PROOF COVER SHEET

Journal acronym: CHAS
Author(s): Michael S. Matthews
Article title: Effects of schoolwide cluster grouping and within-class ability grouping on elementary school students' academic achievement growth
Article no: 846251
Enclosures: 1) Query sheet
2) Article proofs

Dear Author,

1. Please check these proofs carefully. It is the responsibility of the corresponding author to check these and approve or amend them. A second proof is not normally provided. Taylor & Francis cannot be held responsible for uncorrected errors, even if introduced during the production process. Once your corrections have been added to the article, it will be considered ready for publication.

Please limit changes at this stage to the correction of errors. You should not make insignificant changes, improve prose style, add new material, or delete existing material at this stage. Making a large number of small, non-essential corrections can lead to errors being introduced. We therefore reserve the right not to make such corrections.

For detailed guidance on how to check your proofs, please see
http://journalauthors.tandf.co.uk/production/checkingproofs.asp

2. Please review the table of contributors below and confirm that the first and last names are structured correctly and that the authors are listed in the correct order of contribution. This check is to ensure that your name will appear correctly online and when the article is indexed.

Sequence	Prefix	Given name(s)	Surname	Suffix
1		Michael S.	Matthews	
2		Jennifer A.	Ritchotte	
3		Matthew T.	McBee	

Queries are marked in the margins of the proofs.

AUTHOR QUERIES

General query: You have warranted that you have secured the necessary written permission from the appropriate copyright owner for the reproduction of any text, illustration, or other material in your article. (Please see http://journalauthors.tandf.co.uk/preparation/permission.asp.) Please check that any required acknowledgements have been included to reflect this.

AQ1	Please check and confirm the affiliation details for author 'Ritchotte'.
AQ2	Please check that the heading levels have been correctly formatted throughout.
AQ3	The reference citation 'Slavin (1993)' has been changed to 'Slavin (1987)' to match the author date in the reference list. Please check here and in subsequent occurrences, and correct if necessary.
AQ4	The reference citation 'Feldhusen and Moon (1992)' has been changed to 'Feldhusen and Moon (2004)' to match the author date in the reference list. Please check here and in subsequent occurrences, and correct if necessary.
AQ5	The reference 'Pierce et al. (2010)' is cited in text but not provided in the reference list. Please provide the full reference details or delete it from the citation.
AQ6	Please provide editors name for reference 'Feldhusen and Moon (2004)'.
AQ7	The reference 'Field (2009)' is listed in the references list but is not cited in the text. Please either cite the reference or remove it from the references list.
AQ8	Please provide significance of * in 'Table 6'.

How to make corrections to your proofs using Adobe Acrobat

Taylor & Francis now offer you a choice of options to help you make corrections to your proofs. Your PDF proof file has been enabled so that you can edit the proof directly using Adobe Acrobat. This is the simplest and best way for you to ensure that your corrections will be incorporated. If you wish to do this, please follow these instructions:

1. Save the file to your hard disk.

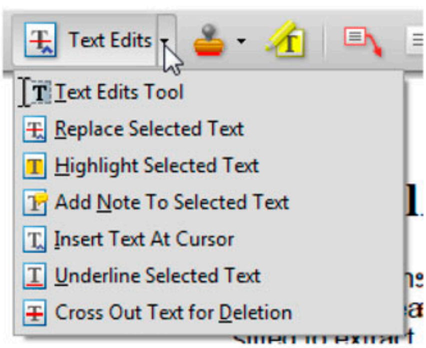
2. Check which version of Adobe Acrobat you have on your computer. You can do this by clicking on the "Help" tab, and then "About."

If Adobe Reader is not installed, you can get the latest version free from http://get.adobe.com/reader/.

- If you have Adobe Reader 8 (or a later version), go to "Tools"/ "Comments & Markup"/ "Show Comments & Markup."

- If you have Acrobat Professional 7, go to "Tools"/ "Commenting"/ "Show Commenting Toolbar."

3. Click "Text Edits." You can then select any text and delete it, replace it, or insert new text as you need to. If you need to include new sections of text, it is also possible to add a comment to the proofs. To do this, use the Sticky Note tool in the task bar. Please also see our FAQs here: http://journalauthors.tandf.co.uk/production/index.asp.



4. Make sure that you save the file when you close the document before uploading it to CATS using the "Upload File" button on the online correction form. A full list of the comments and edits you have made can be viewed by clicking on the "Comments" tab in the bottom left-hand corner of the PDF.

If you prefer, you can make your corrections using the CATS online correction form.

High Ability Studies, 2013
http://dx.doi.org/10.1080/13598139.2013.846251

Effects of schoolwide cluster grouping and within-class ability grouping on elementary school students' academic achievement growth

Michael S. Matthews[a]*, Jennifer A. Ritchotte[b] and Matthew T. McBee[c]

[a]*Department of Special Education and Child Development, University of North Carolina at Charlotte, Charlotte, NC, USA;* [b]*Special Education, University of Northern Colorado, Greeley, CO, USA;* [c]*Department of Psychology, East Tennessee State University, Johnson City, TN, USA*

(*Received 3 October 2012; final version received 16 September 2013*)

We evaluated the effects of one year of schoolwide cluster grouping on the academic achievement growth of gifted and non-identified elementary students using a piecewise multilevel growth model. Scores from 186 non-identified and 68 gifted students' Measures of Academic Progress Reading and Math scores were examined over three school years. In 2008–2009 within-class ability grouping was used. In 2009–2010 schoolwide cluster grouping was implemented. In 2010–2011 students once again were grouped only within classrooms by ability and students identified as gifted were spread across all classrooms at each grade level. Results suggest that schoolwide cluster grouping influenced student performance in the year following its implementation, but only for mathematics and not the area of reading.

Keywords: cluster grouping; within-class ability grouping; gifted; K-6; measures of academic progress

Tracking and ability grouping

Ability grouping in schools has been the subject of heated debate. Opponents have claimed that ability grouping does not produce academic gains for gifted students and impedes the learning of typical students (Oakes, 1987; Slavin, 1987). Oakes (1987) noted the inequity of teacher and instructional quality for students grouped into low-track classrooms, though this likely is an implementation issue rather than a feature of ability grouping per se.

Ability grouping as a whole is often equated to tracking, but these terms are considered to have different meanings in the gifted education context. Some researchers in response refer to ability grouping as "flexible ability grouping" (Feldhusen & Moon, 2004; Neihart, 2007; Tieso, 2003). Whereas tracking implies a permanent assignment to a sequence of courses for students at a certain ability level, these authors and others have emphasized that flexible ability grouping allows students to move into and out of grouping assignments at any time given demonstration of new capabilities (Feldhusen & Moon, 2004).

*Corresponding author. Email: michael.matthews@uncc.edu

In a recent study on this topic, Kalogrides and Loeb (2013) examined data from over 900 schools, and concluded that "most but not all of the variation in peer and teacher characteristics occurs between schools, rather than within schools, and much of the within-school sorting by race, ethnicity, and poverty is explained by differences in prior achievement" (p. 313). Furthermore, these authors note that "rigorous studies on the effects of tracking on student achievement have found little evidence that tracking hurts lower ability students" (p. 314). Though its effects actually appear to be relatively small, tracking remains controversial. The current study refers to ability grouping, not tracking, as student placement into classrooms was reassessed at least annually in the study setting.

Researchers have found that ability grouping benefits students at all ability levels (Feldhusen & Moon, 2004; Gentry & Owen, 1999; Kulik & Kulik, 1992; Neihart, 2007; Nomi, 2010; Pierce et al., 2011). For example, Kulik and Kulik (1992) conducted a meta-analytic review of the literature on ability grouping and found students in high ability groups *and* low ability groups made academic gains when grouped homogeneously. Despite arguments by both sides of the debate, recent data to examine best practices for gifted and typical students remain scarce.

Two types of ability grouping

Ability grouping can be classified into two categories based on the level at which the grouping occurs: between-class ability grouping and within-in class ability grouping. Between-class ability grouping is the assignment of students to classrooms based on academic ability and/or prior performance, also known as schoolwide cluster grouping. Pull-out gifted programming is an example of this type of grouping that is usually implemented as a part-time rather than full-time model (Peters, Matthews, McBee, & McCoach, 2013).

Within-class ability grouping is the assignment of students to groups within each classroom based on interest, skill, ability, and various other factors. These groups may be heterogeneous or homogeneous; assignment is decided by the teacher (Gentry & MacDougall, 2009). While both types of grouping may be used concurrently, often in practice only one of the two is used at any given time.

From an instructional point of view, either grouping practice reduces variability in the group's learning abilities and needs, allowing a given level of instruction to be relevant to a greater proportion of learners than it would be in the absence of grouping. In addition to reducing demands on the teacher's preparation time, having fewer different instructional levels also increases the proportion of the teachers' time spent interacting with students at any given instructional level.

Cluster grouping

One specific approach to ability grouping is schoolwide cluster grouping. It generally can be defined as the placement of "several high achieving, high ability, or gifted students in a regular classroom with other students and a teacher who has received training or has a desire to differentiate curriculum and instruction for these 'target' students" (Gentry & MacDougall, 2009, p. 3). This practice has demonstrated effectiveness in meeting the academic needs of not only high-achieving students, but students of other achievement levels as well (Gentry & MacDougall, 2009; Gentry & Owen, 1999).

Total school cluster grouping is a specific cluster grouping model that "operates on the premise … the gifted education program will enhance the entire school" (Gentry & MacDougall, 2009, p. 8). This model considers the achievement levels of all students in a school and provides yearly placements for them. Although students are initially assigned a category (i.e. high achieving, above average, average, low average, low, and special education), category assignment is expected to change as student achievement levels increase in response to appropriately differentiated curriculum (Gentry & Mann, 2008). Category clusters are distributed evenly within classrooms, with the exception of the high ability cluster, which is assigned to one specific classroom with an appointed teacher. To provide opportunities for every student to benefit from gifted programming, all teachers receive ongoing professional development on gifted education curriculum, instruction, and strategies (Gentry & MacDougall, 2009).

Ability grouping with students identified as gifted

In the past 10 years, relatively few studies have investigated the effects of different ability grouping practices on gifted and typical students' academic achievement. In one study, using a pre-test–post-test, quasi-experimental design, Tieso (2005) examined the effects of within and between-class ability grouping combined with a differentiated curriculum on the math achievement of elementary school students. Both between- and within-class ability grouping resulted in significant, sustainable academic gains for high ability students, though low ability students did not make statistically significant academic gains under either form of grouping.

In another study, Nomi (2010) conducted a causal-comparative study using propensity score matching to examine the relationships between a school's likelihood of adopting ability grouping, the school's demographic characteristics, and the reading achievement of first-grade students. Because of limitations in the national data-set used, it was not possible to determine grouping practices at the within-classroom level, so Nomi's analysis categorized schools as using or not using ability grouping, and assumed such grouping took the form of within-class grouping in classrooms within each of these schools. While ability grouping benefitted or did no harm in schools at all levels of likelihood for using ability grouping, as measured by statistical significance, notably it was low-ability students attending schools in affluent settings – the settings least likely to report using ability grouping – who benefited the most academically from these practices.

Gentry and Owen (1999) found students of *all* ability levels made statistically significant academic gains in a school where the total school cluster grouping model was implemented. Using a longitudinal, causal-comparative design, these authors compared students' achievement in an elementary school using the total school cluster grouping model with achievement at a demographically comparable school that had not participated in the model and did not offer special programming for the gifted. Quantitative findings (i.e., increased reading achievement, higher math achievement, and increased numbers of students identified as high-achieving in the treatment school) indicated the use of total school cluster grouping had a positive impact on all students.

In a descriptive action research study, Brulles, Saunders, and Cohn (2010) examined the effects of within-class cluster grouping on the achievement of gifted students who were taught by teachers extensively trained in gifted education and those who were not. Student learning gains were higher when gifted students received services in a gifted cluster class (i.e. with a teacher trained in gifted education and differentiation). Overall differences were large, as indicated by a partial η^2 effect size of .31, though it is known that this effect size measure tends to over-estimate in comparison to other effect size estimators such as ω^2. Critically, because Brulles et al. were not able to provide equivalent teacher training in grouped and non-grouped schools, some portion of the effects they observed in the cluster-grouped classrooms may have been due to teacher training rather than to the cluster grouping intervention itself.

Most recently, Pierce et al. (2011) examined the effect of schoolwide and within-class cluster grouping combined with teacher training in specific differentiated gifted curricula on the mathematics achievement of gifted and typical elementary students. Using a longitudinal pre-test–post-test design, the largest performance gains during two initial years of a planned six-year study were found for both gifted and typical students in the cluster-grouped classrooms; all students benefited from differentiated instruction in the cluster context. Teacher training in this study included a four-day summer training workshop plus an unspecified number of professional development days during the school year.

Teacher training in ability grouping research

Prior research predominately has examined specific forms of ability grouping in combination with some type of professional development. For example, Tieso (2005) examined the two major components of ability grouping, between and within-class grouping, and teaching training on students' math achievement. Gentry and Owen (1999) studied the effects of the total school cluster grouping model on the reading and math achievement of students instructed by teachers extensively trained in gifted education, and both recent studies we located (Brulles et al., 2010; Pierce et al., 2011) also included extensive teacher training.

In the current study, students identified as gifted were clustered within a single classroom at each grade level for a single year, and teachers did not receive supplemental training in differentiated instruction. Thus, the purpose of this study was to investigate the effects of schoolwide cluster grouping, alone, on the academic achievement growth of elementary school gifted and typical students during a single year of clustering and in the absence of a corresponding professional development component. The primary research question addressed by this study was: does schoolwide cluster grouping yield distinct effects on the academic achievement growth of elementary school gifted and typical students, in comparison to within-class ability grouping alone?

Method

Participants and setting

The setting was a dual-language immersion charter school which serves 360 students in grades K-6 in a southeastern US state. Because the immersion language is uncommon in US schools, we have chosen not to identify it here, as doing so would potentially reveal the school's identity. This is why there are relatively few Latino or Asian students in the school population, as native speakers of the immersion language would be reported as White.

Students were identified as eligible for gifted services at the end of second grade in one of two ways, both of which used the composite age percentile score on the Cognitive Abilities Test (CogAT; Lohman & Hagen, 2005). Students scoring at or above the 97th percentile were automatically eligible for AIG placement. Students whose percentile score fell between the 90th and 96th percentile qualified for placement if they also obtained a score of 95th percentile or higher on the reading total, math total, or total battery of the Iowa Tests of Basic Skills.

This school is academically rigorous, and though admission is conducted via lottery, nearly one-third of the school population in grades 3–6 is formally identified as academically gifted under the school's criteria. Most students were identified as gifted in both reading and math, though a few were identified in only one of these areas; we grouped students for analysis so that students gifted in math were not included in the gifted group for the reading analysis, and vice versa.

Based on a gifted education consultant's recommendations to school leaders, cluster grouping was implemented across classrooms within each grade level (3–6) during the second year of the study. Though teachers reported satisfaction with the more focused instruction that cluster grouping allowed, parents of students not in the gifted cluster classrooms were unhappy with their child not being in the gifted placement, leading administrators the following year to revert to the previous practice of evenly spreading students identified as gifted across all classrooms on each grade level.

During the cluster-grouped year, students identified as gifted were placed together into a single one of the three or four classrooms at their grade level for reading, mathematics, or both areas, based on the area(s) in which they had been identified. When there were not enough identified students to fill a classroom, other high-ability learners whose scores had not quite met gifted criteria, but whose Northwest Evaluation Association Measures of Academic Progress (MAP) scores were comparable to the scores obtained by the gifted learners, were also placed into the same cluster classroom.

Gifted services in all three years were delivered by the classroom teacher, as the school did not have a dedicated gifted resource teacher. Two teachers (both at fourth grade) were enrolled in graduate coursework in gifted education during the study period, but there was no other specialized training offered at the school in differentiation or other gifted education practices.

Some differentiation was evident in the gifted cluster classes observed during the cluster-grouped year. Specifically, two instructional groups were evident in mathematics, and their teacher shared that the higher of these groups was able to benefit from instruction that presented content more rapidly and in greater depth than the lower group, though both groups worked on more advanced content than students did in the same teacher's two non-clustered classrooms. In the other content areas, observation suggests that differentiation was more commonly provided in the form of individualized options for projects and other class assignments rather than by specific instructional groupings.

During both non-cluster years (before and following the cluster-grouped year), students identified as gifted were spread evenly throughout all classrooms at each grade level and no schoolwide grouping by ability occurred. In all three years, within-class grouping practices were left to the discretion of each individual teacher.

The target population was gifted and typical elementary students who took the MAP assessment prior to and during the academic year that schoolwide cluster

Table 1. Participant demographics.

Gifted ($n = 68$)			Non-identified ($n = 186$)		
Characteristic	*n*	%	Characteristic	*n*	%
Gender:			*Gender:*		
Male	39	33.96	Male	87	48.25
Female	29	66.04	Female	99	51.75
Race/Ethnicity:			*Race/Ethnicity:*		
White	54	79.41	White	155	83.33
Asian	8	11.76	Asian	3	1.61
Multiracial	1	1.47	Multiracial	5	2.69
Hispanic	1	1.47	Hispanic	9	4.84
Black	4	5.88	Black	14	7.53

Table 2. Descriptive statistics for mathematics and reading outcomes by grade, gifted status, and cluster grouping.

			Mathematics			Reading		
Grade	Gifted	Cluster	*n*	Mean	SD	*n*	Mean	SD
1	0	0	35	180.8	13.8	37	174.1	10.8
	1	0	14	204.1	8.6	13	192.4	11.7
2	0	0	42	195.0	10.4	42	189.6	10.6
		1	43	188.8	13.4	43	181.8	11.6
	1	0	15	212.8	7.6	16	208.1	10.4
		1	15	209.3	6.0	15	200.3	9.0
3	0	0	32	205.2	11.7	33	202.3	12.2
		1	88	199.9	11.0	88	198.7	11.6
	1	0	6	215.5	7.7	12	210.3	7.0
		1	35	216.7	5.9	35	213.1	6.3
4	0	0	24	220.0	13.0	25	210.4	15.0
		1	74	211.0	10.9	74	205.4	10.4
	1	0	6	230.7	8.9	7	224.9	2.9
		1	43	226.9	7.4	43	219.7	6.7
5	0	1	63	222.1	10.7	63	213.3	11.4
	1	1	30	236.2	7.6	30	225.4	7.2
6	0	1	37	229.9	12.6	37	218.2	14.1
	1	1	10	245.8	5.2	10	234.5	3.8

Note: This table presents longitudinal data in which most subjects were measured at multiple grades for purposes of the statistical analysis, whereas Table 1 counts unique individuals; therefore the numbers will not agree across the two tables.

grouping was implemented, and who currently were only receiving instruction in within-class ability groups when the study took place. The entire population meeting the selection criteria was sampled ($n = 254$). Of the 254 unique participants, 186 students were not identified as gifted and 68 students were identified as gifted (Table 1). Typical and gifted students were predominately White (83 and 79%, respectively). Among typical students, slightly over half were female, while the identified-gifted student group was approximately two-thirds female (Table 1). Table 2 presents descriptive statistics for mathematics and reading outcomes (MAP scores) by grade, gifted status, and cluster group, but is based on a longitudinal count (i.e. multiple measures per individual) and therefore is not directly comparable to the unique count expressed in Table 1.

Procedures

Data collection

A database was set up prior to collecting MAP Reading and Math scores at the school site. Existing MAP data from 2008–2009, 2009–2010, and 2010–2011 were collected and de-identified. Data obtained from the school were further cleaned by the research team. Duplicate records (representing students tested twice during the same month) were resolved following these procedures: for duplicate scores, the most recent score was kept and the earlier attempt discarded (retesting was assumed); for Reading MAP data, Survey with Goals scores were kept, while PRI-READ-Survey w/Goals scores were discarded; and these scores primarily were associated with students who were too young for the gifted identification process to have been completed. Visual examination of MAP scores (via histograms produced in SPSS statistical software) revealed no obvious outliers; also, scores were not "piled up" at the top end of the distribution, suggesting that ceiling effects were not an issue within the study population.

Data were cross-checked by two researchers to guard against the possibility of errors in translating the complex output files into Excel and then into SPSS, which was used initially to tabulate demographic data. For the linear growth model portion of the analysis, a macro (or short section of programming code used to automate a repetitive task) was used to read study data directly from multiple comma-separated value (CSV-format) files into SAS for analysis. This method also was used to double check the unique demographic data reported in Table 1.

Instrumentation

The MAP (Northwest Evaluation Association, 2010) is a computer-adaptive test used to assess achievement in reading, math, and language. For gifted students, ceiling effects often exist because assessments are limited in range and unable to accurately measure the performance of high-achieving students (cf. McBee, 2010). However, because the MAP content ranges into high school-level items, ceiling effects are unlikely for examinees at the elementary school level.

MAP test and retest studies typically have looked at scores from the same students after a lapse of 7–12 months. Despite this relatively long interval, reliability indices have consistently been above what is considered statistically significant (Cronin, 2005). Furthermore, MAP scores are on a vertically equated scale, meaning scores are directly comparable across different ages and grade levels. Because of these features, the MAP can be considered suitable for a longitudinal analysis such as the current study.

Measurement

The effects of grouping on the academic achievement of gifted and typical elementary school students were measured in the form of test scores over time. MAP scores from 2008–2009 when the gifted-identified students were spread across all classrooms served as baseline data. Following the year of schoolwide cluster grouping in 2009–2010, the school returned to a policy of within-class grouping only. The data for 2009–2010 under schoolwide cluster grouping were compared to MAP scores from 2010–2011 (under within-class ability grouping only), as well as to 2008–2009 baseline test score data (also under within-class ability grouping only) in order to estimate the effects of the one year during which schoolwide cluster grouping was implemented.

Research design

An *ex post facto* design (Campbell & Stanley, 1963) was used for this study. In other words, the variation in the dependent variable(s) as a result of an independent variable has already occurred. This design is used to examine cause-and-effect relationships between independent and dependent variables when random assignment is not possible. Although there are weaknesses inherent in *ex post facto* research design, Gentry and Owen (1999) discuss ways it can be beneficial. First, this design allows the researcher to investigate a practice implemented in an actual school setting. Second, since the practice already has happened or is in progress, the school assumes ownership for the practice and is likely to have implemented the practice with some degree of fidelity. Third, intact groups may offer more stability over time and assist in longitudinal comparisons of students. Finally, *ex post facto* designs are recommended when little research exists on a topic, in order to develop a foundation for carefully controlled experimental and quasi-experimental research.

Data analysis

A piecewise multilevel growth model was fit to the data. The model allowed for an overall curvilinear trajectory over time. The linear portion of the slope was allowed to change at the beginning of schoolwide cluster grouping and once again after cluster grouping. Thus, the trajectory could take on three unique slopes representing three distinct statuses, which include prior to cluster grouping, during schoolwide cluster grouping, and after cluster grouping, with these linear slopes being situated in the context of the overall curvilinear trajectory. Gifted and typical students were allowed to have different intercepts, different slopes, and differing trajectories before, during, and after the schoolwide cluster grouping. The model was fit separately to the reading and mathematics MAP score outcomes. Since model fit metrics for hierarchical linear models provide relative rather than absolute information and we did not test or attempt to compare multiple models, model fit information is not provided.

Student grade level at assessment provided the time metric and was based on both the student's current grade as well as the semester of test administration. For example, a third grade student being assessed in the fall would be assigned a grade level of 3.2; a winter assessment would have a grade of 3.4, and a spring assessment a grade of 3.8 (decimals were based on the month of school during which the assessment was administered during this third grade year). Cluster grouping status was denoted by a time-varying dummy variable with values of zero prior to cluster grouping and values of one during and after the year of schoolwide cluster grouping. Post-cluster-grouped time periods were denoted by a second time-varying dummy variable taking values of zero prior to and during cluster grouping and one after cluster grouping. Grade (the time metric) was grand-mean centered prior to analysis.

Model-implied estimates were computed to predict the expected scores of prototypical students under different conditions. Estimates for a prototypical student not identified as gifted and never exposed to cluster grouping at the end of fourth grade and the end of sixth grade were computed. Similarly, estimates also were computed for a prototypical non-identified student exposed to cluster grouping during fourth grade, at the end of fourth grade, and the end of sixth grade. Tests of the cluster effect could be computed by contrasting estimated means of students exposed to cluster grouping vs. students not exposed to cluster grouping at each of these two time points. The same estimates and contrasts were performed for a prototypical student identified as gifted.

Results

Reading

Table 3 displays results from the piecewise growth model for the reading outcome. As expected, the linear and quadratic (grade squared) components of the trajectory are sizable and statistically significant. The positive linear slope with negative quadratic slope component describes an overall increase in reading scores over time with a decelerating rate of change in higher grades. The large positive main effect for gifted status ($B = 13.43$, $p < .0001$) indicates that gifted students scored, on average, 13.43 points higher in reading at the time point at which the grade (time) variable was centered, which as stated previously is the grand mean. The non-significant gifted-by-grade and gifted-by-grade squared interaction terms indicated that gifted students increased in reading ability at the same rate over time as typical students who had not been identified as gifted.

The negative and statistically significant coefficient for the grade by cluster interaction ($B = -1.19$, $p = .029$) indicates that typical students grew at a slightly slower rate than did gifted students during the cluster-grouped year. The non-significant interaction term for grade by gifted by cluster indicates that gifted students do not differ from typical students in the change in their slope during the cluster-grouped year; in other words, they too grew at a slightly slower rate in reading ability while cluster grouped.

Table 4 provides the model-implied estimates and helps to contextualize the meaning of the parameter estimates. The first pair of estimates describe expected

Table 3. Piecewise growth model results for reading outcome ($n = 1304/253$).

Variance components				
Level-1 (Residual)	36.45			
Level-2 intercept	88.72			
Level-2 slope	1.14			
Level-2 covariance	4.87			
Fixed effects	*Estimate*	*Std err*	*p*	*Sig*
Intercept	205.50	.797	<.0001	***
Grade (Time)	10.40	.647	<.0001	***
Grade^2	−1.93	.314	<.0001	***
Gifted	13.43	1.493	<.0001	***
Grade × gifted	−.79	1.274	.5368	
Grade^2 × gifted	.82	.623	.1889	
Grade × cluster	−1.19	.545	.0294	*
Grade × post cluster	1.19	.744	.1095	
Grade × gifted × cluster	−.36	1.056	.7356	
Grade × gifted × post cluster	.64	1.433	.6565	

*indicates $p < .05$.
***indicates $p < .001$.

CHAS 846251
27 September 2013
Initial
CE: BB QA: RB
Coll: QC:

Table 4. Model-implied estimates for reading outcome (n = 1304/253).

Label	Estimate	Std err	p	Sig	d
Typical non-cluster mean @ end of 4th grade	212.01	.869	–	–	
Typical cluster mean @ end of 4th grade	211.15	.823	–	–	
Typical non-cluster mean @ end of 6th grade	219.51	3.604	–	–	
Typical cluster mean @ end of 6th grade	219.52	1.554	–	–	
Gifted non-cluster mean @ end of 4th grade	225.3	1.448	–	–	
Gifted cluster mean @ end of 4th grade	224.18	1.342	–	–	
Gifted non-cluster mean @ end of 6th grade	236.86	6.27	–	–	
Gifted cluster mean @ end of 6th grade	237.64	2.662	–	–	
Typical cluster effect at end of 4th grade	−.86	.394	.0294	*	−.07
Typical cluster effect at end of 6th grade	.01	2.827	.998		.00
Gifted cluster effect at end of 4th grade	−1.12	.653	.0877		−.09
Gifted cluster effect at end of 6th grade	.77	4.672	.8688		.05

*indicates $p < .05$.
p values not provided for estimated means because the tests are uninformative.
Cohen's d effect sizes were computed using conditional sd of scores at each grade, ignoring giftedness or cluster grouped status.

reading scores for two typical students at the end of fourth grade, assuming that one of those hypothetical students had been cluster grouped during grade four and the other not cluster grouped. The row in the table labeled "typical cluster effect at end of fourth grade" provides a contrast of these two means. The estimate (M = −.859, p = .029) indicates that the typical cluster-grouped student would be expected to score nearly one point lower at the end of the cluster-grouped year than the non-cluster-grouped student. However, by the end of sixth grade, this deficit would have disappeared (M = .007, p = .998). There were no other significant comparisons. For gifted students, the point estimate of the effect of cluster grouping on reading scores at the end of fourth grade is negative and of similar magnitude as the estimate

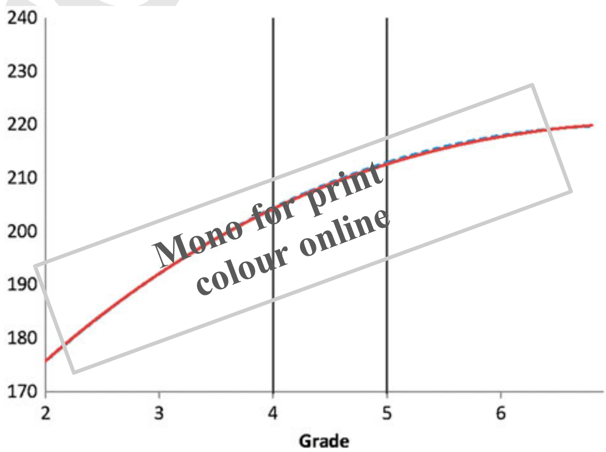


Figure 1. Estimated trajectories over time for cluster-grouped vs. non cluster-grouped *typical* students on reading. Dashed line indicates clustered group, which shows complete overlap with the non cluster-grouped solid line in this figure.

CHAS 846251
27 September 2013
Initial
CE: BB QA: RB
Coll: QC:

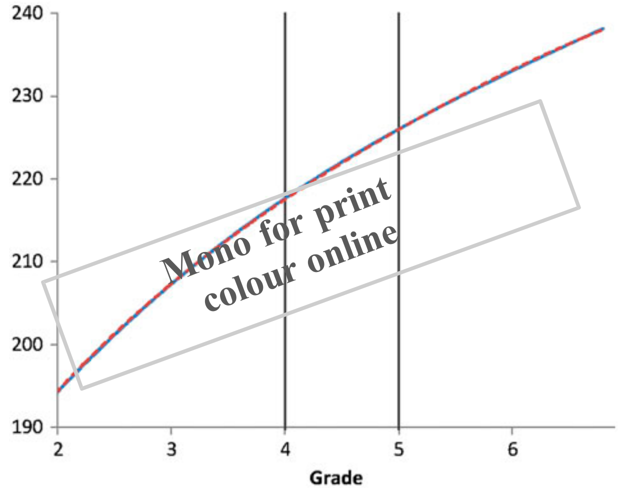


Figure 2. Estimated trajectories over time for cluster-grouped vs. non cluster-grouped *gifted* students on reading. Dashed line indicates clustered group, which shows complete overlap with the non cluster-grouped solid line in this figure.

for typical students. However, because of the smaller number of gifted students in the sample, their model-implied means are estimated with less precision (larger standard errors and 95% confidence intervals), and the contrast is non-significant.

To help visualize the pattern of differences, plots of the model-implied trajectories in reading are provided for typical and gifted students (Figures 1 and 2, respectively). Though the grouping is depicted as having taken place during time (grade year) 4–5, this is for illustrative purposes, and these graphs incorporate information from all grade levels during the three study years. The plots for cluster (dashed line) and non-cluster-grouped students (solid line) are so similar as to overlap.

Table 5. Piecewise growth model results for mathematics outcome (n = 1269/253).

Variance components				
Level-1 (Residual)	32.22			
Level-2 intercept	94.74			
Level-2 slope	3.72			
Level-2 covariance	1.46			
Fixed effects	*Estimate*	*Std err*	*p*	*Sig*
Intercept	211.66	.808	<.0001	***
Grade (Time)	11.14	.627	<.0001	***
$Grade^2$	−1.50	.297	<.0001	***
Gifted	15.53	1.523	<.0001	***
Grade × gifted	−1.23	1.242	.3204	
$Grade^2$ × gifted	.65	.595	.2729	
Grade × cluster	−.89	.526	.0903	
Grade × post cluster	2.45	.707	.0005	*
Grade × gifted × cluster	.07	1.013	.9411	
Grade × gifted × post cluster	−.04	1.375	.9744	

*indicates $p < .05$.
***indicates $p < .001$.

Mathematics

Table 5 displays results from the piecewise growth model for the mathematics outcome. The linear and quadratic (grade squared) components of the trajectory are sizable and statistically significant, once again describing an overall increase in math scores over time with a decelerating rate of increase in higher grades. The large positive main effect for gifted status ($B = 15.53$, $p < .0001$) indicates that gifted students scored, on average, 15.53 points higher in math at the grand mean grade. The non-significant gifted-by-grade and gifted-by-grade squared interaction terms indicated that gifted students increased in math ability at the same rate over time as typical students.

The non-significant coefficients for the grade by cluster interaction and the grade by cluster by gifted interactions indicate that neither typical nor gifted students grew at different rates during the cluster-grouped year. However, the positive and significant grade by post-cluster interaction term indicates that typical students grew at a substantially faster rate during the year following cluster grouping. The non-significant grade by gifted by post-cluster interaction indicates that gifted students did not grow at a different rate after cluster grouping than did typical students; in other words, both groups grew at faster rates after, but not during, the cluster-grouped year. From an advocacy perspective (cf. Wiskow, Fowler, & Christopher, 2011) that takes into account the body of literature supporting the use of schoolwide cluster grouping, it appears that both groups have benefitted from cluster grouping in mathematics, but not until the year following the cluster-grouped instruction. Regardless of the grade level during which the schoolwide clustering happened, the deviation of scores always happens *after* cluster grouping.

Table 6 provides the model implied estimates. There were no significant comparisons. Both the gifted and typical cluster-grouped sixth-grade means are approximately four points higher than their non-cluster-grouped counterparts; however, these comparisons were not statistically significant.

To help visualize the pattern, Figures 3 and 4 provide plots of the model-implied trajectories for typical and gifted students, respectively, for mathematics. The plots for cluster (dashed line) and non-cluster-grouped students (solid line) are so similar

Table 6. Model-implied estimates for mathematics outcome ($n = 1269/253$).

Label	Estimate	Std err	*p*	Sig	*d*
Typical non-cluster mean @ end of 4th grade	218.92	.876	–	–	
Typical cluster mean @ end of 4th grade	218.27	.831	–	–	
Typical non-cluster mean @ end of 6th grade	230.86	3.453	–	–	
Typical cluster mean @ end of 6th grade	235.1	1.492	–	–	
Gifted non-cluster mean @ end of 4th grade	233.9	1.456	–	–	
Gifted cluster mean @ end of 4th grade	233.31	1.357	–	–	
Gifted non-cluster mean @ end of 6th grade	247.86	6.092	–	–	
Gifted cluster mean @ end of 6th grade	252.19	2.577	–	–	
Typical cluster effect at end of 4th grade	−.64	.38	.09		−.05
Typical cluster effect at end of 6th grade	4.25	2.703	.117		.32
Gifted cluster effect at end of 4th grade	−.59	.625	.345		−.04
Gifted cluster effect at end of 6th grade	4.33	4.515	.338		.33

AQ8
*indicates $p < .05$.
p values not provided for estimated means because the tests are uninformative.
Cohen's *d* effect sizes were computed using conditional sd of scores at each grade, ignoring giftedness or cluster grouped status.

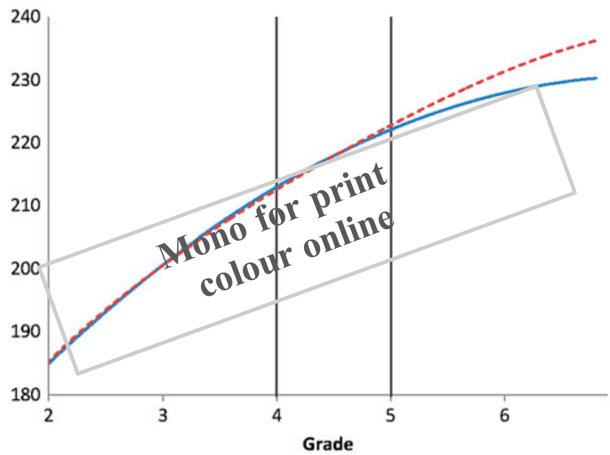


Figure 3. Estimated trajectories over time for cluster-grouped vs. non cluster-grouped *typical* students on mathematics. Dashed line indicates non-cluster grouped students. Data points extend through Grade = 6.8.

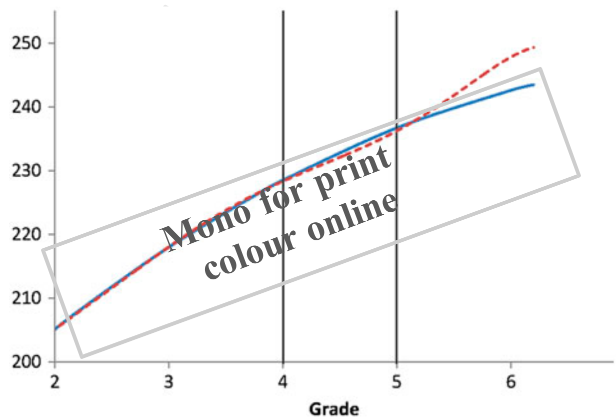


Figure 4. Estimated trajectories over time for cluster-grouped versus non cluster-grouped *gifted* students on mathematics MAP RIT scores. The dashed line indicates cluster grouped students; solid line indicates non-cluster grouped students.

as to overlap on these figures through the end of the cluster grouping period, but following the schoolwide cluster grouping period they begin to diverge. This appears to indicate a delayed effect of the cluster grouping in mathematics. Although the trajectories for the gifted and non-gifted students appear to diverge in the period following cluster grouping, statistically these differences in trajectory are not significantly different. However, the increase in rate of skill acquisition in the post-cluster time period was significant for both groups.

Discussion

For reading, both gifted and typical learners showed similar rates of growth under the two grouping conditions. Estimated achievement growth trajectories in reading

(Figures 1 and 2) did not diverge. In contrast, findings suggest that total school cluster grouping may be beneficial in mathematics.

This study found no evidence that cluster grouping in reading is associated with any lasting harm or any lasting benefits. Despite a negative and statistically significant decrease in the rate of skill acquisition for non-gifted students during the cluster-grouped year, the practical or educational significance of this deficit, which was less than one point, is null. By the end of 6th grade, this small deficit for the non-gifted students has disappeared entirely. Therefore, results taken in isolation might be used to argue that the decision to cluster group or not cluster group in reading should not be based on concerns about academic achievement.

In mathematics, the study found both statistically and practically significant increases in the rate of skill acquisition for both gifted and typical students in the years following cluster grouping, but curiously, not during the year of cluster grouping itself. This suggests that the benefits of cluster grouping may take more than one year to manifest themselves, and in fact, prior studies by both Gentry and Owen (1999) and Pierce et al. (2011) both involved multiple years of implementation of the cluster grouping intervention. Both of these studies found the greatest gains in student achievement during the latter years of the intervention. In the current study, the benefits of grouping did not differ by gifted status; in other words, both gifted and non-gifted students benefitted by the same amount. Though the follow-up test of math scores at the end of 6th grade did not identify statistically significant differences in score between hypothetical cluster and non-cluster-grouped students, this contrast suffered from low power. The point estimates of 4.25 points for non-gifted and 4.33 points for gifted students would be educationally significant effects, representing nearly half a year's worth of instructional gain. Of course, these comparisons were not statistically significant so they cannot be reliably distinguished from sampling error, but the difference in rates of skill acquisition was significant. Because we included the quadratic term for time in the model, these changes in the trajectory corresponding with the post-cluster grouping period are not being driven by erroneously modeled curvature in the slope. They are real effects.

Gentry and Owen (1999) did document greater growth in mathematics under total school cluster grouping than in reading. Unfortunately, power considerations due to the relatively small population of the school in the present study precluded separating analyses by grade level, so any grade level differences in the effectiveness of the schoolwide cluster grouping would not have been captured in the current study.

Implications for practice

Our results suggest that total school cluster grouping may be beneficial to both gifted and typical learners, at least in the area of mathematics, though it appears that this intervention many not yield observable results during the first year of implementation. Given the emphasis placed on standardized testing in US schools, combined with the suggestion that math achievement ultimately increases when all students, especially gifted students, are grouped in schoolwide as well as within-class ability clusters, consideration should be given to this grouping practice when determining school-level assignments to mathematics classrooms in grades 3–6.

CHAS 846251
27 September 2013
Initial
CE: BB QA: RB
Coll: QC:

Suggestions for future research

Because *ex post facto* research cannot rule out sources of confounding and therefore cannot reliably distinguish between causal and non-causal effects, future studies should utilize true experimental designs with random assignment or stronger quasi-experimental designs, such as regression discontinuity, instrumental variables, or propensity score matching, to evaluate the possibility of a causal relationship between total school cluster grouping and student achievement. Additionally, studies conducted in multiple schools or even within a single larger school would allow for tests of moderation by grade level or student characteristics, which was not feasible in this study. For example, given Nomi's (2010) conclusion that ability grouping boosts achievement gains for students of *all* ability levels attending schools with advantageous characteristics, comparisons could be made between schools in high, moderate, and low socioeconomic settings. Also, more attention should be given to differences between reading and math growth under different grouping conditions.

Another critical area for future research concerns the role professional development plays in the success of cluster grouping. The majority of cluster grouping studies have combined cluster grouping with some type of teacher training component AQ5 (Brulles et al., 2010; Gentry & Owen, 1999; Pierce et al., 2010). The significant results obtained from these studies are supported by Gentry and MacDougall's (2009) claim that professional development is essential to the success of cluster grouping, and are consistent with the statement by Brulles et al. that "teacher training remains a critical component to the success of the cluster program model" (p.346). Future research should examine the extent to which cluster grouping combined with professional development may be more effective than cluster grouping alone. This is a particularly critical area for future study because among these two components, the training constitutes most of the overall implementation cost. Finally, due to its reduced ceiling effects, ongoing use of the MAP test is recommended.

Limitations

Causal-comparative research is inherently flawed by its inability to establish causal relationships and control for threats to validity; thus, our results must be viewed with caution. Even if not confounded, our results may not represent "marginal" causal effects due to omitted moderators. For example, this study was conducted at a language immersion charter school; therefore, the results may not be generalizeable to other schools, especially those whose composition and characteristics differ (i.e. regular public schools, schools that lack dual-immersion language programming, or potentially even dual-immersion schools using other languages). Though not all students had scores for all the three study years, missing values are not a limitation in this case because results in HLM are not adversely affected by unbalanced data.

Unknown selection effects may have influenced our results due to the relatively small sample size. Because many students not identified as gifted in this setting already perform above grade level, overall achievement levels may have influenced results in some way. The second language component of the school's curriculum could have confounded the results specifically via unanticipated effects on reading scores. Unlike the studies by Brulles et al. (2010), Gentry and Owen (1999), Pierce et al. (2011), and Tieso (2005), teachers in our study did not receive targeted professional development during the cluster grouping intervention. Presumably, such professional development would increase the effectiveness of this method of grouping. Because it was difficult to separate effects ascribed to professional development, differentiated instruction, and grouping, this too is a potential limitation of the study. Finally, there was no way to monitor the effectiveness of individual teachers' responses, though comfort levels with differentiation likely varied across individual teachers.

Conclusion

Due to limited research, causal statements cannot yet be made about the impact specific types of ability grouping (e.g. total school cluster grouping) have on student achievement growth. As additional studies accrue, meta-analysis eventually may be possible, and future studies should aim toward this goal. If informed decisions are to be made about ability grouping, future research efforts must attempt to shed light on the contextual variables that influence the effectiveness of different approaches to ability grouping, including total school cluster grouping, for gifted and typical students.

References

Brulles, D., Saunders, R., & Cohn, S. J. (2010). Improving performance for gifted students in a cluster grouping model. *Journal for the Education of the Gifted, 34*, 327–350.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 1–76). Chicago, IL: Rand-McNally.

Cronin, J. (2005). *NWEA reliability and validity estimates: Achievement level tests and measures of academic progress* (NWEA Report). Retrieved from http://center.uoregon.edu/conferences/CCSA/CCSA08/uploads/KEY_1042180/Tilton_ResearchNWEAReliabilityandValidity.pdf

Feldhusen, J., & Moon, S. (2004). Grouping gifted students: Issues and concerns. In *Grouping and acceleration practices in gifted education* (pp. 81–90). Thousand Oaks, CA: Corwin Press. Retrieved from PsycINFO database.

AQ6 AQ7

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Thousand Oaks, CA: SAGE.

Gentry, M., & MacDougall, J. (2009). Total school cluster grouping: Model, research, and practice. In J. S. Renzulli, E. J. Gubbins, S. K. McMillen, R. D. Eckert, & C. A. Little (Eds.), *Systems & models for developing programs for the gifted & talented* (2nd ed., pp. 211–234). Mansfield Center, CT: Creative Learning Press.

Gentry, M., & Mann, R. L. (2008). *Total school cluster grouping and differentiation: A comprehensive, research-based plan for student achievement and improving teacher practices*. Mansfield Center, CT: Creative Learning Press.

Gentry, M., & Owen, S. (1999). An investigation of the effects of total school flexible cluster grouping on identification, achievement, and classroom practices. *Gifted Child Quarterly, 43*, 224–243. doi:10.1177/001698629904300402

Kalogrides, D., & Loeb, S. (2013). Different teachers, different peers: The magnitude of student sorting within schools. *Educational Researcher, 42*, 304–316. doi:10.3102/0013189X13495087

Kulik, J. A., & Kulik, J. C. (1992). Meta-analytic findings on grouping programs. *Gifted Child Quarterly, 36*, 73–77. doi:10.1177/001698629203600204

Lohman, D. F., & Hagen, E. P. (2005). *Cognitive abilities test (CogAT)*. Rolling Meadows, IL: Riverside Publishing.

McBee, M. T. (2010). Modeling outcomes with floor or ceiling effects: An introduction to the Tobit model. *Gifted Child Quarterly, 54*, 314–320. doi:10.1177/001698621037909

Neihart, M. (2007). The socioaffective impact of acceleration and ability grouping: Recommendations for best practice. *Gifted Child Quarterly, 51*, 330–341. doi:10.1177/0016986207306319

Nomi, T. (2010). The effects of within-class ability grouping on academic achievement in early elementary years. *Journal of Research on Educational Effectiveness, 3*, 56–92. doi:10.1080/19345740903277601

Northwest Evaluation Association. (2010). *Measures of academic progress MAP basics overview*. Portland, OR: Author. Retrieved from http://www.nwea.org/sites/www.nwea.org/files/resources/MAP%20Basics%20Overview_0.pdf

Oakes, J. (1987). Tracking in secondary schools: A contextual perspective. *Educational Psychologist, 22*, 129–153.

Peters, S. J., Matthews, M. S., McBee, M. T., & McCoach, D. B. (2013). *Beyond gifted education: Designing and implementing advanced academic programs*. Waco, TX: Prufrock Press.

Pierce, R. L., Cassady, J. C., Adams, C. M., Speirs Neumeister, K. L., Dixon, F. A., & Cross, T. L. (2011). The effects of clustering and curriculum on the development of gifted learners' math achievement. *Journal for the Education of the Gifted, 34*, 569–594.

Slavin, R. E. (1987). Grouping for instruction in the elementary school. *Educational Psychologist, 22*, 109–127.

Tieso, C. (2003). Ability grouping is not just tracking anymore. *Roeper Review, 26*, 29–36. doi:10.1080/02783190309554236

Tieso, C. (2005). The effects of grouping practices and curricular adjustments on achievement. *Journal for the Education of the Gifted*, *29,* 60–89. Retrieved from PsycINFO database.

Wiskow, K., Fowler, V. D., & Christopher, M. M. (2011). Active advocacy: Working together for appropriate services for gifted learners. *Gifted Child Today, 34,* 20–25.