

When schools cheat on exams Additional funds for education make a difference – if they are substantial

Benny Geys

When you spend more money buying something, you generally expect to receive higher quality in return. This could mean better food in a restaurant, better service in a hotel or nicer fabric when it comes to clothes. As there is no reason to believe that politicians think any different, governments increasing the budget on, for instance, education might likewise expect class sizes to decrease, the teaching environment to improve or more (and better) teachers to be hired. Ultimately, the government would expect children to receive a better education – and we would most likely share that expectation since it is our taxes funding the increase in the education budget. This, however, leads to an obvious question: Are resource-driven public policies effective in improving educational outcomes?

This turns out to be a very difficult question, and well over five decades of academic research have only led to ambiguous results. Evidently, there may be many potential reasons why empirical studies addressing this question have produced widely varying results across time and space. One such reason is the fact that two things might happen when a school receives more government funding. On the one hand, the school could use these additional resources to expand and/or enhance the teaching environment with the ultimate aim of improving students' education. Even though the outcome in terms of improved student performance – most often measured via test scores – necessarily remains uncertain at the time the investments are made, this may be seen as the ideal scenario.

On the other hand, the pressure on schools receiving more resources to show value for money might induce them to game the system and generate better achievements by inflating their grades. That is, rather than invest in lower class sizes and/or better teachers, schools simply award their students higher grades for an unchanged performance level. In such a scenario, all uncertainty regarding improved student performance is eliminated through schools' – rather than students' – cheating on the exams.

A policy intervention in the Netherlands offers a unique opportunity to analyze the effects of increased spending. The central government designated 40 districts in 18 large and medium-sized Dutch cities as *power districts* (*krachtwijken* in Dutch) and received substantial additional block grants totaling 250 million euro per year (ranging from €1.2 million to €29.3 million across districts, or €333 to €3.995 per inhabitant in the districts). These additional funds were earmarked to finance public investments in social policies such as education as of the summer of 2007, and the responsible minister explicitly made the improvement of educational outcomes one of the core aims of the program. Interestingly, although the selection of long-listed districts was driven by a set of 18 indicators of socio-economic development, public disorder, and public safety within the area, the final decision to include or exclude districts was taken by the responsible minister. As a result, many districts in the Netherlands with often quite similar characteristics to those finally selected did not receive additional funding under the program.

The Netherlands is fairly unique in combining elements of a decentralized system, in which construction and evaluation of the exam is done by the school

Kurz gefasst: Was passiert, wenn der Staat in einzelne Brennpunktschulen unter der Bedingung investiert, dass die Noten dann besser werden? Mогeln Schulen, indem sie den Notenschnitt künstlich heben? Oder erhöhen sich die Leistungen wirklich? In den Niederlanden lässt sich das aufgrund der Kombination von Noten bei zentralen Prüfungen und dezentraler Notengebung durch die Lehrer erforschen. Das Ergebnis: Bei relativ geringen Investitionen steigt der Notenschnitt, ohne dass die Leistungen wirklich anziehen. Wenn allerdings viel Geld investiert wird, werden die Leistungen tatsächlich besser.

Summary: What happens when the government increases funding for individual schools in poor neighborhoods under the condition that testing results eventually improve? Do results really improve after a certain period, or do schools cheat by simply elevating their students' grades? The answer is offered by the unique situation in the Netherlands, where scores from standardized exams are combined with grades in class exams that depend on the individual teacher: When the amount of additional funding is modest, schools tend to cheat. When the amount of extra funding is substantial, scholastic achievement indeed improves over time.

teacher, and a centralized system, which clearly limits the opportunity for teachers and/or schools to affect the grading scheme and inflate grades when resources are increased and policy-makers expect students' achievements to improve accordingly. In primary and secondary education, pupils' school-leaving test results in the Dutch education system are determined by both standardized national exit exams and school exams.

Specifically, at the secondary level (the focus on the analysis below), all students have to take two exams for each course (with few exceptions). The first exam is a national assessment constructed by the Central Institute for Test Development (Cito). This exam's content is externally screened by professors and a prior test on a sample of students is taken to measure and monitor its difficulty, which is thereby guaranteed to remain at the same level over time. Correction of this central exam is based on a uniform correction model and there is a teacher from a different school acting as a second corrector. The second exam – the "school exam" – has fewer quality controls in its construction and evaluation as it is set up and corrected only by the school teacher. Part of the grade on the school exam is also earned during the year in the form of intermediate tests and assignments provided by the same teacher. The student's final grade at the end of secondary education consists of the arithmetic average of both exams.

This Dutch setting creates an environment that enables us to test whether increased resources improve student performance (which should be reflected in improving central exam results) or stimulate cheating by schools (which should be reflected in a growing gap between both exams). In other words, since schools only have discretion over the difficulty and grading standard of the school exam, the results of the central exam can be employed as a benchmark uniformly applied to all pupils in all schools, against which to set the school exam results. Hence, one can evaluate the intertemporal evolution of schools' grading standards relative to those of the central exam (which are fixed), thereby allowing a clear test of what happens when school resources increase.

Combining the unique characteristics of the Dutch education system discussed above with the quasi-experimental setting created by the policy intervention, one obtains an ideal scenario to address the question we started out with: Are resource-driven public policies effective in improving educational outcomes? Particularly, one can exploit this setting to compare educational outcomes (i.e. students' results on the central and school exams) in Dutch schools inside the 40 districts covered by the new legislation (the "treated" group) with those not covered by the new legislation (the "control" group) before/after the policy intervention in 2007. In technical terms, this approach is called a difference-in-differences approach because one effectively estimates whether the treatment (here, the policy intervention) causes a significant difference in the difference (if any existed) between the treated and untreated groups before and after a treatment. The careful reader will notice at this point that the selection of the control group – i.e., the set of schools with which those in the "treated" districts are compared – is critically important for this procedure to generate valid inferences. We therefore implemented various selection procedures to establish our control group – but our results remained similar to those reported below in all cases.

Our findings show that on average there is a stronger decline in central exam results in schools in "treated" districts (compared to schools in the control group). While initially surprising, this finding may be linked to the fact that being singled out as a "disadvantageous neighborhood" by the government of one's country may have some stigmatizing and de-moralizing effect on students and teachers. Interestingly, however, these same schools on average show a limited improvement in school exams (once again compared to schools in the control group). Schools in "treated" districts thus seem to have maintained their school grade average despite their students performing worse on the centrally administered exam, which strongly suggests that they engaged in a substantial inflation of their school exam grades after the policy intervention. This confirms that higher resources obtained from the policy program result in a pressure to show better performance. If this cannot be obtained sufficiently quickly, schools

resort to adjusting their grading structure to suggest an unrealized improvement.

When the size of the additional resources (ranging, as mentioned, from €1.2 million to €29.3 million, or €333 to €3.995 per resident) is accounted for, the results are somewhat more nuanced. The surprising negative effect on central exam results discussed above indeed changes into a positive effect when the investment surpasses approximately €17 million (or circa €2.000 per resident). In other words, resources appear beneficial in terms of improving central exam results when the additional funds are sufficiently elevated, but tend to induce grade inflation when the resources were limited. From a more general policy perspective, this suggests that policy programs aimed at improving education outcomes may easily fail to reach pre-set targets when the apportioned resources are overly limited. True, rather than feigned, success requires sufficiently elevated funds.

References

Bonesrønning, Hans: "The Variation in Teachers' Grading Practices: Causes and Consequences". In: Economics of Education Review, 1999, No. 18, pp. 89–105.

De Witte, Kristof/Geys, Benny/Solondz, Catharina: "Public Expenditures, Educational Outcomes and Grade Inflation: Theory and Evidence from a Policy Intervention in the Netherlands", 2012, mimeo.

Hanushek, Eric A.: "The Failure of Input-Based Schooling Policies". In: Economic Journal, 2003, Vol. 113, February, F64–F98.

Holmlund, Helena/McNally, Sandra/Viarengo, Martina: "Does Money Matter for Schools?" In: Economics of Education Review, 2010, Vol. 29, No. 6, pp. 1154–1164.

Wikström, Christina/Wikström, Magnus: "Grade Inflation and School Competition: An Empirical Analysis based on the Swedish Upper Secondary Schools". In: Economics of Education Review, 2005, Vol. 24, No. 3, pp. 309–322.



Benny Geys ist Associate Professor an der Norwegian Business School BI in Oslo und Gastwissenschaftler im WZB-Projekt The Future of Fiscal Federalism in Germany and Europe, das er bis zu dessen Abschluss Ende 2012 koordiniert. (Foto: Nina Bonge).

benny.geys@bi.no