University of Groningen

The impact of student composition on schools' value-added performance

Timmermans, Anneke C.; Thomas, Sally M.

Published in:
School Effectiveness and School Improvement

DOI:
[10.1080/09243453.2014.957328](https://doi.org/10.1080/09243453.2014.957328)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Timmermans, A. C., & Thomas, S. M. (2015). The impact of student composition on schools' value-added performance: a comparison of seven empirical studies. *School Effectiveness and School Improvement*, 26(3), 487-498. <https://doi.org/10.1080/09243453.2014.957328>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

The impact of student composition on schools' value-added performance: a comparison of seven empirical studies

Anneke C. Timmermans^{a*} and Sally M. Thomas^b

^a*GION Education/Research, University of Groningen, Groningen, The Netherlands;* ^b*Graduate School of Education, University of Bristol, Bristol, UK*

(Received 21 July 2014; final version received 21 July 2014)

In many countries, policy makers struggle with the development of value-added indicators of school performance for educational accountability purposes and in particular with the choice whether school context measured in the form of student composition variables should be included. This study investigates differences between 7 empirical studies that have looked at changes in value-added indicators through the inclusion of school composition variables. Differences were found in the strength of the association between Type A (student-level covariates) and Type B (student- and school-level covariates) school effect models in various regions, varying between 0.74 and 0.99 for secondary education. Important differences were found between the estimate of a school's effectiveness by using a Type A or a Type B value-added model in secondary education in The Netherlands, Belgium, Chile, and Australia and in US and UK primary education. However, few differences were found in other cases.

Keywords: value-added models; school composition effects; educational accountability

Introduction

In many countries, policy makers struggle with the development of indicators measuring school performance for educational accountability purposes. A variety of approaches are used including raw examination scores, adjusted examination scores, and value-added measures (Thomas, Salim, Munoz-Chereau, & Peng, 2012). However, typically only value-added approaches utilize longitudinal data and are generally considered to provide the most valid and appropriate measures of school effectiveness. Therefore, in a world of arguably imperfect measures, there is at least some clear justification for interpreting value-added measures as “an indication of the extent to which any given school has fostered the progress of all students in a range of subjects during a particular time period in comparison to the effects of other schools in the sample” (Sammons, Thomas, & Mortimore, 1997, p. 24). And of course, it is always important to remember that value-added estimates reflect the school effect as well as any other relevant factors not accounted for in the analysis. Yet, in the practice of educational accountability many different value-added indicators are used. One of the many issues in developing value-added indicators is the choice of which control variables should be included in the statistical model in order to control sufficiently for differences between students at entry and differences in school populations. A specific aspect regarding the choice of control variables concerns whether school context in the form of student composition variables

*Corresponding author. Email: a.c.timmermans@rug.nl

should be included (Keeves, Hungi, & Afrassa, 2005). For example, in the English secondary school league tables (published by newspaper media and now also online via the Department for Education website <http://www.education.gov.uk/>) the UK government has shifted over the last 20 years from initially only using raw scores (no control variables) to value-added scores (VA, only student-level control variables) to contextualized value-added scores (CVA, student- and school-level control variables) and then back to value-added scores (VA) but for different groups of students.

The matter of the inclusion of student composition variables is closely related to the different types of school effects as described by Raudenbush and Willms (1995). In the Type A school effect, adjustments are made for prior achievement and student-level background variables (Willms & Raudenbush, 1989). The Type A school effect model therefore answers the question of what the difference is between the performance of students in a particular school and the average school for children with comparable characteristics. The resulting Type A school effect is not only influenced by the quality of the teaching staff and educational practices but might also be influenced by a more positive or negative surrounding of the school or the school's population. Willms and Raudenbush (1989) argue that it would not be fair to compare the performance of schools on the basis of this Type A school effect because the quality of teaching is only a part of the estimated school effect.

Additionally, the Type B school effect was described, in which both student- and school-level context variables are included. Examples of such school-level variables are the average prior achievement and socioeconomic status (SES) of the students in a school as possible indicators of the school composition, as well as any possibly wider social influences such as denomination, school size, and neighbourhood deprivation as school context indicators (Opdenakker & Van Damme, 2006). Conceptually, the Type B school effect isolates the effects of practices and policies of schools on a child's performance and seems the best model for educational accountability (De Fraine, Van Damme, & Onghena, 2002; Timmermans, Doolaard, & De Wolf, 2011). The idea is that schools should not be held accountable for the context in which they operate, but they can be held accountable for effective or ineffective educational practices. One of the requirements for the inclusion of school context variables is that they should be outside the control of the school. However, this is not clear cut given that there may be international differences with respect to the factors that do lie outside the control of the school and in terms of how much contextual measures actually overlap with or reflect the quality of teaching and learning and other key process factors within schools (De Fraine et al., 2002; Thomas & Mortimore, 1996).

The choice of whether these school composition variables need to be included for a valid estimation of value added for educational accountability purposes depends on several factors. In the first place, including more control variables leads to a more complex statistical model that becomes more difficult to understand for stakeholders in the educational accountability context. Second, the choice depends also on the strength of the school compositional effects, also known as "school mix" or intellectual composition of the school. In the case of significant school composition effects, including these variables might change the estimated value added of secondary schools. Otherwise, the inclusion might not lead to a relevant change in the estimated value added and would lead to unnecessary complexity.

An additional complication for developing value-added models in educational accountability is that the strength of school composition effects may vary across countries or school systems. In the case of differences in school composition effects between

countries, clearly the choice concerning the inclusion of school composition variables needs to be based on research within the particular context. A comparison of value-added models across regions and educational systems may therefore assist policy makers in understanding how different aspects of regional or country school systems may potentially influence the estimation of value added. In a previous study where several regions were compared, it appeared that in regions with larger between-school differences in raw scores and a greater extent of student selection between schools, larger differences were found in size of value-added school effects (Thomas, 2001). Leckie and Goldstein (2009) also observed substantial differences between a number of studies in different country contexts and between their own findings from England and those of Thomas, Peng, and Gray (2007) in the estimated stability of time trends in schools' value-added performance and in part attributed this to whether student composition factors were controlled for in the analysis or not.

School composition effects are usually defined as the effects of the student body of a school on the outcomes of students in addition to individual pupil characteristics (Harker & Tymms, 2004; Lauder, Kounali, Robinson, & Goldstein, 2010; McEwan, 2003). This means that pupils with higher prior achievement scores tend to get higher final achievement scores, but also that two pupils with similar achievement prior scores but attending different schools can be predicted to have different final achievement scores depending on the average scores of the other pupils in their schools (Harker & Tymms, 2004). Educational effectiveness research has repeatedly established that school compositional variables explain differences in educational outcomes after taking individual student performance and background characteristics into account (McEwan, 2003; Opdenakker & Van Damme, 2006; Palardy, 2008; Peetsma, Van der Veen, Koopman, & Van Schooten, 2006; Schindler Rangvid, 2007; Van Ewijk & Slegers, 2010a, 2010b), although interestingly this is not always found to be the case as in Thomas and Mortimore's 1996 study.

The current debate focuses on how the school composition effects need to be interpreted. Three common explanations of school composition effects are given that might operate simultaneously, namely, (a) being an artefact of not measuring all relevant student-level variables or measurement error in achievement scores (Harker & Tymms, 2004; Nash, 2003; Thomas & Mortimore 1996; Thrupp, Lauder, & Robinson, 2002); (b) the variation in teaching quality between schools is related to variation in school composition or, in other words, school mix is a proxy for unmeasured school processes (Lauder & Hughes, 1990); or (c) a real social phenomenon operating in schools.

The aim of the current study, however, is to compare several similar studies that investigated differences among value-added models with and without student compositional variables. The focus of this study lies therefore on the impact of the choice of control variables on value-added indicators in educational systems where judgments are made about the performance of schools based on performance or attainment data of students. Furthermore, we aim to investigate whether the impact of including school compositional variables in value-added measures differs between regions.

Selection of studies

For an adequate comparison of studies, several criteria were formulated. The criteria comprised:

- (1) The study should be conducted based on longitudinal data collections (Goldstein, 1997; Thrupp et al., 2002) in secondary or primary education.

- (2) The study should be an empirical comparison of Type A and Type B value-added models.
- (3) Prior achievement and other student-level variables related to student performance need to be available and included in the Type A and Type B school effects models (Aitkin & Longford, 1986; Nash, 2003; Thrupp et al., 2002; Willms, 1992).
- (4) Hierarchical linear models need to be applied to take account of the clustering of students within schools (Aitkin & Longford, 1986; Goldstein, 1997; Snijders & Bosker, 2012; Thrupp et al., 2002).
- (5) The type of statistical models in the study needs to be consistent over value-added models for different types of school effects to prevent distortions through model specifications (Wei, Hembry, Murphy, & McBride, 2012).
- (6) The study needs to report some type of association or difference (correlation or graphically) between the estimations of value added based on the different types of school effects.

Seven studies were found that met these criteria. In five of these studies, correlations are presented between the estimated value added of schools for the different types of effects, and in two other studies differences between the Type A and Type B value-added estimates are presented graphically. A summary of the available studies and reported associations between Type A and Type B school effects are presented in Table 1. Four of the five studies compared various value-added estimates of schools in secondary education (China, UK, Chile, and The Netherlands), and one study compared value-added models in primary education (US).

Differences between value-added models with or without school-level control variables

The correlations of the estimated Type A and Type B effects of schools in secondary education range between 0.74 and 0.99. All in all, these are very high correlations, indicating that generally schools with high estimated Type A school effects tend to also have high estimated Type B school effects. However, especially in the case of developing value-added models for educational accountabilities, even relatively small deviations from a perfect association can have a large impact on specific individual schools.

In the study of Thomas and Mortimore (1996) in the context of secondary education in the Lancashire region (UK), a correlation of .98 between the estimated Type A and B school effects was found. Once adequate account was taken of control variables on the student level, the school-level variables had little impact on the remaining variation in student outcomes. It is argued that in the case when rich student-level data are available, it can be unnecessary (as in their reported findings) but also to some extent inappropriate and misleading to utilize school context variables due to the potential link with school process variables. However, school compositional variables (such as % Free school meals) might be useful as a proxy in the case where pupil intake data are lacking (Thomas & Mortimore, 1996).

Somewhat different conclusions were drawn from the study of Dutch secondary education (Timmermans et al., 2011). In this study, the average SES of the students appeared to be a statistically significant predictor of student performance after taking account of various student-level control variables. In this study, the correlations between Type A and Type B school effects were somewhat smaller and ranged between 0.90 and

Table 1. Overview of the association (Pearson's r) between estimated types of school effects.

Study	Context	Dependent variable	Correlations Type A & Type B	Significant school-/class-level variables
Darmawan and Keeves (2006)	Australian secondary education	Science achievement		Average prior achievement, class size
De Fraine et al. (2002)	Belgium secondary education	Achievement in mathematics and Dutch language		Class average prior achievement, SES, achievement motivation, school level
Munoz (2013)	Chilean secondary education	Achievement in mathematics and language	0.74 (mathematics) 0.88 (language)	Average previous attainment
Peng and Klieme (2014)	Chinese secondary education	Achievement in mathematics, Chinese, and English	0.85–0.99	Average prior achievement and average SES
Tekwe et al. (2004)	US primary education	Mathematics, Reading	0.61–0.95 (mathematics) 0.72–0.87 (reading)	
Thomas and Mortimore (1996)	UK secondary education	GCSE examination results	0.98	None
Timmermans et al. (2011)	Dutch secondary education	Average examination grades	0.90–0.94	Average SES

0.94. In this particular context of Dutch secondary education, it was shown that including compositional variables at the school level in the estimation of value added can have a large impact on the estimation of the effects of individual schools and on the classification of schools as underperforming, average, and overperforming.

More variation in correlations between Type A and Type B school effects appeared in a study of Chinese secondary education (Peng & Klieme, 2014). The association between the Type A and Type B effects ranged between 0.85 in the 1st year of secondary education to 0.99 in the 3rd year of secondary education. The average prior achievement and average SES of students were only statistically significant predictors in the estimation of a school's value added for the 1st year. This latter study deviated from the previous ones as in this study value added was estimated for separate years in secondary education, while the first studies (Thomas & Mortimore, 1996; Timmermans et al., 2011) investigated value added based on the entire length of secondary education.

Furthermore, a recent study was conducted in secondary education in Chile (Munoz, 2013). Several value-added models were investigated for a cohort of students that took the tests from the Education Quality Measurement System (SIMCE) in Grades 8 and 10 in 2004 and 2006. The correlation between the Type A and Type B value-added estimates of secondary schools in Chile seemed somewhat weaker than in the aforementioned studies, namely 0.74 for mathematics and 0.88 for language. From this study, it was shown that the choice of type of value-added model will lead to empirical differences, for example, in the classifications of individual schools as poor, regular, satisfactory, or good. The main conclusion from this study was that choice of which value-added model to use to judge school performance in Chile is a crucial technical as well as policy decision that may have practical implications for nearly all schools.

A final study in which a comparison of various value-added models was made and that reported correlation is situated in the context of US primary education (Tekwe et al., 2004). The correlations between the schools' estimated Type A and Type B school effect range between 0.61–0.95 for Mathematics and between 0.72–0.87 for Reading. In general, it seems that these lower correlations indicate that the choice of control variables matters even more in primary education.

Not all studies provided correlations amongst various value-added models, but two studies provided other means to compare the differences between the estimates of Type A and Type B school effects. The first of these studies is conducted in the context of Australian secondary education (Darmawan & Keeves, 2006). In Figure 1, several types of school effects are presented. From this figure, it appears that the choice of control variables can have a big impact on the estimation of the effect of individual schools. Furthermore, the difference between the Type A and Type B effect can range up to 0.5 point on the residual scale.

The second study was conducted in the context of Belgian secondary education. In Figure 2 (De Fraine et al., 2002), the triangles and associated confidence intervals represent the Type A effects for a sample of classes. The dots in the figure represent the estimates of the Type B class effects. For some classes, the Type A and Type B effects are very similar, as can be seen by a small distance between the point estimates of the Type A and Type B class effects. However, for a substantial group of classes the differences between the estimated Type A and Type B effect are relatively large. The final conclusion of this particular study was that the effects of classes and schools are to a certain extent due to group composition.

Figure 1. Types A, B, X, and Z school residuals (Darmawan & Keeves, 2006).

Note: Reprinted from “Accountability of Teachers and Schools: A Value-Added Approach,” by I. G. N Darmawan and J. P. Keeves, 2006, *International Education Journal*, 7, pp. 174–188.

Figure 2. Class residuals for mathematics (De Fraine et al., 2002).

Note: Reprinted from “Accountability of Schools and Teachers: What Should Be Taken into Account?,” by B. De Fraine, J. Van Damme, and P. Onghena, 2002, *European Educational Research Journal*, 1, pp. 403–428.

Besides the previous seven studies, an eighth study (Lauder et al., 2010) might be worthwhile in this particular context although it did not fit within the previously described inclusion criteria, because no Type A and Type B value-added models were compared. This particular study investigated in UK primary education the difference between the well-known contextualized value added (CVA), which can be considered a Type B value-added model, and a more extended Type B value-added model in which more detailed information on the socioeconomic background of the students and schools was included. The results show small composition effects of the social class composition of schools on the reading performance of students and small positive composition effects of the prior achievement composition on both reading and mathematics. Maybe the most important finding of the study is that the inclusion of the more detailed compositional variables quite clearly changed the rank order of primary schools.

Discussion

The aim of the current study was to investigate differences between countries in school composition effects on the performance of students, for the particular purpose of informing the development of any value-added model for educational accountability. The studies presented concern longitudinal studies from various countries and educational systems, such as the UK, US, Belgium, Australia, China, Chile, and The Netherlands. Differences were found in the strength of the association between Type A and Type B school effect models in various regions, varying between 0.74 and 0.99 in secondary education. Important differences were found between the estimate of a school's effectiveness by using a Type A or a Type B value-added model in secondary education in The Netherlands, Belgium, Chile, and Australia and in primary education in the US and UK. In these studies, the average prior achievement and average SES were the most common contextual variables that significantly predicted student performance after taking account of student-level variables. In secondary education in Lancashire (UK) and the second and third grades of secondary education in China, hardly any differences were found between a Type A or a Type B value-added model.

Consequently, differences were found in the conclusions drawn from these seven studies. In general, the studies that found significant school composition effects conclude that school effects might partly be due to group composition differences and that is important to include school composition variables if schools are compared in performance, recognizing difficulties that arise from including these variables (De Fraine et al., 2002; Lauder et al., 2010; Timmermans et al., 2011). In the studies in which no significant school composition effects were found, the general conclusion is not to include school compositional variables in the estimation of value added given the drawbacks of including school composition. These previous findings from the comparison of seven empirical studies imply that the choice concerning the inclusion of school composition variables needs to be based very carefully on the stated purpose of the resulting measures and on research within the particular context.

In interpreting the differences between the studies, several limitations need to be kept in mind. In case of differences between value-added models based on the types of school effects, there can be several explanations that relate to methodological differences between the studies. In the first place, the studies differ in the number and type of covariates used in the value-added analysis of school effects. The differences in sets of control variables at the student level could arguably account for differences in the association found between Type A and Type B school effects. This difference in student-level control variables implies that the explanation that a significant school compositional variable may be an artefact of not measuring all relevant student-level variables may be more plausible in some studies than in other studies. Similarly, the studies differ in the number and method of measuring school composition. For example, in the studies of Lauder et al. (2010) and Timmermans et al. (2011), multiple school composition variables were included in the model, related both to the prior achievement and the social economic composition and including both measurements of the average composition as well as indicators of mix in composition (homogeneous versus heterogeneous). Furthermore, the studies differ in the dependent variable that is used to estimate the value added of schools. The study of Thomas and Mortimore (1996) and the study of Timmermans et al. (2011) use scores of students at the final examination of secondary education. Other studies have used the achievement of students as measured on standardized tests, in some cases because there are no national examinations (De Fraine et al., 2002).

Second, the studies are conducted in the context of very different educational systems. The most prominent difference between educational systems is comprehensive versus tracked. The studies of De Fraine et al. (2002) and Timmermans et al. (2011) are based on a sample of students in one or two separate tracks, whereas the other studies are conducted in comprehensive educational systems. The studies in tracked systems therefore give information on the effectiveness of a school for a specific group of students, namely, only those students in the tracks analysed, which may tend to inflate reported associations among various value-added models. This is because one of the requirements of investigating student composition effects as described by Thrupp and colleagues (2002) is that samples need to be used in which the full spectrum of between-school differences in composition is present, because school compositional effects might only arise under specific circumstances. Whether statistically significant effects of school context or school compositional variables are found may depend on the range and extent of differences between schools in school composition. This may be more likely in regions where selection policies are employed by a minority of schools (Sammons et al., 1997; Thomas, 2001) or in large urban regions where school choice is a real option for many families (Harker & Tymms, 2004). Whether the datasets analysed in the studies of interest meet this requirement is not clear. However, in the studies in tracked educational systems where only a specific track is investigated it might be questionable whether the full spectrum of school compositional differences was present.

Although this current study is specifically based on studies that conducted empirical comparisons of value-added models, there are some similarities with previous meta-analysis on school compositional effects. Results from meta-analyses indicate that school compositional effects for ethnic diversity seem to vary across countries or contexts (Van Ewijk & Sleegers, 2010b); however, only little evidence was found for differences in effect sizes of the socioeconomic composition of schools between countries (Van Ewijk & Sleegers, 2010a). In a review of the literature concerning socioeconomic school composition effects by Peetsma and colleagues (2006), inconsistent findings are highlighted, suggesting that several studies do find SES school composition effects, but several other studies failed to find this relationship. In a review by Thrupp and colleagues (2002), it was concluded that, with respect to school composition in terms of the student's prior achievement, there was remarkably little consensus on the nature and size of school composition effects (Thrupp et al., 2002). Research has also noted that the direction of statistically significant school compositional effects on student achievement is not always consistent. Many studies have reported positive or otherwise positive but nonsignificant effects (Munoz, 2013; Thomas & Mortimore, 1996; Timmermans et al., 2011), while others have indicated a negative impact especially on non-academic student outcomes – sometimes referred to as the “big-fish-little-pond” effect (Marsh & O'Mara, 2010; Marsh, Trautwein, Lüdtke, Baumert, & Köller, 2007). Given these findings from reviews, it might not be surprising that studies that investigated differences between Type A and Type B value-added models resulted in inconsistent findings.

Besides the methodological issues as outlined above, several other explanations might shed light on inconsistent findings of school composition effects. In the first place, school practices and policies may be related to a greater or lesser extent to school composition possibly due to the extent of additional public funding or resources provided by wealthier parents. Second, there might be a joint effect on student performance (Iturre, 2005; Opdenakker & Van Damme, 2001; Thrupp et al., 2002), and the inclusion of such contextual variables in a value-added model might therefore remove some of the joint

effects of school composition and school policy and practice (Opdenakker & Van Damme, 2007; Willms, 1992). The inclusion of school compositional variables in value-added analyses therefore is likely to lead to an underestimation of the real differences in effectiveness between schools (De Fraine et al., 2002; Timmermans et al., 2011).

There are no easy solutions to the dilemma of whether school context factors should or should not be controlled for in value-added indicators, and value-added measures can never provide a perfect instrument for accountability. There will always be risks in either case. First, if school context is controlled for, it may be more difficult to identify where good or poor teaching and learning and other key processes are happening in schools. Second, if school context is not controlled for, schools with highly disadvantaged intakes may not be treated fairly because school effectiveness and school improvement are limited by the initial nature of the schools' intake but also by how the students in the school subsequently interact with one another and with the general context of their schools (Thrupp et al., 2002). In this case, the responsibility of the schools for the performance of students is less clear.

In conclusion, although some regional differences appear from this comparison and in some regions the school context can be considered appropriate to be included in value-added measures, the estimation of Type B school effects is clearly not without problems. This underlines the need for careful consideration by policy makers of relevant local research that uses both qualitative and quantitative empirical approaches – to consult local stakeholders on methodology and in selecting value-added indicators for accountability systems. Even more care is required in using such indicators for high-stakes purposes such as identifying “failing” schools, especially in the absence of any other evidence.

References

- Aitkin, M., & Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 149, 1–43. <http://www.jstor.org/stable/2981882>
- Darmawan, I. G. N., & Keeves, J. P. (2006). Accountability of teachers and schools: A value-added approach. *International Education Journal*, 7, 174–188.
- De Fraine, B., Van Damme, J., & Onghena, P. (2002). Accountability of schools and teachers: What should be taken into account? *European Educational Research Journal*, 1, 403–428. doi:10.2304/eeerj.2002.1.3.2
- Goldstein, H. (1997). Methods in school effectiveness research. *School Effectiveness and School Improvement*, 8, 369–395. doi:10.1080/0924345970080401
- Harker, R., & Tymms, P. (2004). The effects of student composition on school outcomes. *School Effectiveness and School Improvement*, 15, 177–199. doi:10.1076/sesi.15.2.177.30432
- Iturre, R. A. C. (2005). The relationship between school composition, school process and mathematics achievement in secondary education in Argentina. *International Review of Education*, 51, 173–200. doi:10.1007/s11159-005-1843-7
- Keeves, J. P., Hungi, N., & Afrassa, T. (2005). Measuring value added effect across schools: Should schools be compared in performance? *Studies in Educational Evaluation*, 31, 247–266. doi:10.1016/j.stueduc.2005.05.012
- Lauder, H., & Hughes, D. (1990). Social inequalities and differences in school outcomes. *New Zealand Journal of Educational Studies*, 25, 37–60.
- Lauder, H., Kounali, D., Robinson, T., & Goldstein, H. (2010). Pupil composition and accountability: An analysis of English primary schools. *International Journal of Educational Research*, 49, 49–68. doi:10.1016/j.ijer.2010.08.001
- Leckie, G., & Goldstein, H. (2009). The limitations of using school league tables to inform school choice. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172, 835–851. doi:10.1111/j.1467-985X.2009.00597.x

- Marsh, H. W., & O'Mara, A. J. (2010). Long-term total negative effects of school-average ability on diverse educational outcomes. *Zeitschrift für Pädagogische Psychologie*, 24, 51–72.
- Marsh, H. W., Trautwein, U., Lüdtke, O., Baumert, J., & Köller, O. (2007). The big-fish-little-pond effect: Persistent negative effects of selective high schools on self-concept after graduation. *American Educational Research Journal*, 44, 631–669.
- McEwan, P. J. (2003). Peer effects on students achievement: Evidence from Chile. *Economics of Education Review*, 22, 131–141. doi:[10.1016/S0272-7757\(02\)00005-5](https://doi.org/10.1016/S0272-7757(02)00005-5)
- Munoz, B. (2013). *Value added measures of school performance in Chile* (Unpublished doctoral dissertation). University of Bristol, Bristol, UK.
- Nash, R. (2003). Is the school composition effect real? A discussion with evidence from the UK PISA data. *School Effectiveness and School Improvement*, 14, 441–457. doi:[10.1076/sesi.14.4.441.17153](https://doi.org/10.1076/sesi.14.4.441.17153)
- Opdenakker, M.-C., & Van Damme, J. (2001). Relationship between school composition and characteristics of school process and their effect on mathematic achievement. *British Educational Research Journal*, 27, 407–432. doi:[10.1080/0141192010071434](https://doi.org/10.1080/0141192010071434)
- Opdenakker, M.-C., & Van Damme, J. (2006). Differences between secondary schools: A study about school context, group composition, school practice, and school effects with special attention to public and Catholic schools and type of schools. *School Effectiveness and School Improvement*, 17, 87–117. doi:[10.1080/09243450600264457](https://doi.org/10.1080/09243450600264457)
- Opdenakker, M.-C., & Van Damme, J. (2007). Do school context, student composition and school leadership affect school practice and outcomes in secondary education? *British Educational Research Journal*, 33, 179–206. doi:[10.1080/01411920701208233](https://doi.org/10.1080/01411920701208233)
- Palardy, G. J. (2008). Differential school effects among low, middle, and high social class composition schools: A multiple group, multilevel latent growth curve analysis. *School Effectiveness and School Improvement*, 19, 21–49. doi:[10.1080/09243450801936845](https://doi.org/10.1080/09243450801936845)
- Peetsma, T., Van der Veen, I., Koopman, P., & Van Schooten, E. (2006). Class composition influences on pupils' cognitive development. *School Effectiveness and School Improvement*, 17, 275–302. doi:[10.1080/13803610500480114](https://doi.org/10.1080/13803610500480114)
- Peng, P., & Klieme, E. (2014). *One-year value-added school effects from various models and their inter-temporal variability: Evidence from China*. Unpublished manuscript. Retrieved from www.invalsi.it/invalsi/ri/improving_education/Papers/peng/151.doc
- Raudenbush, S. W., & Willms, J. D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, 20, 307–335. doi:[10.3102/10769986020004307](https://doi.org/10.3102/10769986020004307)
- Sammons, P., Thomas, S., & Mortimore, P. (1997). *Forging links: Effective schools and effective departments*. London: Paul Chapman.
- Schindler Rangvid, B. (2007). School composition effects in Denmark: Quantile regression evidence from PISA 2000. *Empirical Economics*, 33, 359–388. doi:[10.1007/s00181-007-0133-6](https://doi.org/10.1007/s00181-007-0133-6)
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.
- Tekwe, C. D., & Carter, R. L., Ma, C.-X., Algina, J., Lucas, M. E., Roth, J., ... Resnick, M. B. (2004). An empirical comparison of statistical models for value-added assessment of school performance. *Journal of Educational and Behavioral Statistics*, 29, 11–36. doi:[10.3102/10769986029001011](https://doi.org/10.3102/10769986029001011)
- Thomas, S. (2001). Dimensions of secondary school effectiveness: Comparative analyses across regions. *School Effectiveness and School Improvement*, 12, 285–322. doi:[10.1076/sesi.12.3.285.3448](https://doi.org/10.1076/sesi.12.3.285.3448)
- Thomas, S., & Mortimore, P. (1996). Comparison of value-added models for secondary-school effectiveness. *Research Papers in Education*, 11, 5–33. doi:[10.1080/0267152960110103](https://doi.org/10.1080/0267152960110103)
- Thomas, S., Salim, M., Munoz-Chereau, B., & Peng, W.-J. (2012). Educational quality, effectiveness and evaluation: Perspectives from China, South America and Africa. In C. Chapman, P. Armstrong, A. Harris, D. Muijs, D. Reynolds, & P. Sammons (Eds.), *School effectiveness and improvement research, policy and practice: Challenging the orthodoxy?* (pp. 125–145). London: Routledge.
- Thomas, S. M., Peng, W.-J., & Gray, J. (2007). Modelling patterns of improvement over time: Value added trends in English secondary school performance across ten cohorts. *Oxford Review of Education*, 33, 261–295.
- Thrupp, M., Lauder, H., & Robinson, T. (2002). School composition and peer effects. *International Journal of Educational Research*, 37, 483–504. doi:[10.1016/S0883-0355\(3\)0016-8](https://doi.org/10.1016/S0883-0355(3)0016-8)

- Timmermans, A. C., Doolaard, S., & De Wolf, I. (2011). Conceptual and empirical differences among various value-added models for accountability. *School Effectiveness and School Improvement*, 22, 393–413. doi:[10.1080/09243453.2011.590704](https://doi.org/10.1080/09243453.2011.590704)
- Van Ewijk, R., & Sleegers, P. (2010a). The effect of peer socioeconomic status on achievement: A meta-analysis. *Educational Research Review*, 5, 134–150. doi:[10.1016/j.edurev.2010.02.001](https://doi.org/10.1016/j.edurev.2010.02.001)
- Van Ewijk, R., & Sleegers, P. (2010b). Peer ethnicity and achievement: A meta-analysis into the compositional effect. *School Effectiveness and School Improvement*, 21, 237–265. doi:[10.1080/09243451003612671](https://doi.org/10.1080/09243451003612671)
- Wei, H., Hembry, T., Murphy, D. L., & McBride, Y. (2012). *Value-added models in the evaluation of teacher effectiveness: A comparison of models and outcomes* (Pearson Research Report). Retrieved from <http://images.pearsonassessments.com/images/tmrs/ComparisonofValue-AddedModelsandOutcomes.pdf>
- Willms, J. D. (1992). *Monitoring school performance: A guide for educators*. Washington, DC: The Falmer Press.
- Willms, J. D., & Raudenbush, S. W. (1989). A longitudinal hierarchical linear model for estimating school effects and their stability. *Journal of Educational Measurement*, 26, 209–232.