



## Malleability in educational effectiveness: what are realistic expectations about effect sizes? Introduction to the special issue

Jaap Scheerens & Gary N. Marks

To cite this article: Jaap Scheerens & Gary N. Marks (2018): Malleability in educational effectiveness: what are realistic expectations about effect sizes? Introduction to the special issue, Educational Research and Evaluation, DOI: [10.1080/13803611.2017.1455280](https://doi.org/10.1080/13803611.2017.1455280)

To link to this article: <https://doi.org/10.1080/13803611.2017.1455280>



© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 10 Apr 2018.



Submit your article to this journal [↗](#)



Article views: 77



View related articles [↗](#)



View Crossmark data [↗](#)

# Malleability in educational effectiveness: what are realistic expectations about effect sizes? Introduction to the special issue

Jaap Scheerens<sup>a</sup> and Gary N. Marks<sup>b</sup>

<sup>a</sup>University of Twente, Enschede, The Netherlands; <sup>b</sup>Office of Government, Policy and Strategy, Australian Catholic University, Fitzroy, VIC, Australia

## ABSTRACT

Educational effectiveness research separates hypothetical causes of performance differences into “given”, “contextual”, “endogenous”, or simply “prior” conditions, on the one hand, and malleable factors, or treatments, on the other hand. Recent studies indicate that the effects of background conditions tend to be bigger, and those of malleable variables and interventions smaller, than usually expected. These findings give reason to pose “limited malleability” as the central hypothesis of the special issue. This hypothesis is addressed in the 5 articles that make up this special issue. The themes addressed in these articles are respectively: optimizing the choice of adjustment variables, the development of a nomological network of educational achievement at country level, the stability of system-level educational performance, modelling approaches to the estimation of size, stability, and consistency of school effects, and treatment effects in schooling. The final article makes up the balance on the “limited malleability” thesis and discusses implications for educational policy and practice.

## KEYWORDS

Educational effectiveness; school effects; adjusting for prior achievement; system-level policy-amenable and contextual factors; treatment effects

## Introduction

A key issue in educational effectiveness research is separating the effects of malleable, policy-amenable factors on student achievement from those of “given” background conditions at system and student level. Recent studies indicate that the effects of background conditions tend to be bigger, and those of malleable variables and interventions smaller, than usually expected (Marks, 2016; Scheerens, 2016). This combination of findings, a relatively strong influence of given background conditions of various kinds and a relatively weak influence of malleable, policy-amenable variables, gives reason to pose “limited malleability” as the central hypothesis of this special issue. Prior achievement has a much stronger relationship with subsequent student performance than is often acknowledged. The correlations of prior achievement with subsequent achievement or the standardized effects of prior test scores are between 0.6 and 0.8 (Aubrey, Dahl, & Godfrey, 2006, p. 35;

**CONTACT** Jaap Scheerens  j.scheerens@utwente.nl  University of Twente, Enschede, The Netherlands

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Duckworth, Quinn, & Tsukayama, 2012, p. 443; Kriegbaum, Jansen, & Spinath, 2015; Lu & Rickard, 2014, p. 32; Marks, 2014; Parsons, 2014, p. 36; Reynolds & Walberg, 1992). These correlations are far too high to be ignored and will limit the estimates for malleable educational factors relating to schools, teachers, or programmes. Scheerens (2012, 2014) conducted a series of meta-analyses in which the effect sizes of factors considered to enhance educational effectiveness, such as instructional time, frequent evaluation, and educational leadership, appeared to be, on average, no higher than .10. In analyses based on international studies, the success of often considered policy-amenable variables, such as increased school autonomy, facets of evaluation and accountability, and free school choice is also disappointing (Scheerens, 2016). Comparative studies also show that, generally, the performances of educational systems are quite stable across time, and that reforms take long to have effects, if they have effects at all (Scheerens, (2016). In-depth studies of individual systems and comparative case studies suggest that this stability is due to historically developed structural characteristics of education systems and cultural conditions (Sahlgren, 2015).

This special issue consists of papers that address the theme of “limited malleability” at different aggregation levels (national systems and schools) and focus on different treatment variables and control variables. The results of the various contributions are discussed in a final paper (Scheerens, this issue), in which the results are compared to current high expectations of educational reforms, as expressed by international agencies and consultancy firms.

## Scope of the special issue

Malleability in educational effectiveness research is the extent that treatments and effectiveness-enhancing conditions “in the field” affect student performance. The limited malleability thesis is that the “net” impact of school systems, schools, teachers, and programmes is limited in comparison to non-policy-amenable student background variables and contextual conditions. The thesis includes the impact of school and teacher characteristics, school and teacher practices, and system-level reforms.

This special issue is placed broadly in the research field of educational effectiveness, which encompasses system-level, school-level, and teacher/teaching-level effectiveness. At the higher aggregation levels (national system, school), experimental intervention studies are generally not possible, so studies frequently depend on “naturally” occurring variation in educational studies. This is reflected in the four empirical studies in this special issue, which are all “correlational”. Various methods exist to separate the influence of malleable variables (“treatment”) and “given” contextual conditions that have a positive influence on student achievement. Empirical studies using different methods to adjust for contextual variables yield effect sizes of varying size.

The four empirical studies in this special issue examine the importance of contextual variables at system and school level, and shed light on implications for the scope of malleable variables to affect performance. The fifth and concluding article connects the above issues in determining “net” school effects to considerations about measuring treatment variables, study characteristics of school effectiveness research, and discussion of the state of substantive knowledge in this field. Implications for educational policy and practice and future research are considered as well.

## Overview of contents

The first paper, by Marks, addresses the question of optimizing the choice of adjustment variables in determining value-added school effects. As expected, the estimates of effect sizes differ depending on the adjustment variables. For primary school, across the different domains adjusting for prior achievement is sufficient since the addition of prior achievement in other domains or general aptitude makes little difference to the effect sizes and the distribution of school effects. One possible exception is “writing”, in which student performance is less reliably measured than in the other domains. For reading, writing, and grammar in secondary school, it appears that the most appropriate model uses a combination of same-domain prior achievement and a measure of more general scholastic aptitude as adjustment variables. In contrast, for the analysis of numeracy and spelling in secondary school, the additional covariates do not substantially change the estimated effect sizes or the distribution of school effects. The paper incorporates studies from behavioural genetics to account for these differences.

The study described in the second paper, by He, Van de Vijver, and Kulikova, establishes a nomological network of educational achievement at the country level, with clusters of country-level variables derived from psychological, sociological, and other country-comparative research. Achievement data were compiled from all cycles of the Programme for International Student Achievement (PISA) and the Trends in International Mathematics and Science Study (TIMSS) for Grade 4 and Grade 8 students. The clusters of country-level characteristics relate to country affluence, diversity, intelligence, cultural orientations (on the basis of taxonomies from Hofstede, 2009; House, Hanges, Javidan, Dorfman, & Gupta, 2004; Inglehart, Basafiez, Diez-Medrano, Halman, & Luijkx, 2004; Schwartz, 2009) and teacher self-reports. Some patterns of correlations generally conformed to theoretical expectations and earlier research, for example, a positive association between country-level indicators of affluence indicators and country-level achievement. Contrary to expectations, country-level conscientiousness, one of the Big Five personality traits, had a negative association with student achievement. In comparison to mainstream educational effectiveness research, the paper by He et al. addresses a wide range of country-level characteristics including structural and cultural characteristics that are outside the range of policy levers within the educational province, such as curriculum characteristics and accountability arrangements. Structural country-level characteristics are indirectly malleable, since they are a function of national economic policies (e.g., the indicators associated with national affluence). In contrast, culture characteristics are not directly malleable in educational policy or indirectly influenced by macro-economic policies. To the degree that such contextual system-level variables affect student achievement, they can be seen as limiting the scope for malleable educational policy variables to explain variance in country-level educational achievement.

The article by Aloisi and Tymms is based on a study that sought to contrast the ability of policy-malleable variables to affect PISA scores to that of non-policy-malleable variables. Country-level student performance and non-policy malleable variables were analysed using data from six waves of the PISA study (2001–2015). The focus was on “curriculum” as an educational policy-malleable variable. The core quantitative analyses analysed mathematical literacy. County-level performance across waves was analysed by multilevel growth-curves techniques. Three research questions are addressed:

- (1) What is the relationship between changes in the socioeconomic and demographic characteristics of the PISA cohorts, and changes in country outcomes?
- (2) What is the relationship between changes in the curricular provision of PISA-participating countries and their outcomes?
- (3) What is the relative effect of non-policy-malleable factors (student SES and demographics) on PISA scores, compared to policy-malleable factors (curricular changes)?

The results of this study speak directly to the central theme of this special issue: the degree to which contextual variables, relative to policy-malleable variables, influence student performance. The results were obtained from international longitudinal data, allowing for insight into the stability over time of gross and adjusted performance. Main outcomes of this study address the stability of country-level performance over time, the relatively strong influence of contextual conditions, not directly malleable by means of educational policy, and the effect of curriculum change, which was very small with an effect size of 0.02. The relative performance of countries appeared to be extremely stable over the period of PISA testing 2000 to 2015.

The paper by Timmermans and Van der Werf analyses differences in effectiveness between schools by using the learning gain of students over three points in time by means of growth curve modelling. The empirical data that were used in the study are from a Dutch data set which includes students' scores on reading comprehension, spelling, and mathematics tests, taken in Grades 4, 5, and 6 across three cohorts. Students in each cohort were followed for 3 consecutive years (from Grade 4 until Grade 6, age approximately 9–12 years). Within each domain, the students' test results on each particular grade-specific test were calibrated to the other grade-specific tests by means of item response models, more specifically, by means of the one parameter logistic model, which assumes a one-dimensional underlying latent scale per domain. Gross and value-added school effects were estimated within the same multivariate multilevel growth curve model. The results indicated considerably larger value-added school effects than are usually found by means of covariance adjustment models, while the stability and consistency of school effects were not high, which is consistent with other studies. These outcomes stimulate further discussion about the meaning and comparability of effect sizes in terms of growth as compared to "adjusted performance level".

The concluding article of this special issue, by Scheerens, starts out with a review of definitions and operational criteria of school-effect measures. The different ways to estimate school effects depend on the way "gross" school effects are adjusted to what is usually referred to as "value-added" effects. In most applications, "value-added" school effects are adjusted performance levels, but in other cases progress or growth in achievement over time. The article also brings in substantive research outcomes from individual studies and meta-analyses, to conclude on the magnitude of treatment effects. The conclusion is that the most suitable adjustment variables, for example, prior achievement and general intelligence or aptitude, generally produce relatively small value-added or "net" school effects. The implication of this finding is that there is limited scope for effectiveness-enhancing factors when the margins for malleability are so small. Complicating factors in assessing treatment effects include study characteristics such as the nature of the test, sample size, and research design. Such study characteristics might partly explain the rather divergent results from meta-analyses focusing on similar

effectiveness-enhancing conditions. Diversity and questionable quality of treatment measures are discussed as additional challenges for reliably assessing treatment effects of schooling. The discussion section considers implications of small treatment effects and limited malleability for policy and research.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## References

- Aubrey, C., Dahl, S., & Godfrey, R. (2006). Early mathematics development and later achievement: Further evidence. *Mathematics Education Research Journal*, 18(1), 27–46. doi:10.1007/BF03217428
- Duckworth, A. L., Quinn, P. D., & Tsukayama, E. (2012). What No Child Left Behind leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades. *Journal of Educational Psychology*, 104(2), 439–451. doi:10.1037/a0026280
- Hofstede, G. (2009). Dimension data matrix. Retrieved February 3, 2001, from <http://www.geerthofstede.eu/dimension-data-matrix>
- House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (Eds.). (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Thousand Oaks, CA: Sage.
- Inglehart, R., Basafiez, M., Diez-Medrano, J., Halman, L., & Luijkx, R. (Eds.). (2004). *Human beliefs and values: A cross-cultural sourcebook based on the 1999–2002 values surveys*. Mexico City, Mexico: Siglo Veintiuno Editores.
- Kriegbaum, K., Jansen, M., & Spinath, B. (2015). Motivation: A predictor of PISA's mathematical competence beyond intelligence and prior test achievement. *Learning and Individual Differences*, 43, 140–148. doi:10.1016/j.lindif.2015.08.026
- Lu, L., & Rickard, K. (2014). *Value added models for NSW government schools*. Sydney: Centre for Education Statistics and Evaluation, Education and Communities.
- Marks, G. N. (2014). Demographic and socioeconomic inequalities in student achievement over the school career. *Australian Journal of Education*, 58(3), 223–247. doi:10.1177/0004944114537052
- Marks, G. N. (2016). Explaining the substantial inter-domain and over-time correlations in student achievement: The importance of stable student attributes. *Educational Research and Evaluation*, 22(1–2), 45–64. doi:10.1080/13803611.2016.1191359
- Parsons, S. (2014). *Childhood cognition in the 1970 British Cohort Study*. London: Centre for Longitudinal Studies, Institute of Education, University of London
- Raudenbush, S. W., & Willms, J. D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, 20(4), 307–335. doi:10.3102/10769986020004307
- Reynolds, A. J., & Walberg, H. J. (1992). A process model of mathematics achievement and attitude. *Journal for Research in Mathematics Education*, 23(4), 306–328. doi:10.2307/749308
- Sahlgren, G. H. (2015). *Real Finnish lessons: The true story of an education superpower*. London: Centre for Policy Studies.
- Scheerens, J. (Ed.). (2012). *School leadership effects revisited: Review and meta-analysis of empirical studies*. Dordrecht: Springer.
- Scheerens, J. (Ed.). (2014). *Effectiveness of time investments in education: Insights from a review and meta-analysis*. Dordrecht: Springer.
- Scheerens, J. (2016). *Educational effectiveness and ineffectiveness: A critical review of the knowledge base*. Dordrecht: Springer.
- Schwartz, S. H. (2009). Culture matters: National value cultures, sources, and consequences. In R. S. Wyer, C.-y. Chiu, & Y.-y. Hong (Eds.), *Understanding culture: Theory, research, and application* (pp. 127–150). New York, NY: Psychology Press.