Abstract Title Page

Title

First-grade retention: Effects on children's actual and perceived performance throughout elementary education

Author(s)

Mieke Goos, Jan Van Damme, Patrick Onghena, & Katja Petry
Department of Educational Sciences
Katholieke Universiteit Leuven
Belgium
E-mail: mieke.goos@ped.kuleuven.be

Abstract Body

Background

Within any class in early elementary education, there is typically considerable variation in students' cognitive and psychosocial skills. The question often raised by parents, educators, and policy makers is what to do with children who do not succeed well in passing through these early years of education. Some nations privilege social promotion and 'on-time' educational interventions (for example, Scandinavian countries). Many other countries (like the United States, France, and Belgium) respond to this challenge with early-grade repetition.

Yet, despite its popularity, early-grade retention remains a controversial practice with both proponents and opponents, who all adduce their own theory-based arguments.
Proponents of early-grade retention hypothesize that …

- it grants children more time to catch up on prerequisite knowledge and skills, preventing later school failure (maturationist developmental theory; Gesell, 1940; Piaget & Inhelder, 1962);
- it gives weak performing children the opportunity to refresh, relearn, and experience new successes throughout the retention year, resulting in a feeling of being competent, more pleasure in learning, a higher motivation, and a better school liking (self-determination theory; Deci & Ryan, 2000);
- it motivates repeaters since the alternative option (i.e., promoting them to the next grade while they are not proficient enough in the subject matter of the current grade) sends them the (wrong) message that little is expected of them and rewards them undeservedly for having not worked hard (theory of operant conditioning; Skinner, 1953);
- it stimulates repeaters' self-concept and self-confidence during the retention year because they start with a small advantage in knowledge and skills over their new, younger classmates (social comparison theory; Festinger, 1954); and
- it results in more homogeneous classes and thus more efficient to instruct and on average better performing classes (socio-cultural developmental theory; Vygotsky, 1978).

Opponents of early-grade retention hypothesize that …

- it deprives retainees of access to meaningful intellectual challenges on a continual basis – since repeating equals sheer rehearsal – consequently disrupting and impeding children's academic growth (contextual developmental theory; Bronfenbrenner, 1980; Vygotsky, 1978);
- it does not solve children's learning problems since during the retention year neither learning objectives nor instruction methods are adapted to the retained student's individual needs (theory of direct instruction; Carroll, 1963);
- it gives children feelings of 'having failed', humiliation, and shame, when comparing themselves with their previous, promoted classmates, with negative implications for their self-concept and self-confidence (self-determination theory; Deci & Ryan, 2000; social comparison theory; Festinger, 1954);
- it demotivates, frustrates, and punishes repeaters, because they have to do all subjects again including those they were already sufficiently skilful at (self-determination theory; Deci & Ryan, 2000);
- it influences children's existing peer relationships and thus their feeling of relatedness in a negative way, since early-grade repeaters have to leave their friends behind (self-determination theory; Deci & Ryan, 2000);

- it increases the risk of repeaters being socially rejected or bullied by their new, younger classmates given that early-grade retention is associated with several negatively loaded labels like 'stupidity' and 'abnormality' (labeling theory; Becker, 1963; Lemert, 1967); and
- it increases the risk of retainees being treated socially different by their teacher (i.e., more reprimands, less reinforcement of positive actions, less tolerance) because teachers, in general, have negative perceptions and low expectations on behalf of repeaters (theory of the Pygmalion effect; Rosenthal & Jacobson, 1992).

Purpose of this study

Since in Flanders (i.e., the Dutch-speaking part of Belgium) early-grade repetition, and especially first-grade repetition, is widely used and more or less socially approved by parents, educators, as well as policy makers, we would like to investigate this practice in more detail, while particularly paying attention to the methodology used. More specifically, the current study investigates the effects of repeating first grade on children's further academic growth, by tracking the actual performance and the teacher-rated performance of a cohort of Flemish first-graders until the end of elementary school. Two research questions are raised:

- How do first-grade repeaters, at the cost of one extra year of education, develop in comparison to younger children with whom they will eventually finish elementary school (i.e., same-grade comparison), taking into account their propensity of repeating first grade?
- How would first-grade retainees have developed, had they been promoted to second grade instead (i.e., same-age comparison)?

Setting

This study is part of the large-scale longitudinal SiBO[1] project, designed to investigate Flemish children's development and school trajectory throughout elementary education (Maes, Ghesquière, Onghena, & Van Damme, 2002). The SiBO project involves a random stratified sample of 122 Flemish primary schools, which was found to be representative for the entire Flemish school population. A cohort of approximately 4,000 children (219 classes) was recruited and was followed for the whole duration of the project, from the beginning of first grade (school year 2003-2004) until the end of sixth grade.

Participants

In the present study, all first-time first-graders (N = 3,707) were examined. At the beginning of first grade, the mean age of all children was 6 years and 3 months. There were 50.6% boys. 22% of the children had at least one parent having a foreign nationality and 26% of the children were targeted by the Equal Educational Opportunities Act. In school year 2004-2005, a subset of 298 pupils (8%) repeated first grade (experimental group), while 3,326 children (90%) were promoted to second grade (control group). The remaining students (2%), moving to third grade or special education, were excluded from our analyses.

[1] SiBO is the Dutch acronym for 'Schoolloopbanen in het BasisOnderwijs' (School trajectories in Elementary Education).

Practice

In Flanders, about 7% of the children repeat first grade. What makes our school system rather unique internationally is that no formal rules exist regarding grade promotion: each school can decide for itself how to measure children's mastery of the curriculum. Next, though elementary schools officially have the final word in the decision to retain a child, this decision is rather a decision made by teachers and parents jointly. Besides, as already mentioned, the practice of repeating first grade is more or less accepted. This implies that being a first-grade retainee in Flanders might have a different connotation than it has in the United States (rather negative overtone) and thus also different effects.

Research Design

Since our data are observational in nature, we made use of propensity score stratification to balance retained and promoted children with respect to their distribution of prior characteristics (Rosenbaum, 2002), thereby creating a quasi-randomized experiment. More specifically, we had 169 preretention child-, class-, and school-level covariates to our disposal (see below). Only those covariates related to both treatment (i.e., first-grade retention) and outcome (i.e., children's academic growth) were included in the propensity score model (Austin, Grootendorst, & Anderson, 2007; Brookhart et al., 2006; Judkins et al., 2007; Newgard, Hedges, Arthur, & Mullins, 2004), a three-level logistic regression model. Stratification was based on the deciles of the propensity score logit.

Data Collection and Analysis

Academic growth throughout elementary education. To follow children's academic performance through the elementary years, several achievement tests were administered at the end of each school year. Math tests, covering multiple domains in mathematics, were administered annually. The math scores were vertically equated using item response modeling (IRT). A reading fluency test was administered in first, second, and third grade. The reliabilities of all achievement tests, assessed by Cronbach's alpha coefficient, were considered good to high (ranging from .88 to .94 across all waves). Secondly, teachers were asked to rate the overall math and language skills of each child in their class, by filling in a questionnaire at the end of each school year. Both items were rated on a five-point Likert scale.

Propensity score. The SiBO data provide a comparatively comprehensive set of potential confounders of first-grade retention: data were available on 68 prior child characteristics (e.g., child demographics, academic achievement, psychosocial functioning, parental school involvement), 59 prior class characteristics (e.g., teacher demographics, class composition, instructional practices), and 42 prior school characteristics (e.g., principal demographics, school composition, school culture, school policy). Data were obtained from official records, achievement tests, the intelligence test Standard Progressive Matrices, a teacher questionnaire about the child, extensive parent questionnaires, a teacher questionnaire about teacher didactics, and a school staff questionnaire.

Analyses. Our analyses involved four steps, to a large extent in line with the procedure Hong and Raudenbush (2005; 2006) and Hong and Yu (2007; 2008) used[2]. In step 1, we identified the 'true' confounders of first-grade retention: those observed preretention covariates that are related to both treatment and outcome[3]. This yielded 55 child characteristics and 3 class characteristics. Step 2 included the estimation of each child's conditional propensity to repeat first grade as a function of these confounders. For that purpose, we constructed a three-level logistic regression model, with students nested within classes, nested within schools. The results can be found in Figure B1. In step 3, all pupils were stratified based on the logit of their estimated propensity score. Ten strata of equal size were created, of which the results are shown in Table B1. Within-stratum balance was achieved in the propensity score, as well as in at least 95% of the observed preretention covariates. Finally, in step 4, we estimated the average academic effects of first-grade retention. To that end, we constructed a number of three-level growth curve models, with the repeated measures of academic achievement at level 1, students at level 2, and schools at level 3. We carried out two sets of growth curve analyses: one based on the same-grade approach (i.e., comparing point B versus the 'to-be-promoted' subset in point A and so on in Figure B2) and one based on the same-age approach (i.e., comparing point B versus C and so on in Figure B3).

Findings

The graphical representations of retained and promoted children's academic growth trajectories based on the same-grade approach are displayed in Figure B4. Our results, on average, reveal that first-grade repeaters start their retention year with an advantage in math and reading fluency compared to their new, younger grade-mates who are at similar risk of being retained. Their teachers, however, rate them as equally (bad) skilled in math and language. What is more, over time first-grade repeaters grow significantly slower, making them even end up with a relative disadvantage at the end of elementary school, both in actual performance and as perceived by the teacher.

Same-age comparisons, on the other hand, consistently show that first-grade repeaters would have achieved better in math and reading fluency and would have been rated more positively by their teacher in terms of math and language skills, had they been promoted to second grade instead, both in the short and long run. The graphical representations of these growth curve models can be found in Figure B5.

Conclusions

The main contribution of our study to the field of education is twofold. Firstly, to our knowledge, this is the first study to look at the academic outcomes of retained and promoted students in terms of both actual performance and performance as perceived by the teacher. Though we

[2] Hong and colleagues utilized this procedure to make same-age comparisons of retained and promoted kindergartners and first-graders. We apply this methodology to both same-age and same-grade comparisons, and focus on first-grade retention only. Moreover, our propensity score model is more restricted, in the sense that only covariates that relate to both treatment and outcome are included.

[3] For the sake of comparison, we also reanalyzed our data relying on propensity scores as originally defined by Rosenbaum and Rubin (1983; 1984; 1985), meaning based on all covariates related to the treatment, irrespective of their relation to the outcome. This yielded similar estimates of first-grade retention effects.

acknowledge that we have not tested any explanatory mechanisms in the relationship between first-grade retention and students' achievement trajectories throughout elementary school, our findings provide some preliminary insights into this issue. Our results indicate, as already mentioned, that during the retention year, first-grade repeaters outperform their younger grade-mates who are at similar risk of being retained. But, teachers do not rate or perceive them as doing so. This pattern even worsens throughout time. Moreover, teacher ratings or perceptions would have been more positive in general, had first-grade repeaters not repeated their grade. This might be an indication that a Pygmalion effect might be hold responsible for the negative effects of early-grade repetition found in the literature so far. Again, we did not test for such an explanation. It would be interesting to have a closer look at this issue, as well as other explanatory mechanisms (see first paragraph), in future research.

Secondly, our results have practical implications. Our findings seem to suggest that making children who are not keeping up repeat first grade, on average, is not helpful for their performance in math and reading fluency. Moreover, it seems that teachers (perhaps unconsciously) have severe negative perceptions with regard to repeaters. Combining these findings with the results we presented at the SREE 2010 conference (i.e., first-grade repetition is ineffective for repeaters' psychosocial growth throughout elementary school as well), we have reasons to start calling this educational practice into question.

Appendices

Appendix A. References

Austin, P. C., Grootendorst, P., & Anderson, G. M. (2007). A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Statistics in Medicine, 26*, 734-753. Retrieved from ISI:000244236700003

Becker, H. S. (1963). *Outsiders: Studies in the sociology of deviance*. New York: Free Press.

Bronfenbrenner, U. (1980). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J., & Sturmer, T. (2006). Variable selection for propensity score models. *American Journal of Epidemiology, 163*, 1149-1156. Retrieved from ISI:000238424900011

Carroll, J. (1963). A model of school learning. *Teachers College Record, 64*, 723-733.

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*, 227-268. Retrieved from WOS:000166046400001

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117-140.

Gesell, A. (1940). *The first five years of life*. New York: Harper.

Hong, G., & Yu, B. (2007). Early-grade retention and children's reading and math learning in elementary years. *Educational Evaluation and Policy Analysis, 29*, 239-261. Retrieved from ISI:000251195400001

Hong, G. L., & Raudenbush, S. W. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. *Educational Evaluation and Policy Analysis, 27*, 205-224. Retrieved from ISI:000235127900001

Hong, G. L., & Raudenbush, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association, 101*, 901-910. Retrieved from ISI:000240158700004

Hong, G., & Yu, B. (2008). Effects of kindergarten retention on children's social-emotional development: An application of propensity score method to multivariate, multilevel data. *Developmental Psychology, 44*, 407-421. doi:doi: DOI: 10.1037/0012-1649.44.2.407. Retrieved from http://www.sciencedirect.com/science/article/B6WYC-4S9KYNS-B/2/78f7f82a8d40626e9d5a5c9d3c22508e

Judkins, D. R., Morganstein, D., Zador, P., Piesse, A., Barrett, B., & Mukhopadhyay, P. (2007). Variable selection and raking in propensity scoring. *Statistics in Medicine, 26*, 1022-1033. Retrieved from ISI:000244422100007

Lemert, E. (1967). *Human deviance, social problems and social control*. Englewood Cliffs, NJ: Prentice-Hall.

Maes, F., Ghesquière, P., Onghena, P., & Van Damme, J. (2002). *Longitudinaal onderzoek in het basisonderwijs: Van doelstellingen tot onderzoeksopzet [A longitudinal study in elementary education: Research objectives and planning]* (Report No. 1). Leuven, Belgium: Steunpunt Loopbanen doorheen Onderwijs naar Arbeidsmarkt. Retrieved from: http://www.steunpuntloopbanen.be/rapporten/LOA-rapport_01.pdf

Newgard, C. D., Hedges, J. R., Arthur, M., & Mullins, R. J. (2004). Advanced statistics: The propensity score - A

method for estimating treatment effect in observational research. *Academic Emergency Medicine, 11*, 953-961. Retrieved from ISI:000223732700009

Piaget, J., & Inhelder, B. (1962). *The psychology of the child*. New York: Basic Books.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*, 41-55. Retrieved from ISI:A1983QH66900005

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association, 79*, 516-524. Retrieved from ISI:A1984TK97500004

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control-group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician, 39*, 33-38. Retrieved from ISI:A1985ABK0900009

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

Rosenthal, R., & Jacobson, L. (1992). *Pygmalion in the classroom*. New York: Irvington.

Skinner, B. F. (1953). *Science and human behavior*. Oxford, England: Macmillan.

Vygotsky, L. S. (1978). *Mind in society. The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Appendix B. Tables and Figures

Table 1
Within-Stratum Balance in the Propensity Score Logit

	Promoted students			First-grade repeaters			Total		
Decile	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
1	241	-6.27	0.22	1	-6.78		242	-6.28	0.23
2	243	-5.71	0.13	0			243	-5.71	0.13
3	240	-5.34	0.10	2	-5.38	0.08	242	-5.34	0.10
4	242	-5.02	0.09	1	-5.11		243	-5.02	0.09
5	240	-4.68	0.10	3	-4.64	0.13	243	-4.68	0.10
6	241	-4.35	0.09	1	-4.22		242	-4.35	0.09
7	242	-4.02	0.11	1	-4.06		243	-4.02	0.11
8	240	-3.62	0.13	2	-3.64	0.07	242	-3.62	0.13
9	237	-3.00	0.24	6	-3.01	0.30	243	-3.00	0.24
10	139	-1.25	1.23	103	0.36	1.15	242	-0.57	1.44
Total	2,305	-4.47	1.29	120	-0.28	1.94	2,425	-4.26	1.61

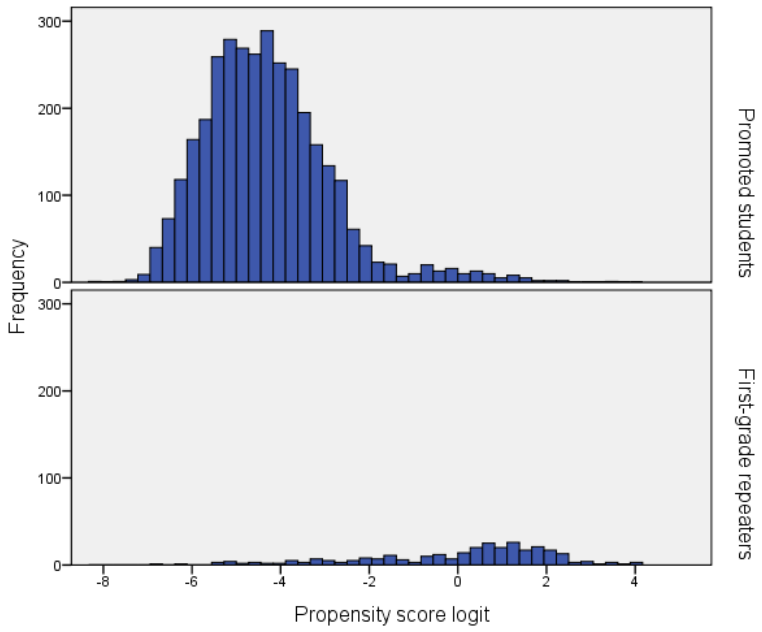

Figure 1. Distribution of propensity score logits among promoted students (above) and first-grade repeaters (below).

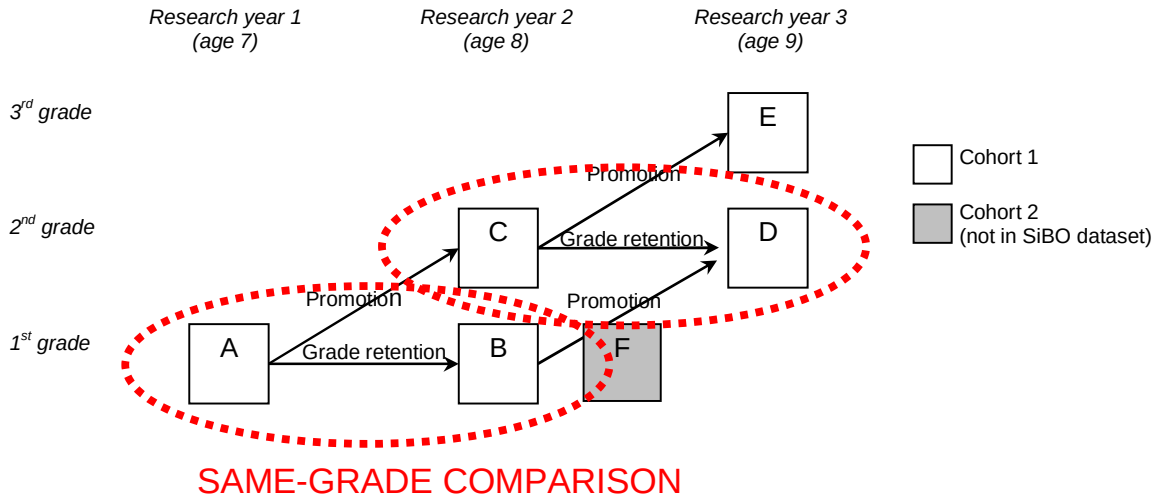

Figure 2. Illustration of same-grade comparisons in the SiBO project.

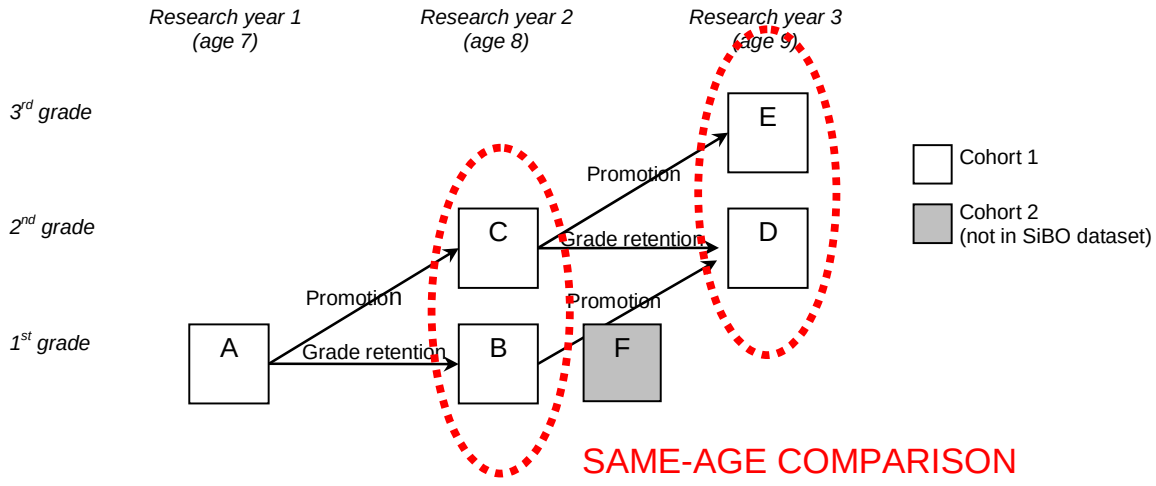

Figure 3. Illustration of same-age comparisons in the SiBO project.

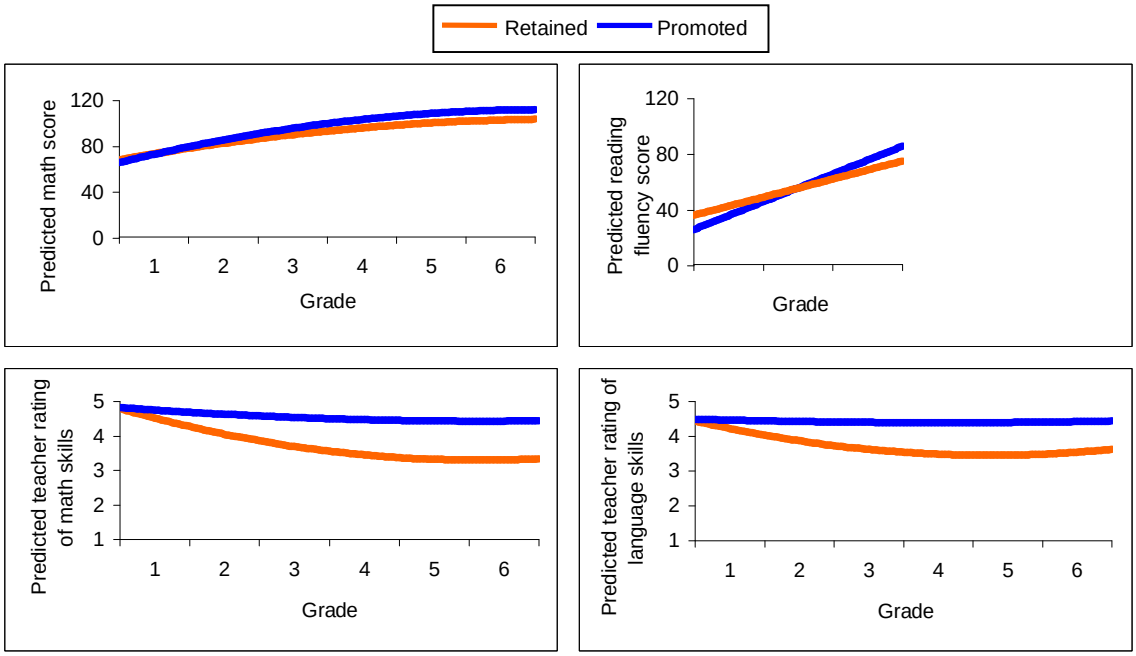

Figure 4. Results of three-level growth curve analyses based on same-grade comparisons. Grade 1 represents the repetition year for first-grade repeaters.

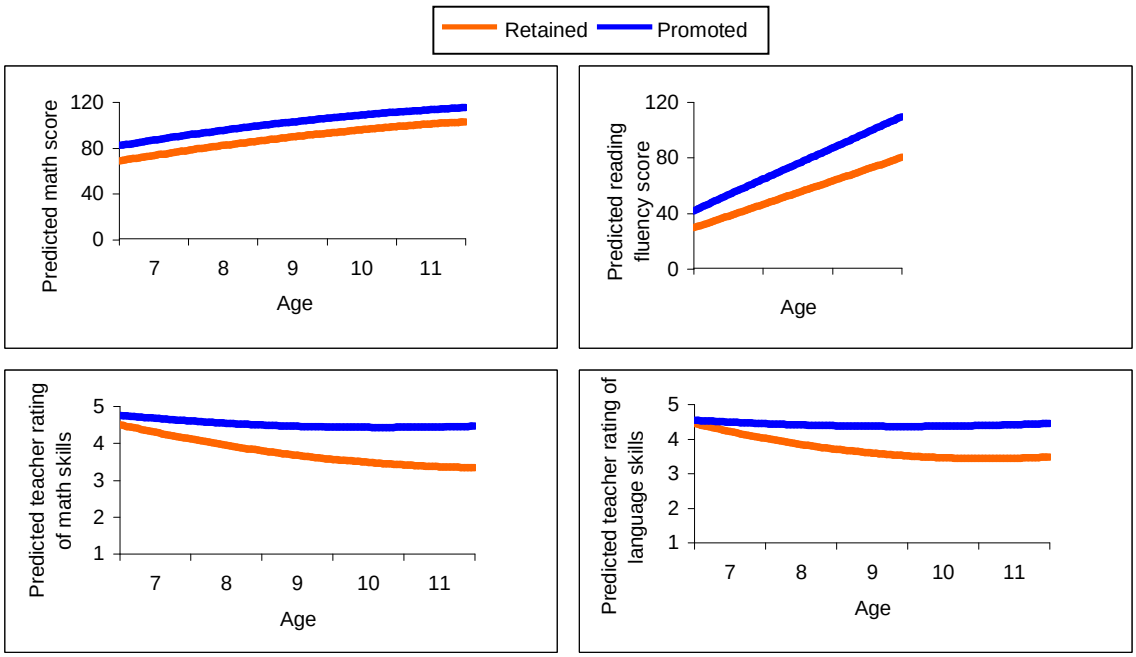

Figure 5. Results of three-level growth curve analyses based on same-age comparisons. Age 7 represents the repetition year for first-grade repeaters.