

On the effectiveness and efficiency of educational innovations

Citation for published version (APA):

de Witte, K. (2016). *On the effectiveness and efficiency of educational innovations*. Maastricht University. <https://doi.org/10.26481/spe.20160324kw>

Document status and date:

Published: 24/03/2016

DOI:

[10.26481/spe.20160324kw](https://doi.org/10.26481/spe.20160324kw)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Unspecified

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

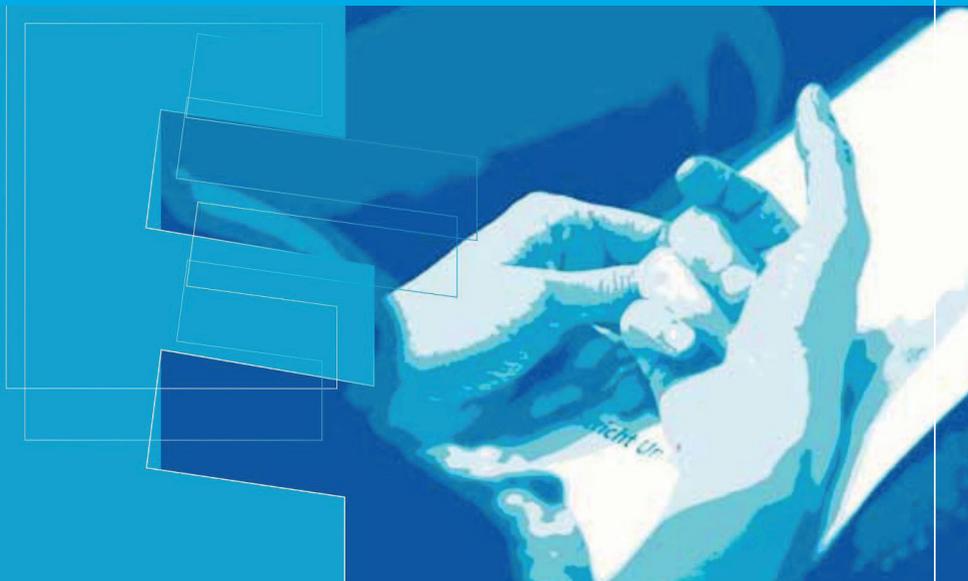
repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Prof. dr. Kristof De Witte

Faculty of Humanities and Sciences

On the effectiveness and efficiency of educational innovations



On the Effectiveness and Efficiency of Educational Innovations

ISBN: 978-90-5681-458-8

NUR: 840

All rights reserved. No part of this publication may be reproduced, modified, stored in a retrieval system or made public without the prior written permission of the author or publisher.

On the Effectiveness and Efficiency of Educational Innovations

Inaugural lecture

presented in an abbreviated version upon the acceptance of the chair 'Effectiveness and Efficiency of Educational Innovations' in the Faculty of Humanities and Sciences, Maastricht University

Maastricht, March 24, 2016

By Prof. dr. Kristof De Witte

Mijnheer de Rector Magnificus, waarde collega's, beste vrienden

Despite the fact that we organize education in a very similar way as decades ago, there are daily innovations in education. Examples of innovations at the micro-level include teachers who experiment with new didactical tools (De Witte and Schelfhout, 2015), who innovate with differentiated instruction (Casteleins et al., 2016), or who train the meta-cognitive abilities of students. At the intermediate meso-level, one can think of introducing more ICT at schools (De Witte, Haerlemans and Rogge, 2015), reducing class sizes or smoothing the transition between schools (De Witte and Cabus, 2013). At the system level, innovations include restructuring school systems, administrative mergers of schools, or closing poor performing schools (De Witte and Van Klaveren, 2014). While most of these innovations are costly, advocates point to their effect on efficiency and effectiveness. The former denotes that outcomes, e.g. educational attainments, can be reached with fewer resources, or that with the same resources higher outcomes can be obtained. The latter -effectiveness- indicates that the expected outcomes are reached.

Regarding the effectiveness, the Dutch Education Council pointed in their influential 2006 report already to the lack of evidence based approaches in Dutch primary and secondary education (Onderwijsraad, 2006). Innovations are too often introduced at a large scale without examining their effect. This has also been acknowledged by the 'Commission Dijsselbloem' in 2008. This commission, installed by the Parliament, suggested, among others, that the government should stimulate 'evidence based' or 'practice based' innovations in schools. It also suggested to choose for small-scale research, which focusses on specific aspects; to account for the significant diversity in schools and in pupils; and to invest more resources and time in the dissemination of existing knowledge. In 2014, the Dutch Education Council examined whether the report by the 'Commission Dijsselbloem' has led to a structural change in education

policy. The Education Council (2014) observed that evidence still plays a minor role in educational innovations.

Regarding the efficiency of educational innovations, education accounts for 11.5% of total public spending in the average OECD country and about 4.6% of GDP in 2012 (OECD, 2015). Given the trend to austerity, given that education inflation is way above the consumer price index, the shortages in the teacher labour market (Ministry of Education, 2011), and the capacity constraints of many schools the call for efficiency in education is increasing.

This chair continues to this stream of ideas by looking at the efficiency and effectiveness of educational innovations in education. During the inaugural lecture of the chair, I guide you through my earlier work. In particular, using insights from earlier papers with my various co-authors, I will answer the question how efficiency and effectiveness are intertwined in educational innovations.

1. Resource-driven policies

Education matters. The quality of education, as measured by learning outcomes of students in cognitive tests is strongly and positively correlated to the economic growth of countries (Barro, 2001; Hanushek and Woessman, 2008). The question is how can we improve the quality of education? A first idea to improve the quality is by increasing the resources for education. In a paper with Benny Geys and Catherina Solondz we explore whether resource expansion can improve actual student performance (De Witte, Geys and Solondz, 2014).

The question whether or not resource-driven policies are effective in increasing educational quality and student performance remains vigorously debated (for reviews, see Hanushek, 2003; Hægeland et al., 2012). Hoxby (2000) argues that this ambiguity may well derive from differing objective functions of teachers, schools or public authorities. Another reason, however, may be that exam systems differ widely. In some systems, exams and grading standards are set by schools, while in others central standards or exams are

set. The latter clearly limits the opportunity for teachers and/or schools to affect the grading structure when resources are increased and policy-makers expect students' achievements to improve accordingly. A change in education spending may therefore have a different observed impact (in terms of exam results) depending on the exam system at hand.

This naturally raises the question what determines grading practices (and whether resource shifts are one of these determinants). In this respect, two key factors are discussed in the existing literature. The first are accountability systems that evaluate teachers' and schools' performance via students' test scores. Apart from exerting a positive effect on students' achievements (Carnoy and Loeb 2002; Koning and van der Wiel 2012), undesired side-effects of such systems range from focusing teaching effort on pupils with achievements close to tests' thresholds (Reback, 2008; Neal and Whitmore Schanzenbach, 2010; Rockoff and Turner, 2010) to the distortion of results and cheating by teachers (i.e. grade inflation; Jacob and Levitt, 2003). The second determinant of grading practices is student 'demand'. Bonesrønning (1999), for instance, argues that rent-seeking students may press for easy grading while DePaola and Scoppa (2010) highlight diverging preferences of high- and low-ability students for precise versus noisy grading (see also Himmeler and Schwager, 2013). However, the only study explicitly linking school resources to grading practices is Backes-Gellner and Veen (2008). They argue that schools have incentives to lower their grading standard if their budget depends on the number of students. Although they do not provide a formal verification, their argument suggests that grading standards might depend on financial constraints. Yet, it does not necessarily imply that this likewise holds for public education expenditures.

Most studies in this vast literature analyze school-level exam results, rather than standardized central exit exams or SAT scores. In De Witte, Geys and Solondz (2014), we argue that this choice of evaluation standard is not innocuous. The reason is that schools can affect observed student performance through their choice of grading

standard, which not only translate students' performance into a given grade, but also affect learning effort. This suggests that resource-driven policies may have both a direct effect on student performance (extensively discussed in the foregoing literature), and an indirect one via schools' endogenous grading structure decisions (disregarded in earlier work). The pressure on schools receiving more resources to show improved outcomes might indeed induce them to 'game' the system and 'generate' better achievements by inflating their grades.

To examine the latter, we exploit a policy intervention in the Netherlands, which features two crucial characteristics. First, it created a quasi-experimental setting where 40 districts in 18 cities received substantial additional block grants from the Dutch central government totaling 250 million euro per year (while other, often quite similar, districts received no such funding). These additional funds were earmarked for investments in social policies such as education as of summer 2007, and the responsible minister explicitly made the improvement of educational outcomes one of the core aims of the program (Tweede Kamer, 2008-2009). Second, pupils' school-leaving test results in the Dutch education system are determined by both standardized national exit exams and school exams. Since schools only have discretion over the grading standard in the school exam, we can employ the results of the central exam as a benchmark (uniformly applied to all pupils in all schools) against which to set the school exam results.

This setting allows for a difference-in-differences (DiD) identification strategy whereby Dutch schools inside/outside the selected districts are compared over the 2004-2006 period before the intervention and the 2008-2009 period after the intervention (see also Gerritsen and Webbink, 2010; Wittebrood and Permentier, 2011). Our findings show that, on average, there is a decline in central exam results, but an (insignificant) relative improvement in school exams, in schools located in districts with additional funding. Accounting for the varying size of the investment program across districts, higher investment is found to significantly dampen the relative decline in central-level exam results,

while leaving school exams unaffected. Hence, increased resources seem to have positively affected central exam results when additional funds were sufficiently elevated, but induced grade inflation when funds were limited. These findings are robust to the level of analysis (i.e. schools or districts), different specifications of the control group and the implementation of a matching estimator exploiting the purposeful assignment to the treatment.

2. Performance versus Efficiency

Evidence Based Education

As De Witte, Geys and Solondz (2014) show that general resource-driven policies do not necessarily increase education quality, measured by student outcomes, question remains how education can be further improved.

The evidence-based education paradigm argues that we should examine the effectiveness of educational innovations. By examining the effectiveness of innovations at a small scale, we should only implement innovations at a larger scale when they are effective.

This is exactly what the Top Institute for Evidence Based Education Research (TIER) is doing. The researchers of this inter-disciplinary and inter-university research group at Maastricht University, University of Amsterdam and University of Groningen run experiments and quasi-experiments to explore the effectiveness of innovations to reduce early school leaving in secondary and higher education, innovations like 'Playing for Success', youth juvenile like HALT, ICT innovations, etc.

Similar innovations have always an effectiveness and efficiency component. Effectiveness refers to the capacity to do the 'right things'. For example, the innovations should reduce early school leaving or improve cognitive learning outcomes. Efficiency indicates that the innovations should be done in the 'right way', i.e., without spoiling resources.

Efficiency versus effectiveness

The concepts of efficiency and effectiveness are clearly interrelated. In a paper with Sergio Perelman and Laurens Cherchye (2015) we argue that the essential difference between these two approaches relates to whether or not resource constraints are taken into account. We may also argue that both approaches provide a partial analysis, which may thus give incomplete and potentially misleading information. First, effectiveness (or performance) scores tell us to which extent targets are fulfilled, but they do not inform us if the service provided is produced at full capacity. Second, productivity scores show whether or not the service is supplied efficiently (by exploiting the available production capacity in an optimal way), but they do not tell us to which extent the general outcome targets are achieved. The main reason behind this incompleteness is data availability. Very often data on outcomes and resources are not observed simultaneously, or they do not match perfectly for the production units under observation (Pestieau, 2009). Cherchye, De Witte and Perelman (2015) show that when data on outcomes and resources are available for a large sample of observations both performance and productivity scores are computable by using an appropriate frontier approach.

We are able to compute a performance/productivity ratio. If this ratio is below unity it indicates that better outcomes can be attained by weakening the resource constraints. In other words, scarcity of resources hampers output performance. On the contrary, a performance/productivity ratio above unity indicates that better outcomes can be achieved with the same resources. This mainly signals unexploited production capacity rather than scarcity of resources.

In an application to Dutch secondary schools, our results indicate that in most cases schools' performance improvement is a matter of unexploited production capacity, while scarcity of resources is a lesser issue. It provides an argument for educational policy makers in times of austerity. While there are schools with resource constraints, the

majority of the schools should first increase their productivity before requesting additional funding.

We have also investigated what are the characteristics of the schools with unexploited capacity versus those faced with resource constraints. Under-utilization of resources is positively related to the number of school managers. This finding is in line with the large increase in the number of middle managers in the Netherlands. Due to a consolidation of the number of school districts (i.e., more schools per governing body), we could observe an increase in the number of school managers. Our analysis suggests that many schools would benefit from a reduction of the number of managers as it correlates to the under-utilization of resources. Scarcity of resources bears a positive association with the number of students per teacher. This observation provides a hands-on tool for policy makers to analyse the over-utilization of resources. Combining the latter two observations, it can be argued that schools with a high number of students per teacher and a low number of school managers operate under serious resource constraints. Larger schools seem to have a scarcity of resources. Those schools are often located in urban areas, such that they face various challenges due to their unfavourable socioeconomic position. In addition, larger schools have a higher complexity such that educational policy makers should compensate these schools with additional resources. Finally, resource over-utilization correlates positively to the number of students who leave school without a higher secondary degree and who are not further enrolled in education or training. This suggests that the more early school leavers a school has, the less unexploited capacity there will be. Schools with stringent resource constraints seem to be unable to monitor and prevent early school leaving. Given the substantial societal costs of early school leaving, this suggests that governments should make sure that they provide a sufficient large resources to schools to prevent such action.

The utilization of resources is also explored in my work with Carla Haelermans and Jos Blank (2012). In this paper we analyse for Dutch schools whether alternative school

resources, such as supporting personnel, school management and materials, can be - to a limited extent - substituted for teachers. Tightening budget constraints and increasing demands force schools to spend resources in terms of employees, management and material in the most productive way. Much is at stake. The increasing requirements for schools (e.g., more pupil counselling, additional extra-curriculum activities, use of school buildings during weekends and summer holidays) induce pressure on the resources which are already in place. Productivity becomes an issue in schools (see Ball & Goldman, 1997; Ministry of Education, 2011, p. 98). With the insights at hand, resources can be reallocated among schools such that higher educational attainments can be obtained.

Haelermans, De Witte and Blank (2012) observe an overutilization of teachers by 0.65% (which is about 400 FTE (Ministry of Education, 2011)) and an underutilization of other personnel resources in 2007. Our findings imply that a different allocation of resources is not necessary for most schools, and that the national teacher shortage cannot be solved or reduced by the reallocation of resources. However, more than 25 per cent of the schools have (far) too many teachers, compared with the optimum and for these schools reallocation of their surplus teachers to the shortage schools could be a solution to the teacher shortage.

3. Issues with modelling efficiency of education

The two previous papers (Cherchye et al., 2015 and Haelermans et al., 2012) rely on efficiency models. Efficiency measurement is an interesting route to explore educational innovations. My earlier work in efficiency analysis focused on two important issues in the efficiency literature: the selection of variables, and the applied models.

First consider the selection of variables. De Witte and Laura Lopez-Torres (2016) review the literature on efficiency in education, and in particular the applied input, output and contextual variables. They show that the earlier applied inputs

and outputs are mainly data driven (i.e., people select those variables which they happen to observe in their data). Consequently, some variables are used as both input, output and contextual variables. In addition, their review shows that in the extensive empirical literature on productive efficiency, one common characteristic is the reliance on final outcomes - e.g., school results, as the main output measure to measure the efficiency of schools.

De Witte and Geys (2012) argue that this choice ignores a key characteristic of the production process for public services: namely, the central role of the underlying population (e.g., students, citizens) as coproducers of services (e.g., Whitaker, 1980; Parks et al., 1981). Such coproduction has received renewed academic and political interest in recent years (Meijer, 2011). Yet, while the academic literature on coproduction discusses both the determinants and consequences of such coproductive activities, it does not highlight its critical importance for the measurement of productive efficiency. In combining the literatures dealing with the measurement of efficiency and with coproduction in service provision, De Witte and Geys (2012) argue that coproduction in public services make that final outcomes are inappropriate output variables in efficiency studies. The reason is that final outputs - due to such coproduction - are at least partly beyond the control of the service provider. That is, schools can "supply little education without inputs from students", police forces cannot maintain community safety without citizens reporting crimes or testifying in court (Parks et al., 1981, 1003) and tax collection is eased with citizens "submitting tax returns" (Alford, 2002, 39). As a consequence, estimates of public service providers' productive efficiency using such output measures may become biased. Briefly stated, low levels of citizen activity (e.g., low study investment by pupils or few requests for library books) imply relatively low levels of final outputs, which leads a high-input service provider to be designated as inefficient. Yet, it might be that this provider is very effective at translating basic inputs into service potential. If so, the use of final outputs as performance measures unduly punishes this service provider for the restricted nature of citizens'

coproductive activity in this area. This is not an argument to support the location of high-cost service providers in areas with low citizen participation, which constitutes a clear waste of public resources (i.e., allocative inefficiency). Rather, the argument is that, from a purely productive efficiency perspective, this service provider should not be described as an underperformer for an element beyond its control (i.e., citizens' coproduction or consumption decisions).

Also with respect to the applied methodologies, there are serious issues. In particular, the efficiency literature pays currently a significant effort to optimize its own models. In doing so, it focusses on minor methodological details and neglects the current important issues in related fields. Therefore, De Witte and Lopez-Torres (2016) propose to focus more on the issue of causality, and on how endogeneity biases the efficiency outcomes. They provide multiple avenues to do so by pointing to some conceptual similarities between efficiency models and some standard econometric techniques. First, the popular matching approach resembles in various ways the conditional efficiency model (see De Witte and Kortelainen, 2013). Further extensions to the latter approach can be inspired by the matching techniques. Second, partial frontier techniques remind us of the quantile analyses. Third, metafrontier techniques could drastically increase their popularity and relevance if they would mimic better the standard Difference-in-Differences ideas. Finally, there are some clear similarities between the measurement of value added in the efficiency literature and the effectiveness literature. By paying attention to the differences and similarities, both literatures might benefit from each other.

4. Measuring educational performance in an accurate way

If additional resources do not help to improve educational outcomes, the question remains how we can improve the performance and productivity of schools. Principle-agent theories argue that incentives play an important role in this. Before incentives can play their role, we need to be able to

measure the performance of the agent in an accurate way. This is often problematic in an education context. I show two examples from my earlier work: the assessment of teaching and the measurement of research performance.

Together with Nicky Rogge, I explored the measurement of teaching performance (De Witte and Rogge, 2011). Students' evaluations of teaching (SETs hereafter) are increasingly used in (higher) education to evaluate teaching performance. Yet, for all their use, SETs continue to be a controversial topic with teachers, practitioners, and researchers sharing the concern that SET-scores tend to be 'unfair' as they fail to properly account for the impact of factors outside the teacher's control. The reason for this concern is twofold. On the one hand, there are the numerous findings in the academic literature which suggest that one or more background conditions (e.g., class size, teacher gender, teacher experience, course grades, timing of the course) may have a significant influence on SET-scores (see, for instance, Centra & Gaubatz, 2000; Marsh & Roche, 2000). On the other hand, there is the practical experience from teachers themselves which indicates that some teaching environments are more constructive to high-quality teaching (and, hence, high SET-scores) while other environments make such a level of teaching less evident. This potential 'unfairness' in mind, several researchers have argued for a cautious interpretation of SET-scores.

It is common practice to calculate scores as an arithmetic mean or as a sum of the ratings on questionnaire items (eventually expressed as a percentage to the maximal attainable overall rating). Essentially, this implies that all teaching aspects are assumed to be of equal importance. Whether such equal weights (and, in general, any set of fixed weights) are appropriate is questionable. Indeed, there are some indications suggesting that equality of weights across teaching aspects and/or over teachers is undesirably restrictive. As an illustration of the latter, teachers value teaching aspects differently in the definition (and, thus, the evaluation) of excellent teaching. These differences could be expected given the different personalities and abilities of teachers. Hence, using fixed weights in the build-up of SET-

scores may be somewhat counterintuitive. Moreover, in the absence of a consensus on how teaching aspects exactly interrelate, any choice of fixed weights will be subjective to some extent. The use of fixed weights can also introduce unfairness in teacher evaluations.

De Witte and Rogge (2011) suggest the use of a conditional Benefit-of-the-Doubt (BoD) model. Thanks to its linear programming formulation, the BoD model allows for an endogenous weighting of multiple outputs/achievements (Melyn & Moesen, 1991). We design the BoD model such that it allows for measurement errors which arrive from the survey data. In particular, we apply insights from the robust order- m efficiency scores of Cazals, Florens & Simar (2002) to our specific BoD setting. As such, the BoD model has three major advantages. First, for each teacher performance under evaluation, the weights on the questionnaire items are chosen in a relative perspective such that the highest possible SET-score is realized. Therefore, teachers with one or more low SET-scores can no longer blame these poor evaluations to unfair weights. Second, the BoD model is flexible enough to incorporate stakeholder opinion (e.g., teachers, students, experts) in the construction of the SET-scores. Among others, Pritchard et al. (1998) strongly argued in favour of developing an evaluation system with such significant and meaningful stakeholder (particularly the teachers) participation. In their opinion, such involvement is a necessary condition for the credibility and acceptance of the evaluation results. Third, the robust specification of the BoD model allows us to account for outlying and wrongly measured questionnaire values. The 'conditional' BoD model (adapted from De Witte and Kortelainen, 2013) allows for environment adjusted SET-scores without assuming a separability between the teacher's performance and the exogenous influences. The conditional model is a non-parametric technique that allows us to include teacher, student and course related influences immediately in the efficiency scores.

As a second example, it is difficult to accurately measure research performance. Current literature on research evaluation mainly employs single-criterion measures, such as

reputational ratings gathered by polls or peer reviews, number of publications (eventually in conjunction with a journal quality index or a citation index) in a predefined set of refereed journals, or citation counts. Recently, several opponents have criticized such simplistic measures doubting whether they are able to accurately convey research performance. In their opinion, the nature of research is by far too complex to be grasped by one single output criterion.

However, the construction of a multi-criteria Research Evaluation Score (RES-score) is an intricate matter with, amongst others, two important conceptual and methodological difficulties to overcome: (1) How should one weight and aggregate the different output criteria? Or, stated differently, how important are the several research outputs in the overall performance evaluation? Is it legitimate to assign a uniform set of weights over the several output criteria (i.e., equal/fixed weights)? Also, is it legitimate to apply a uniform set of weights to all evaluated researchers? Some researchers are clearly specializing in writing international books, while other are specializing in attracting research funding. Using the same weights for all researchers, would be considered as unfair within a research unit. (2) How should the RES-scores be adjusted for the impact of exogenous characteristics which are (often) beyond the control of the researcher? There are numerous findings in the academic literature which suggest that some background characteristics (e.g., age, gender, rank/tenure, time spent on teaching, department policy, etc.) may have a significant impact on the research performance of academic staff. Intuitively, some conditions are more beneficial to productive research while other conditions are more detrimental. Yet, traditional RES-scores do not account for differences in these uncontrollable conditions. Consequently, these scores are inherently biased towards researchers working under more favorable conditions. With this 'bias' in mind, several practitioners and researchers have claimed that uncorrected scores tend to be unfair as they give an advantage to those who work in more constructive conditions. The opposite reasoning holds true for academics who work under less favorable conditions. In their case, it is more difficult to obtain a good performance level (and, hence,

a good RES-score). Stated differently, unadjusted RES-scores are potentially flawed and, therefore, unreliable as a measure of researcher performance. However, we are unaware of any study which corrects RES-scores for heterogeneity in (potentially) influential characteristics and conditions not under the control of the evaluated researchers.

Using a specially tailored version of the Benefit-of-the-Doubt model, De Witte and Rogge (2010) suggest an endogenous weighting of multiple research dimensions. Their proposal allows us to control for the operational environment and flexible weights depending on the strengths of the researchers.

5. Scope economies

Using the more accurate way to compute student evaluation of teaching (SET) and research evaluation scores (RES), we can combine the efficiency and effectiveness literature in examining the scope economies between teaching and research (in De Witte, Rogge, Cherchye and Van Puyenbroeck, 2013). The debate on the precise relationship between teaching and research - in particular whether teaching tasks hinder good research or whether research benefits the teaching effectiveness - is a long-standing and highly controversial one. A key reason for this controversy seems to be the involvement of multiple university stakeholders (such as students, professors themselves, university management, and policy makers) with sometimes diverging interests. Broadly speaking, there are two main streams in the debate: the believers in a positive teaching-research nexus and the sceptics. The believers argue that teaching and research go together in an essential and symbiotic way. Therefore, they believe that it is unwise to separate teaching and research (i.e., assigning activities among professors and/or institutions such that there are only full-time teachers/teaching institutions and full-time researchers/research institutions). The sceptics strongly contest this alleged positive teaching-research nexus. In their opinion, this positive nexus is just a notion invoked by certain stakeholders who benefit from the

presence of this perceived link. Most of the skeptics even question whether there exists any relationship at all between teaching and research. Hattie and Marsh, for example, posited that "the widely held belief that teaching and research are inextricably entwined is an enduring myth. At best, research and teaching are very loosely coupled" (Hattie and Marsh, 1996 p. 529).

We compute the SET and RES-scores in unique data of a university college that can be compared to the 'new' polytechnic universities in the UK and the colleges in the US where a teaching university is only recently research oriented. In this transformation, the research and teaching time has been allocated exogenously and did not depend on the professor's actual and past performance, but rather on the group he/she belongs to. This makes an attractive setting for studying a more causal relationship between teaching and research.

Our results indicate that professors who are full-time researchers (and, hence, currently have no teaching duties) could improve their overall academic performance by investing a small proportion of their time in teaching. However, this favorable effect of increasing teaching time is fairly modest (and only holds if teaching time is less than 20% of total time). In fact, once teaching time exceeds 20%, further increases in teaching duties seem to harm the overall academic performance. Besides for the extremes of full specialization, we observe that specialization in teaching and research correlates with better academic performance.

However, we believe that this finding should be interpreted with caution. A detailed study of our data shows mixed outcomes, with some professors succeeding in balancing a considerable research task with moderate teaching loads (realizing good SET-scores and a high research output) and others being less successful in combining both activities (realizing only good SET-scores, or a high research output, or neither). Next, our results clearly suggest that granting only marginal research time to professors is inefficient in terms of improving overall academic performance. In fact, a large majority of the professors with low research percentages did not succeed in realizing any

research output. Apparently, a minimal amount of research time is needed for a professor to be able to produce research outputs. This finding seems intuitively plausible.

6. Providing incentives

In the measurement of teaching and research performance, it was necessary to account for the operating environment where teachers and researchers are operating at. While often neglected, it is as important to control for heterogeneity when educational outcomes can be easily and accurately measured. Early school leaving (or school dropout), measured in administrative data as students who leave education without a higher secondary degree, provide an example for this. One cannot consider straightaway differences in outcomes as a performance measure, even for a unidimensional, simple and uniform indicator of performance. Although this may seem common knowledge, in practice, many policy incentives are based on raw comparisons of outcomes. The effectiveness of regional policy does not only depend on the regional policy makers themselves, but also depends on the underlying population. If cities or regions are held accountable for their performance on particular dimensions, central government should account for regional heterogeneity in an appropriate manner.

The literature on regulation (e.g., Laffont and Tirole, 1993) shows that uniform incentives can be demotivating. The principal (i.e., the central government) sets out an incentive for the agent (i.e., the regional government). To keep the incentive as transparent as possible, the principal determines a uniform incentive for all agents. If an a priori specified target is obtained, the agent receives the incentive. This uniform incentive is common practice in the health sector (e.g., for quality targets), transport (e.g., keeping time tables), service centers (e.g., time to take a call), etc. From the examples, it is clear that the uniform incentive is wide spread in policy. Nevertheless, for agents (e.g., hospitals, transport companies, etc.) operating in different (geographical) regions, obtaining the same target may be

more requiring in a region with less advantageous characteristics than in a region with more advantageous characteristics. Both 'naming and shaming' and monetary rewards can therefore provide a negative motivation for the region with a disadvantageous population.

I provide two examples on how to correct for similar differences in the comparison of school dropout rates. Together with Sofie Cabus, I explored how the European naming-and-shaming strategy in early school leaving is influenced by economic cycles. As supra-national organizations often have few tools to urge member states to undertake policy reforms, they rely on peer pressure arising from cross-country comparisons. The peer pressure is particularly aimed for in the 'naming and shaming' framework, in which the best performing countries/regions are named as benchmarks and the least performing countries are shamed as lagging behind. Previous literature indicated that naming and shaming is an effective incentive. Particularly the United Nations and the European Commission use this 'sunshine regulatory model' extensively in their evaluation of the policy outcomes of member states (e.g., poverty reduction, unemployment rates, educational attainments, or production volume).

Despite its wide spread use, a mere comparison of outcome variables is nearly meaningless. Various other factors may have influenced the outcome variable of interest. If not properly accounted for, the 'naming and shaming' will be inaccurate and favors countries which are influenced by positive exogenous influences (e.g., economic cycle).

In Cabus and De Witte (2012) we suggest a benchmarking model in which outcomes are corrected for time and country invariant effects as well as economic conditions. This contributes particularly to the productivity literature where the method can be best compared to index numbers such as Laspyres index or Tornquiz index. Index numbers are convenient and heavily applied as they are non-parametric (i.e., they do not assume a priori specifications on the production frontier) and relatively easy to interpret and estimate (Fried et al., 2008). They suffer, however, from difficulties to account for heterogeneity. The suggested

parametric benchmark model proceeds by applying micro-econometric panel data techniques to a macro-economic setting. It is estimated by system ordinary least squares.

Using the example of early school leaving, Cabus and De Witte (2012) provide evidence that relying on output indicators might name and shame the wrong units. Various observed and unobserved variables influence the output indicator. If the general economic climate heavily influences the dropout rate, and if the economic climate is hard to influence by educational policy makers, than one may reasonably argue that this should be accounted for in a 'fair' benchmark regulatory framework. National and supra-national organizations which rely on indicators should thus at least correct for the observed influences (and preferably also for the unobserved time invariant influences).

As a second example, similar observations on the inappropriate comparison among students was observed in work with Chris Van Klaveren. In De Witte and Van Klaveren (2012) we focus on school dropout in some large Dutch cities. One of the policy measures to reduce dropout consists of a uniform monetary incentive of 2500 euro per early school leaver less in comparison to the base year 2005-2006. The Ministry of Education (OCW) allocated 5.4 % of this budget (i.e., 17.04 million euro) to this uniform incentive in 2008. This allocation increases to 11.4 % (i.e., 45.44 million euro) in 2011. It can be expected that regions with more disadvantageous population (e.g., more difficult to reach) receive less subsidy (i.e., the 2500 euro per student less) than regions with a more advantageous population. If this is true, a uniform incentive is not the most appropriate way to stimulate policy makers as a uniform incentive provides a negative motivation in this case. We showed that this was the case.

7. Effectiveness

Besides contributing to the (measurement of) performance of the education system, this chair aims to examine the effectiveness of educational innovations. In line with the

principle of evidence-based education, we should rigorously explore whether educational innovations work, and if so, compute their cost-effectiveness. The example of early school leaving is attractive to discuss some examples of effectiveness analysis.

Early school leaving makes an attractive example as there is much at stake. In a world characterized by innovation and persistent globalization, transversal skills such as critical thinking, initiative taking and problem solving are becoming ever more important. According to CEDEFOP (2012), the proportion of jobs requiring tertiary level qualifications will increase from 29% in 2010 to 34% in 2020. On the contrary, predictions over the same period indicate that the proportion of lower-skilled jobs is to witness a significant plunge, namely from 23% to 18%. Yet, in 2014, 5.6 million EU citizens aged 18 to 24 have not completed upper secondary education and were no longer in education and training. This represents an average EU rate of early school leaving (ESL) or school dropout (we will use the terms interchangeably throughout the paper) amounting to 11.1% (Eurostat, 2015). Likewise, only 79% of the US pupils aged 25 years or below completed upper secondary education; just one percentage point above the OECD average in 2012 (Organization for Economic Cooperation and Development (OECD), 2014). According to the estimates of the European Commission (2011), merely a single percentage point in ESL reduction would lead to nearly half a million additional qualified employees for the EU economy (see also Mazrekaj and De Witte, 2016).

We focus on the Netherlands, which is an interesting case study. Before the academic year 2002-2003, a coherent policy towards early school-leaving was lacking (see De Witte and Cabus, 2013). Because of large differences in the underlying population, different regions and cities require different policy measures. Therefore, a decentralized policy has been developed. A task force on early school leaving within the Ministry of Education (the so-called 'Projectdirectie Voortijdig Schoolverlaten') created 39 regional dropout authorities (RMC) in 2002. At that time, each of the RMC regions could take (or could not take) different actions towards policy goal settings. This can be exploited in research.

Compulsory education age

One straightforward way to prevent students from school dropout is increasing the compulsory education age. By staying longer in school, the idea is that more students eventually obtain a higher secondary diploma. Early school-leaving would be reduced. In Cabus and De Witte (2011) we examine whether an increase of compulsory education age effectively reduces school dropout.

The causal impact of one additional year of compulsory education on dropout is estimated on administrative data of all students going to school in Amsterdam. To estimate the causal impact of compulsory education on early school-leaving, we exploit the implementation of the policy reform in the Netherlands. On the one hand, students who were born before August 1st 1990 and, hence, are 17 year olds on the date of the policy reform are exempted from the policy reform. They have to attend school until their 17th birthday has been reached. On the other hand, students born after August 1st 1990 are liable to the new Qualification Law. The Qualification Law implies mandatory school attendance until the age of 18 has been reached or until a qualification has been obtained. To enforce the implementation of the law, improved registration has been developed. Moreover, truancy officers can fine the parents of the truant (see De Witte & Csillag, 2014 for an evaluation). After the age of 18, and thus after adulthood has been reached, parents are no longer responsible for the school attendance of their child. An adolescent without qualification has the responsibility to obtain a higher secondary diploma, but is no longer obliged to attend school. Consequently, the 2007 Qualification Law implies a one-year increase in compulsory education age: from 17 to 18. It is clear that August 1st serves as a discontinuous date. To find out the effect of a one-year increase in compulsory education age, one can easily compare the average dropout rates of the student group born before August 1st 1990 with the average dropout rates of the student group born after August 1st 1990.

Cabus and De Witte (2011) estimate a structural equation for dropout using difference-in-differences methodology. The estimate suggests a decrease in early school leaving of -2.52 percentage points thanks to the increase in compulsory education. Although the results reveal a significant decrease in dropout thanks to the compulsory education law, some caution is in place. The observed effect arises almost completely from the fact that control group students left school at the legal school-leaving age after the policy reform. Given the economic revival at the time of the policy reform, this may suggest an anticipation of job market opportunities. On the contrary, students liable to the policy reform did not significantly alter their dropout decision. As a major issue for discussion, the results of this paper point to the unintended consequences of straightforward policy making. Even though the new Qualification Law did not intentionally influence control group students, the results suggest that the policy reform caused a labor demand shock for semi- and unskilled students. The effect is local in nature as it only affected a small group of students.

Active School Attendance Intervention

As a second example, Cabus and De Witte (2015) evaluate the impact of an Active School Attendance Intervention (further abbreviated by ASAI), which aims to reduce school dropout by actively focusing on unauthorized truancy. The intervention is a pilot project called "Bewust Aanwezig op School" (i.e., "increased awareness for school attendance"). The intervention aims to reduce school dropout by focusing on school absenteeism. Professional mentors, teachers, and compulsory education age officers are involved in the project. The intervention runs as follows. A mentor and/or social worker has a personal talk with every truant. If necessary, this is followed by a home visit by a social worker. The home visit takes about 3 hours in which both the student and his/her parents are pointed to the importance of schooling. As part of the prevention program, the ASAI also increases the awareness of teachers to the issue of dropout prevention. For instance, specially tailored information sessions at the school are organized by compulsory education age officers. The

sessions contain information on awareness for school attendance, better reporting, uniform registration, and care for at-risk students. As a final part of the intervention, school management, teachers, mentors, social workers, and compulsory education age officers gather on a regular basis at the school. This should stimulate the development of professionalism at the school and the exchange of information and best practices.

The quasi-experimental design in Cabus and De Witte (2015) exploits the way the pilot is set up. We can make a clear distinction between control schools and intervention schools, which, from a policy perspective, only differ in the intervention. Indeed, both control schools and intervention schools belong to the same administrative dropout authority, i.e., a policy level defined by the Ministry of Education to combat school dropout. Therefore, all schools in the sample apply the same set of interventions. While there is no observed selection into the control schools or intervention schools (e.g., due to reputation effects), student characteristics might still differ across the schools. To mitigate the impact of similar differences, we construct a matched sample of students with similar (observed) characteristics as gender, ethnicity, place of residence, and study track. The matching estimation technique allows us to match lookalike untreated students to treated students. Using the matched sample, we apply a difference-in-differences design to estimate the causal effect of the ASAI intervention on school dropout.

The results suggest that the intervention schools have significantly reduced school dropout with -0.6% points in the year of implementation (i.e., the school year 2009–2010). This is a significant reduction compared to the average dropout rate of about 2% among 12- to 16-year-old students.

8. Conclusion and ways forward

In this inaugural lecture, I provided a brief overview of my earlier work to show how efficiency and effectiveness are intertwined. This learned us the following points:

- I argued that general resource-driven policies do not necessarily increase education quality, measured by student outcomes. A way to improve education outcomes is by carefully examining the effectiveness of educational innovations.
- I argued that the concepts of efficiency and effectiveness are interrelated. I discussed that both approaches provide a partial analysis, which may thus give incomplete and potentially misleading information. First, effectiveness analysis tells us to which extent outcomes are reached, but they do not inform us if the service provided is produced at full capacity. Second, productivity scores show whether the available production capacity is used in an optimal way, but they do not tell us to which extent the general outcome targets are achieved.
- Current literature examining educational efficiency mainly focused on student test scores to assess the efficiency of schools. This might be a wrong output variable as it ignores that education services is coproduced by the students, i.e. schools can "supply little education without inputs from students".
- Insights from efficiency analysis allow us to measuring performance in an accurate and more benevolent way. I showed that the assessment of teaching and research performance might be considered as more fair by using the endogenous weighting system from the efficiency literature.
- A popular tool for national and supra-national governments to analyze the effectiveness of interventions is by considering differences in outcomes as a performance measure. By using the example of early school leaving, I argue that a more fair benchmarking model should capture the heterogeneity in the underlying population.
- I showed by two examples that effectiveness analysis can result in surprising outcomes: expensive interventions can be cost-effective and well-intended policy measures can have undesired outcomes.

This chair is unique in combining the efficiency and effectiveness literature. It explicitly combines insights from efficiency and effectiveness analysis, and applies them to education settings. My work in this chair will rigorously show how both streams of research could mutually benefit from each other. To set the scene, I discuss two examples.

First, I take the example of matching analysis and conditional efficiency analysis (De Witte and Lopez-Torres, 2016). While matching analysis focusses on the effect of a treatment and the conditional efficiency analysis focusses on the relative efficiency of observations, the techniques share similar ideas and aim to interpret the influence of confounding variables in a similar way. The conditional efficiency framework uses non-parametric kernel estimations to attach weights to observations with similar observed characteristics. Observations which are similar to the evaluated observation are in the resampling (cf. order- m) more frequently drawn than observations which are dissimilar. To obtain statistical inference, the conditional efficiency scores are compared to the unconditional efficiency scores. As discussed in Van Klaveren and De Witte (2014), in many ways, this is closely similar to a '1:1 matching' strategy. This statistical technique searches for each treated (cf. evaluated) observation a non-treated observation with similar observed characteristics. As the observed characteristics are the same for both groups, matching allows a researcher to assess the influence of a treatment without reduced bias from confounding variables. The assumption is that by selecting on the observed characteristics, also the unobserved characteristics will be similar. Also the 'kernel matching' is comparable to the weights that are assigned in conditional efficiency analysis. Given that the matching methodology is generally more accepted in the economics of education literature, an interesting way forward is to combine the conditional efficiency approach with insights from matching.

As a second example, I focus on endogeneity arising from measurement errors (see De Witte and Lopez-Torres, 2016). Particularly deterministic efficiency models are prone to measurement errors. Although measurement errors can shift the frontier, they are largely neglected. In the case of

measurement errors, the efficiency scores will be biased due to the increased variability. Some suggested methodologies as order-m (from Cazals et al., 2002) or order-alpha (from Daouia and Simar, 2007) mitigate the influence of measurement errors by resampling the original sample. In the absence of prior information on measurement errors, the latter techniques should be favored on standard deterministic techniques. In addition, the efficiency literature should better discuss the presence of measurement errors and its effect on the efficiency scores. For example, the presence of measurement errors might result in an upper bound estimation of the efficiency scores. By discussing this, the policy relevance of the efficiency study increases.

The latter two examples set the scene on how this chair will actively contribute to bridge the gap between the methodologies applied for efficiency and effectiveness analysis.

Besides methodological advances, in this chair I aim to contribute to some policy relevant empirical applications. I provide two examples. A first empirical research line in this chair will deal with school management. I argued before that under-utilization of resources is positively related to the number of school managers. We also know from earlier research that school district organization matters for education outcomes of students. In this chair, I will further deepen this theme.

A second empirical research line deals with under-achievement in terms of learning outcomes and as an extreme form, early school leaving. I argued before that resource over-utilization correlates positively to the number of early school leavers. Schools with stringent resource constraints seem to be unable to monitor and prevent early school leaving. By examining the cost-effectiveness of interventions to reduce early school leaving, I would like to provide evidence-based policy recommendations on how to reduce early school leaving in an effective and efficient way.

During the earlier workshop on education economics, there were scholars from both fields: efficiency analysis and

effectiveness analysis. I hope that the workshop, and this inaugural lecture, can provide further input for our research community. In particular, in this chair, I would like to stimulate the research community by projects that aim for networking activities (e.g., the new 'Economics of Education Network' project), by introducing PhD-students in the field of education economics (e.g., the new graduate program by TIER), by receiving visiting scholars and introducing them to the community, and, finally, by bringing people physically together by organizing workshops and conferences. I truly believe that, together and as a research community, we can improve the educational system in an effective and efficient way.

Ik heb gezegd

Acknowledgements

About 10 years ago, when I was starting my PhD there was no suspicion at all that I would work on education economics. I would work on water utilities and local municipal efficiency. 'Education Economics' is a field that came rather by accident to me.

Although, by accident seems to be an overstatement. Looking at my family tree, about all family members work in education. My grand-father was a school principal, my father was a school principal and currently member of education inspectorate, two of my uncles are school principals and one aunt is a school teacher. So, Lukas and Jolien are very likely to work in education as well.

I would like to thank the Rector Magnificus, the Executive Board of Maastricht University, the dean of Faculty of Humanities and Sciences, Franz Palm and his predecessor Harm Hospers, and the Foundation 'Stichting Wetenschapsbeoefening UM' for their trust in granting me this position. I would like to express my gratitude to Henriëtte Maassen van den Brink and Wim Groot. I am grateful for their patience with my research enthusiasm, the multiple opportunities they provided me and their encouraging words along the journey. Evidence shows that this mentoring and coaching is of vital importance for any (educational) career. I am grateful to my colleagues at Maastricht University and at KU Leuven. In particular to my earlier mentors Wim Moesen and Laurens Cherchye, and to Luc Sels. I would like to thank the participants of this inaugural lecture and the earlier workshop. My work massively benefitted from colleagues like you. All of you stimulated me tremendously. I enjoyed collaborating with my more than 40 different co-authors, from which many became true friends.

I am exceptionally grateful to my parents. They provided me with plenty of opportunities, a carefree youth and sufficient self-confidence to survive in an ever critical academic environment. Evidence shows that these are critical success factors.

Finally, I would like to thank my wife Karen. She is not only the most perfect mother of my children, but also the strong woman that can handle the duties that come with a position in academia. She never says 'no' when I have to travel to conferences, academic visits or seminars. She is always supportive when I have to work at night. She never complains when I am leaving early in the morning or arrive late at night after teaching in Leuven, Amsterdam or Maastricht. In many ways, she contributes to this chair.

References

- Backes-Gellner, U. & S. Veen (2008). The Consequences of Central Examinations on Educational Quality Standards and Labour Market Outcomes. *Oxford Review of Education*, 34(5), 569-588.
- Barro, R.J. (2001). Human capital and growth. *American Economic Review*, 91(2), 12-17.
- Bishop, J.H. & L. Wößmann (2004). Institutional Effects in a Simple Model of Educational Production. *Education Economics*, 12(1), 17-38.
- Bonesrønning, H. (1999). The Variation in Teachers' Grading Practices: Causes and Consequences. *Economics of Education Review*, 18, 89-105.
- Cabus, S. & De Witte, K. (2011). Does School Time Matter? On the impact of compulsory education age on school dropout. *Economics of Education Review*, 30, 1384-1398.
- Cabus, S. & De Witte, K. (2015). The Effectiveness of Active School Attendance Interventions to Tackle Dropout in Secondary Schools -- A Dutch Pilot Case. *Empirical Economics*, 49 (1), 65-80.
- Carnoy, M. & S. Loeb (2002). Does External Accountability Affect Student Outcomes? A Cross-State Analysis. *Educational Evaluation and Policy Analysis*, 24(4), 305-331.
- Castelein, E., Coens, J., De Witte, K., Houben, A., Lauwers, W., Segers, J. & Van den Branden, K. (2016). *Binnenklasdifferentiatie, een beroepshouding, geen recept*. Acco Leuven – Den Haag. Pp. 152. ISBN 978-94-6292-560-1.
- Cazals, C., Florens, J.P., & L. Simar (2002). Nonparametric Frontier Estimation: A Robust Approach. *Journal of Econometrics*, 106(1), 1-25.
- Centra, J.A., & Gaubatz, N.B. (2000). Is there gender bias in student evaluations of teaching. *Journal of Higher Education*, 71(1), 17-33.
- Cherchye, L., De Witte, K. & Perelman, S. (2015). A unified productivity-performance approach, with an application to secondary schools in the Netherlands. *CES Working Paper Series 2015.19*.

- De Witte, K. & C. Van Klaveren (2014). The influence of closing poor performing primary schools on the educational attainment of students. *Educational Research and Evaluation* 20(4), 290-307.
- De Witte, K. & Cabus, S. (2013). Dropout prevention measures in the Netherlands, an evaluation. *Educational Review*, 65(2), 155-176.
- De Witte, K. & Cabus, S. (2013). Dropout prevention measures in the Netherlands, an evaluation. *Educational Review*, 65(2), 155-176.
- De Witte, K. & Geys, B. (2013). Citizen Coproduction and Efficient Public Good Provision: Theory and Evidence from Local Public Libraries. *European Journal of Operational Research*, 224, 592-602.
- De Witte, K. & Lopez-Torres, L. (2016). Efficiency in Education. A review of literature and a way forward. *Journal of Operational Research Society*. In press.
- De Witte, K. & M. Csillag (2014). Does anybody notice? On the impact of improved truancy reporting on school dropout. *Education Economics*, 22 (6), 549-569.
- De Witte, K. & Rogge, N. (2011). Accounting for exogenous influences in performance evaluations of teachers. *Economics of Education Review*, 30(4), 641-653.
- De Witte, K. & Schelfhout, W. (2015). *Vakdidactiek Economie. Lannoo Campus*. ISBN 978-94-014-2806-4, Pp. 338.
- De Witte, K., Geys, B., & Solondz, C. (2014). Public Expenditures, Educational Outcomes and Grade Inflation: Theory and Evidence from a Policy Intervention in the Netherlands. *Economics of Education Review*, 40, 152-166.
- De Witte, K., Haelermans, C. & Rogge, N. (2015). The effectiveness of a computer-assisted math learning program. *Journal of Computer Assisted Learning*, 31(4), 314-329.
- De Witte, K. and Rogge, N. (2010). To publish or not to publish? On the aggregation and drivers of research performance. *Scientometrics*, 85 (3), 657-680.
- De Witte, K., Rogge, N., Cherchye, L. & Van Puyenbroeck, T. (2013). Economies of scope in research and teaching: a

- non-parametric investigation. *Omega – International Journal of Operational Research*, 41, 305-314.
- DePaola, M. & V. Scoppa (2010). A Signaling Model of School Grades under Different Evaluation Systems. *Journal of Economics*, 101, 199-212.
- European Centre for the Development of Vocational Training (2012). *Future skills supply and demand in Europe*. Brussels: CEDEFOP.
- Eurostat. (2015). *Education and Training – Database*. Retrieved from <http://ec.europa.eu/eurostat/web/education-and-training/data/database>
- Fried, H., Lovell, C. A. K., & Schmidt, S. (2008). *The measurement of productive efficiency and productivity growth*. Oxford University Press., pp. 638.
- Gerritsen, S. & D. Webbink (2010). The Effects of Extra Funds for 40 Disadvantaged Neighborhoods. Paper presented at a Centraal Planbureau Conference at Rijksuniversiteit Groningen, March 3, 2010.
- Haelermans, C., De Witte, K. & Blank, J. (2012). On the allocation of resources for secondary education schools. *Economics of Education Review*, 1(5), 575-586.
- Hanushek, E.A. & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of economic literature*, 46(3), 607-668.
- Hattie, J. & Marsh, H.W. (1996). The relationship between research and teaching: A meta-analysis. *Review of Educational Research*, 66(4), 507-542.
- Himmler, O. & R. Schwager (2013). Double Standards in Educational Standards - Are Disadvantaged Students Being Graded More Leniently? *German Economic Review*, 14(2), 166-189.
- Hoxby, C.M. (2000). The Effects of Class Size on Student Achievement: New Evidence from Population Variation. *Quarterly Journal of Economics*, 115(4), 1239-1285.
- Jacob, B.A. (2005). Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools. *Journal of Public Economics*, 89, 761-796.
- Koning, P. & K. van der Wiel (2013). Ranking the Schools: How School-Quality Information Affects School Choice in

- the Netherlands. *Journal of the European Economic Association*, 11(2), 466-493.
- Laffont, J.J. & Tirole, J. (1993). *A theory of incentives in procurement and regulation*. MIT press.
- Marsh, H.W. & Roche, L. (2000). Effects of Grading Leniency and Low Workload on Students' Evaluations of Teaching, Popular Myth, bias, validity, or innocent bystanders? *Journal of Educational Psychology*, 92(1), 202-228.
- Mazrekaj, D. & De Witte, K. (2016). The effect of modularization on early school leaving in secondary education. Evidence from a quasi-experimental design in Flanders. *Mimeo*.
- Meijer, A.J. (2011). Networked Coproduction of Public Services in Virtual Communities: From a Government-Centric to a Community Approach to Public Service Support. *Public Administration Review*, 71(4), 598-607.
- Melyn, W. & Moesen, W. (1991). Towards a Synthetic Indicator of Macroeconomic Performance: Unequal Weighting when Limited Information is Available. *Public Economics Research Paper*, 17, CES, KULeuven.
- Ministry of Education (2011). *Nota werken in het onderwijs 2011*. Den Haag: Ministry of Education, Science and Culture.
- Ministry of Education. (2011). *Kerncijfers 2006-2010*. Den Haag: Ministry of Education, Culture and Science.
- Neal, D. & D. Whitmore Schanzenbach (2010). Left Behind by Design: Proficiency Counts and Test-Based Accountability. *Review of Economics and Statistics*, 92(2), 263-283.
- Organization for Economic Cooperation and Development. (2012). *Preparing teachers and developing school leaders for the 21st century - Lessons from around the world*. Paris: OECD.
- Organization for Economic Cooperation and Development. (2014). *Education at a Glance 2014 - OECD Indicators*. Paris: OECD.
- Parks, Roger B., Paula C. Baker, Larry L. Kiser, et al. (1981). Consumers and Coproducers of Public Services: Some Economic and Institutional Considerations. *Policy Studies Journal*, 9(7), 1001-1011.

- Pestieau, P. (2009). Assessing the performance of the public sector. *Annals of Public and Cooperative Economy*, 80, 133-161.
- Pritchard, R.D., Watson, M.D., Kelly, K., & Paquin, A.R. (1998). *Helping Teachers Teach Well: A New System for Measuring and Improving Teaching Effectiveness in Higher Education*. The New Lexington Press, San Francisco.
- Reback, R. (2008). Teaching to the Rating: School Accountability and the Distribution of Student Achievement. *Journal of Public Economics*, 92, 1394-1415.
- Rockoff, J.E. & L.J. Turner (2010). Short Run Impacts of Accountability on School Quality. *American Economic Journal: Economic Policy*, 2(4), 119-147.
- Tweede Kamer (2008-2009). *Krachtwijken: Monitoring en verantwoording van het beleid*, Tweede Kamer der Staten-Generaal, 31.723, nr. 1-2, 's Gravenhage.
- Van Klaveren, C. and De Witte, K. (2014). How are teachers teaching? A nonparametric approach. *Education Economics*, 22(1), 3-23.
- Whitaker, G.P. (1980). Co-Production: Citizen Participation in Service Delivery. *Public Administration Review*, 40, 240-246.
- Wikström, C. & M. Wikström (2005). Grade Inflation and School Competition: An Empirical Analysis based on the Swedish Upper Secondary Schools. *Economics of Education Review*, 24(3), 309-322.
- Wittebrood, K. & M. Permentier (2011). *Wonen, wijken en interventies: Krachtwijkenbeleid in perspectief*. Sociaal en Cultureel Planbureau – Den Haag.

