−.18 to .36	.513	−.33	−.53 to −.13	.001
b) SE → Writing skills	−.07	−.31 to .16	.548	−.37	−.58 to −.17	<.001
c) SE → Math skills	.34	.10 to .58	.005	−.15	−.36 to .06	.169
d) SE → Reading motivation	.08	−1.05 to 1.22	.887	−.67	−1.48 to .14	.107
e) SE → Writing motivation	−.33	−1.44 to .79	.565	−.21	−1.00 to .58	.601
f) SE → Math motivation	.65	−.71 to 2.02	.348	−.02	−.89 to .86	.974
M2: Adjusted for Log odds for SE and fixed effects						
a) SE → Reading skills	.05	−.46 to 0.55	.851	.59	−.01 to 1.19	.055
b) SE → Writing skills	−.26	−.58 to .07	.123	−.33	−.67 to .02	.061
c) SE → Math skills	.09	−.27 to .44	.640	−.37	−.71 to −.02	.036
d) SE → Reading motivation	.24	−1.00 to 1.47	.709	−.77	−2.22 to .69	.302
e) SE → Writing motivation	.77	−.98 to 2.53	.388	−.13	−1.70 to 1.44	.873
f) SE → Math motivation	−.06	−.169 to .059	.346	−.06	−.169 to .059	.346

Table 5. Model fit results of the estimated effects of special education on academic performance and task motivation adjusted for the propensity for special education – without and with adjustment for all time-invariant confounders.

	χ^2	*df*	*p*	RMSEA[a] (95% CI)	SRMR[b]	CFI [c]	TLI [d]
M1: Adjusted for Log odds for SE							
a) SE → Reading skills	27.678	8	<.001	.057 (.035 to .082)	.031	.987	.967
b) SE → Writing skills	28.953	8	<.001	.059 (.037 to .083)	.028	.986	.965
c) SE → Math skills	30.534	7	<.001	.067 (.044 to .092)	.033	.984	.956
d) SE → Reading motivation	34.834	9	<.001	.062 (.041 to .084)	.051	.973	.940
e) SE → Writing motivation	25.138	8	<.001	.055 (.033 to .078)	.045	.980	.955
f) SE → Math motivation	38.044	8	<.001	.071 (.049 to .094)	.051	.967	.918
M2: Adjusted for Log odds for SE and fixed effects							
a) SE → Reading skills	9.115	2	.011	.069 (.028 to .117)	.020	.995	953
b) SE → Writing skills	10.232	2	.006	.074 (.034 to .122)	.022	.995	.945
c) SE → Math skills	25.386	4	<.001	.085 (.055 to .118)	.045	.985	.927
d) SE → Reading motivation	12.226	3	.007	.064 (.030 to .104)	.019	.990	.936
e) SE → Writing motivation	11.681	3	.009	.062 (.028 to .102)	.018	.991	.942
f) SE → Math motivation	11.829	3	.008	.063 (.028 to .102)	.018	.990	.936

[a]Root mean square error of approximation; [b]Standardised root mean square residual; [c]Comparative fit index; [d]Tucker Lewis index.

skills in fifth grade (Table 4). No further effects of SE were detected. SE, thus, did not appear to enhance or worsen children's academic skills in reading and writing or their motivation for reading, writing and math. All six models showed acceptable fit to the data (Table 5).

Discussion

The purpose of this research was to evaluate the effectiveness of SE on students' academic achievements and task motivation using three waves of data collected from first to fifth grade on a large community sample of Norwegian children. When controlling for only the propensity score, first grade SE appeared to positively affect children's (increased) math skills in third grade, but adversely influence children's (reduced) skills in

reading and writing from third to fifth grade. When controlling for time-invariant confounders, the initially apparent beneficial effects of SE from first to third grade disappeared, and SE services in third grade adversely affected math skills in fifth grade.

These negative findings are in line with the few studies on SE that pay greater attention to the problems of potential selection bias (Morgan et al. 2010; Sullivan and Field 2013; Dempsey and Valentine 2017; Dempsey, Valentine, and Colyvas 2016; Keslair, Maurin, and McNally 2012; Lekhal 2017). The study presented here extends these efforts by implementing even more rigorous controls for potential selection bias through consideration of unobserved time-invariant factors that confound outcomes of SE and application of a propensity-score-based approach. Collectively, these investigations challenge claims that SE enhances academic and motivational performance, and raise questions for educators and policymakers. This study is hopefully a first step towards preventing the inefficient use of shared economic resources and improving the educational outcome for children; the latter is especially important because a lack of basic academic skills is associated with development of problem behaviours and welfare dependency (Frønes 2016).

The lack of evidence of positive effects of SE in countries that differ substantially in their schooling systems and approaches to SE (e.g. Norway, Australia, the USA) indicates that more universal factors inherent to providing SE may be at work. Although we were not able to address why SE services lack efficacy in the present study, there are multiple indications that the quality of SE provided in elementary schools is limited and several important, related factors may be involved. A lack of high-quality teachers in SE is commonly reported across nations (Norwegian Directorate for Education and Training 2016; Nordahl and Hausstätter 2009; McLeskey and Billingsley 2008; Thomas 2012). In the USA, this paucity has been described as severe, chronic and pervasive and has been on-going since the 1980s (Boe and Cook 2006). Similar problems have been reported in Australia and Norway (Thomas 2012; Nordahl and Hausstätter 2009; Norwegian Directorate for Education and Training 2016). In the current investigation, SE had a negative influence on the math performance of students from third to fifth grade. This may be related to unintended consequence of SE to 'water down' the curriculum (Harrison et al. 2013). For example, rather than providing the instruction needed to improve a math skill, a student may be provided with tasks or books that are too easy and may appear to benefit the student at the time – but in the long run may actually be detrimental to further academic development (Harrison et al. 2013). The tendency of teachers, parents and the student to hold lower academic expectations may be at play here and may impede the students' ability to access and learn the curriculum in regular schooling (McCoy et al. 2016).

Another factor related to the null or negative outcomes of SE could be that research-based knowledge on SE is often not applied (Boardman et al. 2005; Carter, Stephenson, and Strnadová 2012; Hausstätter and Thuen 2014). This is especially disconcerting as we have knowledge of which programmes are the most effective for at least some groups of children with learning (Berkeley, Scruggs, and Mastropieri 2010; Scruggs et al. 2010) and behaviour challenges (Harrison et al. 2013). Notably, the evidence on effective interventions is stronger for children with specific learning problems; medium-to-high effect sizes are found for interventions providing systematic, explicit instruction; learning strategies; spatial organisers and study aids; mnemonic strategies; hands-on activities;

peer mediation; and computer-assisted instruction for children with specific learning difficulties (Scruggs et al. 2010). Children with emotional and behavioural disorders (EBD) and ADHD may benefit from being able to choose between different activities, added interest (i.e. matching academic tasks with the students interests), adaptive furniture (i.e. use of therapy balls as chairs), opportunity to respond (i.e. providing opportunities to actively respond to questions in the class), fast-paced instruction, teacher proximity, shortened task length, and small-group instruction (Harrison et al. 2013). However, studies on interventions tailored for EBD and ADHD children are limited in number, have few participants, and fail to provide information on effect sizes for the outcomes measured (Harrison et al. 2013).

To provide effective SE services, it seems clear that two things must occur. First, more attention and weight must be paid to services for which we have evidence of efficacy. Second, we need more studies on what works for whom, particularly for children with EBD and ADHD, who comprise a considerable proportion of SE students and struggle the most with social adjustment and academic achievement (Nordahl and Sunnevåg 2008). Admittedly, results from efficacy studies targeting narrow groups of students may not be that informative to the overall practice of SE in regular schools. Nevertheless, even if this knowledge is informative in some situations for some groups of students, it may be a challenge for schools to ensure a broad enough range of different teachers with specialised training to educate a highly heterogeneous group of SE students. One possibility could be for schools to have continuing education programmes for their teachers that kept them abreast of and fluent in new research and specialised teaching techniques.

Strengths and limitations

A clear strength of the current study is the rigorous methodology employed, controlling for both the propensity score and unobserved time-invariant confounders. Additional strengths include the large sample size of students and the duration for which we have data on their performance (5 years). In regards to limitations, first, although our sample was relatively large, it did not include distinctions between specific forms of SE services for children with differing disabilities (e.g. ADHD, dyslexia, etc.). It is important to highlight that although we failed to detect any positive effects of SE services when our two-pronged effort to discount selection effects was implemented, it remains possible that significant benefits occurred for some children and positive effects for some specific forms of service do exist that are masked by the heterogeneity in SE's student population. Future studies should focus on addressing these potentially differential benefits of SE services, ideally utilising the rigorous methodology described and implemented herein.

Second, although our findings did not illuminate any positive effects of SE services on academic achievements or task motivation from first to fifth grade under the most rigorous empirical evaluation, we cannot rule out the possibility that SE services may have effects after fifth grade or in domains other than those examined herein (e.g. behaviour problems, language proficiency, self-efficacy, self-esteem).

Finally, academic performance was measured by teacher rating only. Ideally, to overcome the potential bias of subjective teacher evaluations of students' academic achievements, we would have included the results of standardised norm-referenced tests or formal grades if such data had been available.

Conclusions

The findings reported here suggest that students who receive SE services in early elementary school are not better off in terms of their academic achievements or task motivation compared with if they had not received such services – and, in one fact, may be adversely affected by their SE experience. This study adds to the limited body of research that attempts to fully take into account the non-random selection into SE. Notably, we have extended these efforts by, in addition to controlling for the propensity score, discounting effects of unobserved time-invariant factors. Future studies should aim to determine whether different SE initiatives are more helpful or harmful to specific groups of students by differentiating between the types of SE intervention, and the special needs of the student. Such studies should provide teachers and policymakers with important information on which to base the planning of future SE services. Additionally, our findings and those from similarly rigorous studies provide strong grounds for questioning the results of existing meta-analyses of SE efficacy and the quality of SE services.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Marianne Nilsen Kvande http://orcid.org/0000-0001-8101-3759
Jay Belsky http://orcid.org/0000-0003-2191-2503
Lars Wichstrøm http://orcid.org/0000-0003-3199-4637

References

Allison, P. D. 2009. *Fixed Effects Regression Models*. Vol. 160. Thousand Oaks, CA: SAGE publications.

American Psychiatric Association. 2000. *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR*. 4th ed. Washington, DC: American Psychiatric Association.

Angold, A., and E. Jane Costello. 2000. "The Child and Adolescent Psychiatric Assessment (CAPA)." *Journal of the American Academy of Child & Adolescent Psychiatry* 39 (1): 39–48. doi:10.1097/00004583-200001000-00015.

Aunola, K., E. Leskinen, and J. E. Nurmi. 2006. "Developmental Dynamics between Mathematical Performance, Task Motivation, and Teachers' Goals during the Transition to Primary School." *The British Journal of Educational Psychology* 76 (Pt 1): 21–40. doi:10.1348/000709905X51608.

Austin, P. C. 2011. "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies." *Multivariate Behavioral Research* 46 (3): 399–424. doi:10.1080/00273171.2011.568786.

Berkeley, S., T. E. Scruggs, and M. A. Mastropieri. 2010. "Reading Comprehension Instruction for Students with Learning Disabilities, 1995—2006: A Meta-Analysis." *Remedial and Special Education* 31 (6): 423–436. doi:10.1177/0741932509355988.

Boardman, A. G., M. E. Argüelles, S. Vaughn, M. T. Hughes, and J. Klingner. 2005. "Special Education Teachers' Views of Research-Based Practices." *The Journal of Special Education* 39 (3): 168–180. doi:10.1177/00224669050390030401.

Boe, E. E., and L. H. Cook. 2006. "The Chronic and Increasing Shortage of Fully Certified Teachers in Special and General Education." *Exceptional Children* 72 (4): 443–460. doi:10.1177/001440290607200404.

Bollen, K. A., and J. E. Brand. 2010. "A General Panel Model with Random and Fixed Effects: A Structural Equations Approach." *Social Forces; a Scientific Medium of Social Study and Interpretation* 89 (1): 1–34. doi:10.1353/sof.2010.0072.

Carter, M., J. Stephenson, and I. Strnadová. 2012. "Reported Prevalence by Australian Special Educators of Evidence-Based Instructional Practices." *Australasian Journal of Special Education* 35 (1): 47–60. doi:10.1375/ajse.35.1.47.

Dempsey, I., and M. Valentine. 2017. "Special Education Outcomes and Young Australian School Students: A Propensity Score Analysis Replication." *Australasian Journal of Special Education* 41 (1): 68–86. doi:10.1017/jse.2017.1.

Dempsey, I., M. Valentine, and K. Colyvas. 2016. "The Effects of Special Education Support on Young Australian School Students." *International Journal of Disability, Development and Education* 63 (3): 271–292. doi:10.1080/1034912X.2015.1091066.

Ehrhardt, J., N. Huntington, J. Molino, and W. Barbaresi. 2013. "Special Education and Later Academic Achievement." *Journal of Developmental and Behavioral Pediatrics* 34 (2): 111–119. doi:10.1097/DBP.0b013e31827df53f.

Firebaugh, G., C. Warner, and M. Massoglia. 2013. "Fixed Effects, Random Effects, and Hybrid Models for Causal Analysis." In *Handbook of Causal Analysis for Social Research*, edited by S. L. Morgan, 113–132. Dordrecht: Springer.

Frønes, I. 2016. "The Absence of Failure: Children at Risk in the Knowledge Based Economy." *Child Indicators Research* 9 (1): 247–260. doi:10.1007/s12187-015-9309-3.

Goodman, R. 1997. "The Strengths and Difficulties Questionnaire: A Research Note." *Journal of Child Psychology and Psychiatry and Allied Disciplines* 38 (5): 581–586. doi:10.1111/jcpp.1997.38.issue-5.

Hanushek, E. A., J. F. Kain, and S. G. Rivkin. 2002. "Inferring Program Effects for Special Populations: Does Special Education Raise Achievement for Students with Disabilities?" *Review of Economics and Statistics* 84 (4): 584–599. doi:10.1162/003465302760556431.

Harrison, J. R., N. Bunford, S. W. Evans, and J. S. Owens. 2013. "Educational Accommodations for Students with Behavioral Challenges: A Systematic Review of the Literature." *Review of Educational Research* 83 (4): 551–597. doi:10.3102/0034654313497517.

Hausstätter, R. S., and H. Thuen. 2014. "Special Education Today in Norway." In A. F. Rotatori, J. P. Bakken, F. E. Obiakor,S. Burkhardt, U. Sharma (Eds.), *Special Education International Perspectives: Practices across the Globe*, 181–207. Bingley, UK: Emerald Group Publishing Limited.

Hibel, J., G. Farkas, and P. L. Morgan. 2010. "Who Is Placed into Special Education?" *Sociology of Education* 83 (4): 312–332. doi:10.1177/0038040710383518.

IBM. 2016. *IBM SPSS Statistics for Windows, Version 24.0*. Armonk, NY: IBM Corporation.

International Labor Office. 1990. "International Standard Classification of Occupations: ISCO-88." International Labour Office. Accessed January 5. http://www.ilo.org/public/english/bureau/stat/isco/isco88/

Keslair, F., E. Maurin, and S. McNally. 2012. "Every Child Matters? An Evaluation of "Special Educational Needs" Programmes in England." *Economics of Education Review* 31 (6): 932–948. doi:10.1016/j.econedurev.2012.06.005.

Kvande, M. N., J. Belsky, and L. Wichstrøm. 2017. "Selection for Special Education Services: The Role of Gender and Socio-Economic Status." *European Journal of Special Needs Education* 1–15. doi:10.1080/08856257.2017.1373493.

Lekhal, R. 2017. "Does Special Education Predict Students' Math and Language Skills?" *European Journal of Special Needs Education* 1–16. doi:10.1080/08856257.2017.1373494.

Mann, E. A., K. McCartney, and J. M. Park. 2007. "Preschool Predictors of the Need for Early Remedial and Special Education Services." *The Elementary School Journal* 107 (3): 273–285. doi:10.1086/511707.

McCoy, S., B. Maître, D. Watson, and J. Banks. 2016. "The Role of Parental Expectations in Understanding Social and Academic Well-Being among Children with Disabilities in Ireland." *European Journal of Special Needs Education* 31 (4): 535–552. doi:10.1080/08856257.2016.1199607.

McLeskey, J., and B. S. Billingsley. 2008. "How Does the Quality and Stability of the Teaching Force Influence the Research-to-Practice Gap?: A Perspective on the Teacher Shortage in

Special Education." *Remedial and Special Education* 29 (5): 293–305. doi:10.1177/0741932507312010.

Morgan, P. L., M. L. Frisco, G. Farkas, and J. Hibel. 2010. "A Propensity Score Matching Analysis of the Effects of Special Education Services." *The Journal of Special Education* 43 (4): 236–254. doi:10.1177/0022466908323007.

Muthèn, L. K., and B. O. Muthèn. 1998–2015. *Mplus User's Guide*. 6th ed. Los Angeles: Muthen & Muthen.

Nordahl, T., and A. K. Sunnevåg. 2008. *Spesialundervisningen i grunnskolen: stor avstand mellom idealer og realiteter* [The Ideal versus the Reality of Special Education in Elementary School]. Hedmark, Norway: Hedmark University of Applied Sciences.

Nordahl, T., and R. S. Hausstätter. 2009. *Spesialundervisningens forutsetninger, innsatser og resultater. Situasjonen til elever med særlige behov under Kunnskapsløftet* [Special education under the knowledge promotion reform]. Hamar: Høgskolen i Hedmark/Utdanningsdirektoratet.

Norwegian Directorate for Education and Training. 2016. *The Education Mirror 2016 - Analysis of Primary and Secondary Education and Training in Norway*. Oslo: Norwegian Directorate for Education and Training.

Norwegian Directorate of Education and Training. 2013. "Utdanninsspeilet 2013 [The Education Mirror 2013]." Accessed February 16. http://www.udir.no/Tilstand/Utdanningsspeilet/Utdanningsspeilet/Utdanningsspeilet-2013/2-Ressurser/21-Kommunene-brukte-59-milliarder-kroner-pa-grunnskolen/

Nurmi, J.-E., and K. Aunola. 1999. *Jyväskylä Entrance Into Primary School Study (JEPS)*. Jyväskylä, Finland: University of Jyväskylä.

Nurmi, J.-E., and K. Aunola. 2005. "Task-Motivation during the First School Years: A Person-Oriented Approach to Longitudinal Data." *Learning and Instruction* 15 (2): 103–122. doi:10.1016/j.learninstruc.2005.04.009.

Parrish, T., J. Harr, J. Anthony, A. Merickel, and P. Esra. 2003. *State Special Education Finance Systems, 1999–2000: Part II: Special Education Revenues and Expenditures*. Palo Alto, CA: Center for Special Education Finance.

Rosenbaum, P. R., and D. B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (1): 41–55. doi:10.2307/2335942.

Scruggs, T. E., M. A. Mastropieri, S. Berkeley, and J. E. Graetz. 2010. "Do Special Education Interventions Improve Learning of Secondary Content? A Meta-Analysis." *Remedial and Special Education* 31 (6): 437–449. doi:doi:10.1177/0741932508327465.

Sullivan, A. L., and S. Field. 2013. "Do Preschool Special Education Services Make A Difference in Kindergarten Reading and Mathematics Skills?: A Propensity Score Weighting Analysis." *Journal of School Psychology* 51 (2): 243–260. doi:10.1016/j.jsp.2012.12.004.

The Norwegian Directorate for Education and Training. 2008. "Our Responsibilities." The Norwegian Directorate for Education and Training. Accessed May 13. http://www.udir.no/Stottemeny/English/A-brief-introduction-to-the-Norwegian-Directorate-for-Education-and-Training/

Thomas, T. 2012. "The Age and Qualifications of Special Education Staff in Australia." *Australasian Journal of Special Education* 33 (2): 109–116. doi:10.1375/ajse.33.2.109.

Union of Education. 2016. "Spesialundervisning [Special Education]." Accessed February 16. https://www.utdanningsforbundet.no/Hovedmeny/Vi-mener/Spesialundervisning-og-tidlig-innsats-ma-forsterkes/

Wechsler, D. 1999. *Manual for the Wechsler Abbreviated Intelligence Scale (WASI)*. San Antonio, TX: Psychological Corporation.

Wichstrøm, L., E. Penelo, K. Rensvik Viddal, N. de la Osa, and L. Ezpeleta. 2017b. "Explaining the Relationship between Temperament and Symptoms of Psychiatric Disorders from Preschool to Middle Childhood: Hybrid Fixed and Random Effects Models of Norwegian and Spanish Children." *Journal of Child Psychology and Psychiatry, and Allied Disciplines*. doi:10.1111/jcpp.12772.

Wichstrøm, L., J. Belsky, and S. Steinsbekk. 2017a. "Homotypic and Heterotypic Continuity of Symptoms of Psychiatric Disorders from Age 4 to 10 Years: A Dynamic Panel Model." *Journal of Child Psychology and Psychiatry* 58 (11): 1239–1247. doi:10.1111/jcpp.12754.

Wichstrøm, L., T. S. Berg-Nielsen, A. Angold, H. L. Egger, E. Solheim, and T. H. Sveen. 2012. "Prevalence of Psychiatric Disorders in Preschoolers." *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 53 (6): 695–705. doi:10.1111/j.1469-7610.2011.02514.x.