CPB Netherlands Bureau for Economic Policy Analysis

CPB Discussion Paper | 347

Are good researchers also good teachers? The relationship between research quality and teaching quality

Ali Palali
Roel van Elk
Jonneke Bolhaar
Iryna Rud

Are good researchers also good teachers? The relationship between research quality and teaching quality

Ali Palali*[1], Roel van Elk[1], Jonneke Bolhaar[1], and Iryna Rud[2]

[1]CPB Netherlands Bureau for Economic Policy Analysis
[2]The Top Institute for Evidence Based Education Research, TIER, Maastricht University

March 15, 2017

Abstract

We investigate the relationship between research quality and teaching quality using data from Maastricht University, the Netherlands, where students are randomly allocated to different teachers within the same course. We measure research quality by the publication records of the teachers and teaching quality by both student evaluations of the teachers and final student grades. We find that being taught by teachers with high quality publications leads to higher grades for master students. This is not fully reflected in the student evaluations of teachers. Master students do not give higher scores to teachers with high quality of publications, bachelor students give lower scores.

Keywords: research and teaching, student grades, teacher evaluations

JEL codes: I23, I28

*Corresponding author. Email: a.palali@cpb.nl

1 Introduction

There is a continuous discussion, both academic and public, on how research and education in universities are related (Hattie and Marsh (1996); uz Zaman (2004); Jenkins et al. (2007); Elken and Wollscheid (2016)). It is often questioned whether a 'good researcher' implies also a 'good teacher', whether teachers who conduct research are more effective than those teachers who do not do research, and, in general, whether research and teaching activities can complement each other. Even though answers to these questions are important for university stakeholders to find the most efficient way in distributing human resources between research and teaching activities, the current empirical evidence is limited and mixed. Furthermore, evidence on the relationship between research and teaching is important for providing a better insight into the effective production of research output and student learning outcomes.

This study aims to analyse the relationship between research quality and teaching quality. We use individual-level data from the School of Business and Economics at Maastricht University in the Netherlands, where students are randomly allocated to different teachers within the same course. At the end of the course, students take the same exam. This enables us to exploit the exogenous variation in research quality of teachers on student outcomes such as grades and student evaluation scores. We measure research quality by the publication records of the teachers. Our results show that master students who are taught by teachers with high quality publications score higher grades. We do not find any effect for having any publications or total number of publications. Therefore, quality seems to matter rather than the quantity. The results on student grades are not fully reflected in how students evaluate their teachers. Master students do not give higher scores to teachers with higher number of publications or higher quality of publications. Bachelor students give lower scores.

Empirical evidence on the relationship between research quality and teaching quality obtained from using data on randomized experiments is scarce. Our study contributes to the existing literature by using rich individual-level data on students who are randomly assigned to different teachers within the same course. The second contribution of

this study lies in exploring and comparing two different measures of teaching outcomes: student evaluations of the teachers and final student grades.

This paper is organized as follows: In section 2, we discuss possible mechanisms underlying the relationship between research and education. In section 3, we discuss previous literature on the relationship between research quality and teaching quality. In section 4 we provide an overview of the higher education system in the Netherlands and of the Maastricht University, in particular. Section 5 describes data and presents descriptive statistics. In section 6, we describe our empirical strategy. In section 7, we discuss our estimation results. Finally, section 8 concludes.

2 Mechanisms linking research quality and teaching quality

The link between research quality and teaching quality is complex and multidimensional. Based on the previous literature, we distinguish several main mechanisms that can underlie this relationship. Depending on which of the mechanisms dominates, this relationship can vary from positive to null, and even to a negative one.

The first type of mechanisms suggests a positive relationship between research and teaching via complementarity between skills (uz Zaman (2004)). Conducting research can both enhance teacher's proficiency in the subject and keep the teacher up-to-date with regards to the newest developments in the discipline. As a result, research activities would have a positive impact on teaching quality. Such skills transfer can operate not only at the level of teacher but also at the teacher-student level. For example, through involvement in teaching activities and interactions with students during classroom discussions, researchers can transfer their critical thinking and research skills to students.

The second set of mechanisms suggests a negative relationship between research and teaching. Both research and teaching activities require investment of time and effort. Being involved in one activity, for instance, the process of conducting research, usually does not allow for simultaneously spending time and effort on another activity (the process of teaching), unless one activity benefits both research and teaching (e.g. reading

a scientific paper can simultaneously contribute to research ideas and to teaching preparation). Time and effort allocated to teaching and research are also influenced by the system of incentives in academia. Research can be rewarded by universities through promotion more generously than teaching. Therefore, there can be a selective inflow into the profession, or people in academia might choose to prioritize research over teaching, and they might be more likely to build career in academia by doing research, while teaching is often regarded as "punishment" (Walstad and Allgood (2005); Cretchley et al. (2014); De Philippis (2015)). Furthermore, contrary to the first mechanism teaching and research might require different set of skills. If research requires more specific skills (e.g. synthesis, deduction) than teaching (e.g. communication, mentoring), this can lead to disparities between skill transfers. Hence, the relationship between research output and teaching effectiveness might be neutral or even negative. Which one of these mechanisms dominates the others is an empirical question.

3 Previous literature

The literature on the relationship between research quality and teaching quality is rather large. However, most of these studies focus on the correlation between the two without making causal claims. Below we give an extensive review of the literature by first focusing on the research and teaching quality measures and then focusing on the findings between these two measurements. Since the liteture is extensive, we present some of the relevant papers in Table 1.[1]

3.1 Measures of research output and teaching quality from the existing literature

It is generally not straightforward how to measure research quality and teaching quality. Even if the research quality can be to some extent observed and summarized based

[1]Studies are not included to this table if they are (a) published earlier than 1980; (b) descriptive (non-empirical); (c) analyze exclusively the link from teaching to research output; (d) based on the analyzes of teachers(students) believes(views) on the research-teaching relationship; (e) the number of observations in the analyzes is lower than 20.

on (produced) research outcomes, teaching quality cannot be directly observed. In our study we measure the research quality by publication records of the teachers. We have information on not only how many publications that a teacher had in a certain year, but also if these publications appeared in A, B or C level journals. Therefore, we can differentiate between quantity of publications and quality of publications. For teaching quality we use two different measures: student evaluation of the teachers and student grades. Whereas student evaluations capture how students perceive teaching quality of their teachers, student grades capture the final learning experience.

We are aware of the fact that there are several different measures prosposed in the literature. Research quality has been traditionally measured by the number of published and refereed articles (see e.g. Gottlieb and Keith (1997)), by citation scores (see e.g. Rothman and Preshaw (1975)), by impact factor (Saha et al. (2003)) and by the combined measures of quantity and quality of publications (Lanjouw and Schankerman (2004); Hirsch (2005); Bornmann et al. (2008) develop an h-index[2] to characterize the scientific output of researchers, based on both the number of published articles and the impact of these publications). However all of these measures have been criticized and there is no consensus in the literature on which measure of research quality should be considered as universal, but rather the choice of the measure depends on the particular research goals of the study. In our study, we will use publication records to construct different research quality measures that vary in emphasis on quantity and quality of publications.

Teaching quality is related to both teacher's performance and student learning outcomes. Student evaluations of teachers are the most frequently used measure to estimate teaching quality in higher education (Becker and Watts (1999); Becker et al. (2012)). This popularity is mainly explained by the availability of the data on student evaluations. This measure, however, has been increasingly challenged in the literature since it is based on the perceptions of respondents and it might not necessarily reflect true teach-

[2]Even though it is much more complex than the other measures, h-index also has been criticized in later studies for non-completeness. Several researchers, for example, argued that this index does not account for aging of citation Sidiropoulos et al. (2007); Glanzel (2006); Burrell (2007). At the same time, alternatives to the h-index, such as a g-index (Egghe (2006)), an a-index (Jin (2006)) and an ar-index (Jin (2007)), hardly overcome all drawbacks of the h-index (Bornmann and Daniel (2007); Bornmann et al. (2008)).

ing effectiveness. In particular, students often evaluate the teachers on the basis of how they enjoyed the course and on the basis of teachers' personal characteristics, and not necessarily characteristics related to teaching quality (Braga et al. (2014b); McPherson et al. (2009)). Evidence shows that student evaluations are less biased in the populations where high-skill students are over-represented (Braga et al. (2014a)). Another concern is that students are usually not obliged by the institution to provide evaluations, and they are not randomly selected, which leads to a biased assessment of teacher quality. In other words, students who eventually fill in evaluation forms represent a selected sample of all students. Salomons and Goos (2014) quantify the direction and size of selection on both observable and unobservable characteristics of students, teachers and courses. They find that the true evaluation score is lower than the average reported, and thus the selection bias is positive. Moreover, they also conclude that taking student evaluation is not advisable when response rates are low or vary considerably across courses. Hoffmann and Oreopoulos (2009) suggest using the mean of the averages for teacher evaluations across classes, as this ensures that teacher quality measures differ only when instructors differ. Finally, Emery et al. (2003) and Becker and Watts (1999) advise not to use student evaluations as the only measure of teaching quality.

The alternative to student evaluation of teachers is actual student grades which are directly informative about student learning. However, using student grades in empirical analyses is not without problems either. Some studies point at the fact that teachers can inflate grades for the purpose of elevating student evaluations (see e.g. Krautmann and Sander (1999); Johnson (2003); Carrell and West (2010)). This tendency is related to characteristics of departments and teacher-specific characteristics (Jewell et al. (2013)), whereas teacher-specific characteristics explain relatively much more variation in grade inflation. Jewell et al. (2013) explain this by the universal tendency of the universities to use student evaluation scores as inputs into tenure and promotion decisions, and therefore teachers are likely to inflate grades rationally. This causes many universities to collect student evaluations before final exams. Another criticism about using student grades is that student performance can be influenced by different characteristics of students, not related to teaching effectiveness (Berk (1988, 2014)). Keeping these criticisms in mind,

we use both student evaluations and student grades as teaching quality measures. By doing so, we also shed more light on the differences between the two measures.

3.2 Existing evidence on the relationship between research quality and teaching quality

The existing literature on the relationship between research and teaching is primarily limited to correlational studies. From an extensive review of empirical literature, uz Zaman (2004) concludes that the correlation between research and teaching varies from -.4 to +.8. This broad range of findings can be explained by different measures of research quality and teaching quality, by differences in applied empirical strategies and by a variety of exogenous and endogenous factors influencing this relationship, such as discipline or the ability level of students. More recent studies extend the previous correlation literature by controlling for different educational settings (e.g. discipline, institution type, student group size, level of studies), characteristics of teachers (e.g. age, academic rank), and characteristics of students (e.g. gender, ability of students) (see e.g. Zamorski (2002); Bettinger and Long (2005); Arnold (2006); Cherastidtham et al. (2013)). Nevertheless, evidence obtained from these studies is mixed.

The relationship between research quality and teaching quality can differ across countries and educational systems. Whereas the vast majority of previous research on the relationship between research quality and teaching quality has been conducted for the United States, there is recently a growing empirical evidence on this relationship from other countries, such as Korea, Italy, the Netherlands, and Australia (Cherastidtham et al. (2013); Braga et al. (2014a); Arnold (2006); Bak et al. (2015)). For the Netherlands, Arnold (2006) examines the relationship between research quality and teaching quality at the Faculty of Economics, at Erasmus University of Rotterdam. He creates a measure of research quality based on the information whether academic staff meets the criteria for a research fellowship of the graduate school and research institute. He measures teaching quality by student evaluations of teachers. The study finds a negative correlation between research quality and teaching quality for the first and second year

bachelor courses, while this relationship is positive for the third year bachelor courses and for the master courses.

Despite controlling for different observable factors influencing research and teaching and accounting for potential non-linearity of the relationship, most of studies on the relationship between research and teaching suffer from endogeneity problems, in particular due to selection issues (i.e. self-selection of teachers to research and teaching activities and self-selection of students to different teachers). In the recent years, more data on random assignment of students to different teachers in higher education have become available and enabled researchers to analyze different aspects in higher education (see Carrell and West (2010); Braga et al. (2014b,a); Feld and Zulitz (2016)). However, causal research on the relationship between research quality and teaching quality is still scarce. The only exception is a study by Braga et al. (2014a) who use data on students who are randomly assigned to different professors at Bocconi University, Italy to investigate the relationship between research and teaching. They find that professors who are more productive in research are likely to be less effective as teachers, when output is measured by the h-index. The effect is reversed using yearly citations, however it is insignificant.

4 Higher education in the Netherlands and Randomization of students

4.1 Higher education in the Netherlands

The system of higher education in the Netherlands is characterized by self-governance, autonomy of the universities and the unity of research and teaching. In Dutch universities, the share of time spent on research and teaching is usually fixed by the contract.[3]

The data we use in this study come from the School of Business and Economics (SBE) of Maastricht University (UM), one of the biggest higher education institutions in the

[3]Based on self-reported information from academic personnel at Dutch universities (n=4243), it follows that the share of working time spent on conducting research for PhD candidates is above 70 percent, for postdoctoral researchers is above 50 percent, for assistant professors and associate professors is between 20 and 25 percent, and for professors is below 20 percent. The rest of the contract time is usually spent on teaching and organizational tasks (de Goede and Hessels (2014)).

country with over 15 000 students. There are around 4200 students enrolled in one of the programs at SBE with a high percentage of international students (around 40%). The vast majority of the bachelor programs last 3 years in contrast to bachelor programs at, for example, the U.S. Universities which last 4 years. Most of the students continue their studies with a master program which lasts only one year. The teaching strategy at the UM provides a unique opportunity to investigate the effect of research quality on teaching. Students follow weekly or every two weeks lectures in both bachelor and master programs as in most educational institutions, generally taught by the senior staff at the departments, which are called course coordinators. Later in that week students participate in tutorials supervised by other teachers. All tutorials make use of Problem-Based Learning approach, which is an important component of the teaching philosophy of the UM (Bastiaens and Nijhuis (2012)). This approach emphasizes on personal skill development, including problem solving, group work and self-directed learning. Each tutorial can have at most 16 students, which means that each course at SBE can have several tutorials taught by different teachers.

At the end of the course, the students who are taking the same course, also take the same exam even though they participate in tutorials taught by different teachers. The exam is generally prepared by the course coordinator, and it is for almost all courses in the form of a written exam. In order to ensure objectivity, the grading is done collectively by the tutorial teachers. General practice is that each teacher grades a part of the exams from all students instead of grading only the exams of the students in their tutorials. Before taking the final exam students fill in online evaluation surveys to indicate their opinions about course and tutorial teachers as well as the completed course. The teachers receive the evaluation scores after the final grades are published online. The final grades are given in a scale of 1-10. The passing grade is 5.5.

4.2 Randomization of students and teachers

Allocation of students into tutorial groups is done by the Scheduling Department at SBE via a computer program. Before the start of the academic year students register for the courses that they want to follow. In bachelor programs most of the courses are

mandatory for students, whereas in masters programs students can choose among a large variety of courses. Once the online registration closes, all students taking the same course are randomly assigned to tutorial groups by a computer program. Afterwards, tutorial teachers are randomly assigned to tutorial groups within a course.[4] Finally, the list of students in each tutorial group and the corresponding teachers is published by the Scheduling Department. Even though they are assigned to different tutors, all students taking the same course take exactly the same exam at the end of the course. Feld and Zulitz (2016) present more detailed information about the procedure used by SBE and perform several estimations to check the random assignment of students into tutorial groups. The authors show that randomization of students works successfully.

5 Data

We received a data set for more than 9000 students in BA and MA programs at UM in the years 2011, 2012 and 2013. This data set includes information on student grades, courses, programs at which the students were participating and several background characteristics such as age and nationality. In total, this data set has 80 000 student-course-grade observations.

For students who filled in course evaluation forms we also received information on tutorial groups and teachers. However, not every student fills in the evaluation forms, and for those who do not fill in evaluation forms there is no tutorial and teacher information. This means we can only use information about students who fill in evaluation forms, which decreases the number of observations from 80 000 to 28 000. The first panel in Table 2 shows the descriptive statistics for student characteristics for the whole sample and for those who filled in evaluation surveys. The last column presents the p-values for mean differences between the two groups. It shows that the difference between the two samples is significant for many characteristics. Overall students with higher grades, female students and older students are more likely to fill in the evaluation surveys. The

[4]It is expected that certain teachers are assigned to certain courses based on their expertise. However, they are randomly assigned to tutorial groups within one course which does not invalidate our randomization.

difference between average grade for students who fill in the evaluation forms and those who do not is 0.4. Even though better students are more likely to fill the evaluation forms, the difference is not very large considering the grades are given in a scale from 1 to 10.[5]

The publication records are also obtained from SBE. For teachers who have worked at UM the entire observed period (2008-2011), we obtain information on publication record. These records show how many publications that a teacher had in A, B or C level journals in a certain year.[6] Since our measure of research quality entirely depends on such publication records we make certain choices with regards to the measurement. First, instead of using the publication records at each year separately we calculate the total number of publications in the last 4 years for each year so that our measurements would suffer less from possible outliers. This means, for example, for a teacher who was teaching in 2011 we use information on the publication records from 2008 to 2011. SBE does not keep track of the publication records of teachers who did not work at UM for the entire period or that of PhD students. Therefore, we can use only a subset of the initial student data set. This subset consists of 5934 student-course-grade observations. In total there are 176 different courses, which gives 408 course-year combinations. 69 of these course-year combinations have multiple tutorial groups taught by different teachers. There are 1127 tutorial groups taught by 83 different teachers. The second panel in Table 2 shows the descriptive statistics for students who filled in evaluation surveys but excluded due to limited information on publications and for students who are included in the final analytical sample of 5934 observations. Although several characteristics are significantly different, student grades do not differ significantly between both groups. Later in Section 7, we discuss these selections more in detail.[7]

Table 3 shows the distribution of teachers according to the number of publications.

[5]That being sad our results still need to be interpreted with caution as we obtain results for slightly better and maybe more motivated students in general.

[6]The list of all journals and corresponding classifications used by SBE are given in Table 16 in Appendix 9. SBE's main strategy in deciding on journal classification is to use 5-year impact factor of the (S)SCI listed journals.

[7]Note that we do not explicitly deal with students who drop out. We can not directly observe if a student drops the course once s/he learns in which tutorial group s/he sits. In order to have an idea about such cases we assume that a student can be classified as a drop out if that student registers to a course but do not make the exam at the end. In the initial sample of 80000 observations, only 7% of the students register to a course but do not make the exam. This number is less than 1% for the analytical sample of 5934 observations.

There are 83 teachers in total. 35 of them had at least one A publication, 59 had at least one B and 60 had at least one C publications in the last 4 years. 15 teachers had one A publications, 8 teachers had two A publications, and so on. Table 4 shows the interactions between different publications. No teacher had only A publications, 4 teacher had only B publications and 6 teachers had only C publications. Finally, 30 teachers had A, B and C publications in the last 4 years.

In our empirical analyses we make a distinction between bachelor and master students as the research quality of teachers might have heterogeneous effects on students due to the differences in course types (mandatory vs. selective courses; general topics vs. specialized topics), student motivation, etc. Tables 5 and 6 present the descriptive statistics of student and teacher variables used in the empirical analysis by differentiating first year bachelor students, second and third year bachelor students and master students.[8] Descriptions of these variables are given in Table 7. Student grades are on average 6.6 for the first year bachelors, 7.2 for the second and third year bachelors and 7.3 for master students. There is almost no professor teaching in the first year bachelor courses.[9]

5.1 Tests for sample selection and randomization of students to teachers

As noted earlier we have a significant selection in our data because of selection of students and restrictions due to teacher information. In order to investigate the selection of students we perform two descriptive analysis. In the first analysis we regress the probability of filling in evaluation surveys on student characteristics for all of the students and then separately for bachelor and master students. Table 8 presents the results. In all columns the results show that students characteristics are significant in probability to fill in evaluation surveys. In the second analysis we regress the probability of being in the

[8]Relatively high number of observation for master program is due to the fact that we use information on teachers (non-PhD students) who worked at UM for the entire period of 2008-2013

[9]Note that there are two different Professor positions. The difference between the two is that the first group has more management responsibilities, has better publication records and rewarded with a better salary.

analytical sample for those who filled in evaluation forms.[10] The results are presented in Table 9. None of the student characteristics are found to be statistically significant. Therefore, the observations that we lose due to the restrictions on teacher information are not systematically different than those included in our analytical sample.

Since randomization of students into teacher groups is the underlying identifying mechanism, we perform a randomization check. In order to see if the randomization of students to different teachers successfully works, we regress teacher specific publication variables on student characteristics. Table 10 presents the results. In the first column we regress the probability of having any publication in the last 4 years on student characteristics. In the following three columns we regress the probability of having any publications in A, B or C level journals on student characteristics. None of the student characteristics is significant. Therefore we conclude that, in terms of publication performance of teachers, randomization works successfully.

6 Model

We investigate the effect of research quality of teachers on teaching quality measured by student evaluations of the teacher and student grades. For student grades we have individual data, and we use the following regression:

$$G_{ictg} = \beta_0 + \beta_1 P_{tg} + \beta_2 T_{tg} + \beta_3 S_{it} + \beta_4 C_{ct} + u_{ictg} \tag{1}$$

where G_{ictg} is the grade of student i, in course c and year t and at the tutorial taught by teacher g. P_{tg} is the publication record of teacher g, in year t. Similarly, T_{tg} is the set of other teacher characteristics. S_{it} is the student characteristics in year t, which also includes a program fixed effect for the program that students are enrolled in. C_{ct} is the course fixed effect for course c in year t.[11] u_{ictg} is the error term. Since students

[10] Note that only a part of the observations is used in the analytical sample due to unavailability of data for certain teachers.

[11] Since the courses are taught by senior members of the department, course coordinators, and tutorials are taught by different teachers, course fixed effects also capture course coordinator effects. This is important because course coordinators can be different when it comes to how rigorous they are about exam questions, course structure or tutorial guidelines. By controlling for such course fixed effects, we

are randomly assigned to different teachers after they make a course choice and take the same exam at the end of the course, we can interpret the coefficients of P_{tg} and T_{tg} as causal, conditional on C_{ct}. In all estimations we use robust standard errors clustered on the course-year level because of a possible correlation between the outcomes of students choosing the same course.

For student evaluations of the teachers, we have data only on the teacher level unlike the data on student grades. In other words we know the average evaluation score that a teacher receives after the course ends. Therefore, we cannot perform the same individual level analysis as in the student grades analysis. In order to investigate the evaluation scores we use averages of the all variables on tutorial (teacher) level. The regression equation is

$$E_{ctg} = \beta_0 + \beta_1 P_{tg} + \beta_2 T_{tg} + \beta_3 S_{tg} + \beta_4 C_{ct} + v_{ctg} \tag{2}$$

where E_{ctg} is the average teacher evaluation score in course c, year t and at the tutorial taught by teacher g. P_{tg} is the publication record of teacher g, T_{tg} is the set of other teacher characteristics in year t. S_t is the average of student characteristics in tutorial g in year t. C_{ct} is again the set of fixed effects for course c in year t. v_{ctg} is the error term.

The course structure in the bachelor and master programs are different. Courses in bachelor programs tend to be general introduction courses on various topics. Courses in master programs, on the other hand, are mostly specialized courses. When a teacher gives a course on the specific topic that s/he specializes in, we expect the expertise and motivation to be different. Therefore, in our estimations we run the above-mentioned model first for all students, and then for bachelor and master students separately.

7 Results

First, we present the results of individual level student grade estimations. The results are displayed in Table 11. The first column displays the results for all students, the rest

achieve identification through within variation- variation due to the different teachers in different tutorials for the same course.

of the columns present the results for first year bachelor students, second and third year bachelor students and finally for master students, respectively. In these analyses, research quality is a dummy variable which is 1 if the teacher had any publications in the last 4 years. Therefore, we measure the effect of having any publication activity regardless of the quality of publication. For all specifications, there is a small positive but insignificant effect on student grades.

Table 12 presents the results ofl other analyses using different publication variables to measure research quality. In row one, the coefficient estimate for total number of publications in the last 4 years shows that the number of publications has no effect on student performance. In this estimation we measure the effect of total publication activity ignoring the quality of publications. In row 2, the research quality variable is a dummy variable which is 1 if the teacher had any A level publications in the last 4 years. Hence, in row 2 we measure the effect of having a teacher who conducts high quality research. The coefficient estimate for master students shows that there is a significant positive effect on student grades. Having a teacher with at least one A level publication in the last four years in associated with a 0.4 higher student grade. This suggests that in master programs students taught by teachers with high quality of publications perform better whereas students of teachers with more publications do not. Thus, quality seems to be more important than quantity. In row 3 the research quality variable is a dummy variable which is 1 if the teacher had any B level publications in the last 4 years. We find smaller insignificant positive effects. Comparing these results with the one in Table 11 shows that as the research quality of the teacher increases, student performance increases, only for master students. In row 4 we estimate the effect of total number of A publications on student grades. Again, for master students we find a positive significant effect. Having one more A level publication increases the student grades by 0.2 on average. Therefore, quantity of publications only matter for A publications.

Our baseline results show that mechanisms suggesting a positive relationship between research and teaching, such as skill transfers and teacher-student interactions, dominate the ones suggesting a negative relationship such as time and effort allocation. We believe that the discrepancy between the results for bachelor and master students further

strengthens this interpretation. Finding a stronger effect for master students is not particular to our analysis (see Arnold (2006)) and can be explained by the course characteristics in the bachelor and master programs. Most of the courses in the bachelor programs are mandatory courses on introductory level. However, master courses can be elective ones, more specialized on certain topics and followed by students who are more interested and motivated. It is also generally the case that teachers give special topic courses which primarily focus on their field of interest. This can increase the effects of skill transfers and the effects of interactions between teachers and students.

In Table 13 we present the results of some sensitivity analyses where we introduce more covariates using student level and tutor level information. For these sensitivity analysis we use the specification in the second row of Table 12.[12] In panel 1, we introduce tutor-student gender combinations. The coefficient estimates show that male students taught by female teachers perform worse in comparison to male students who are taught by male teachers in the master program. In panel 2, we add peer variables by calculating the average age and percentage of females in the classroom for peers of students. We perform this analysis because peer effects can be important in the classroom. The coefficient estimate of the publication variable remains the same. Finally, in panel 3 we add variables to capture the academic position of the teachers. The reference group is lecturers. The correlation between publications and positions is very high. This is of course expected as the decision to promote an assistant professor to associate professor position, for example, mainly depends on the publication records of the academic. Once we control for the position variable, the variation in the publication variable becomes very small. This is reflected in the higher standard error for the publication variable.[13][14]

In Table 14 we present the results of teacher evaluations. The coefficient estimate for having any publication shows that teachers with publications receive lower evaluation

[12] Although we perform the sensitivity analysis for all of the other specifications, we present only the results of the third specification due to high significant effect of research quality measure. All other results remain the same once we introduce more variables, and they are available upon request.

[13] We choose to include the variables on academic position only in a sensitivity analysis because of correlation between publication records and academic positions. This correlation is not surprising as decisions on promotions/tenure largely depend on publishing performance.

[14] When it comes to the question of whether we can control for experience of the tutor, we can do that partly by including the age of the tutor. In all our estimations we control for the age. Therefore, we believe that we partly control for the work experience.

scores on average although the coefficients are estimated imprecisely.[15] Table 15 presents the results for other publication measures. The results for master students show that coefficient estimates are positive but small, indicating that teachers with high quality publications do not receive higher evaluation scores. The teachers with publications on the other hand receive lower scores in the second and third year of the bachelor programs.

The difference between the results concerning students grades and student evaluations is important. As mentioned earlier when students evaluate the teachers, they do not necessarily evaluate the teaching effectiveness. Evaluation scores might reflect the personality of the teacher or in general personal experience in the classroom. A rigorous demanding teacher for example might end up with a lower score compared to a fun but not much better teacher. This can explain the smaller or even the negative results for evaluation scores estimations. This is in line with some of the previous findings in the literature (see Emery et al. (2003)). Student grades, on the other hand, might reflect true learning experience and can be more informative in measuring teaching effectiveness.

8 Conclusion

There is a continuous debate about the relationship between research quality of academicians and their teaching performance. Are good researchers also good teachers? Answering this question is important not only for scientific merit but also for policy-making, especially for higher education stakeholders as the answer can help them in distributing human resources more efficiently between research and teaching.

In this paper we investigate the relationship between research quality and teaching quality. We use data from Maastricht University, the Netherlands, where students are randomly allocated to different teachers even though they all take the same exam. The research quality is measured by the publication records of the teachers. The teaching quality is measured by both student grades and student evaluations of the teachers. Exploiting the random allocation of students to different teachers, and the fact that students with different teachers make the same exam, we find that master students who are taught

[15]The high positive coefficient estimate for the first year bachelor students is most probably due to the low number of observations.

by teachers with high quality publications score higher grades. However, we do not find any effect for having any publications or total number of publications. This shows that quality matters when it comes to student performances, and the quantity matters only if the quality is good because only for A publications the number of publications has a significant positive effect on student grades. Moreover, we believe that the stronger results for master students strengthen our interpretation of the findings. The vast majority of the courses in the bachelor programs are mandatory courses on introductory level. However, master courses can be elective ones, they are much more specialized on certain topics, generally in the interest areas of teachers, and followed by students who are more interested and motivated. This can increase the effects of aforementioned skill transfers and the interactions between teachers and students in the classrooms.

The results based on course evaluations show that the findings from student grades estimations are not fully reflected in how students evaluate their teachers. Master students do not give higher scores to teachers with higher number of publications or higher quality of publications. Moreover, bachelor students give lower scores to teachers with publications. The difference between the results of student grades and student evaluation scores estimations indicates that the two measures capture different things. Evaluation scores might reflect the personality of the teacher or in general personal experience in the classroom rather than learning. Hence, we conclude that it is useful to use both measures in analyzing teaching quality.

When it comes to the policy implications of our findings, one should interpret the results with caution. Our findings cannot be interpreted as evidence supporting or dismissing the argument that research and teaching at the universities should be separated. Our results also do not answer how much time the teachers should spend on teaching or research. We conclude that excellent research performance contributes to a higher teaching quality in the master programs if the quality of teaching is measured by student grades. This might suggest that if good researchers have indeed time for teaching, then they better should be allocated to courses in the master programs.

References

Arnold, I. (2006). Het beste onderwijs komt uit de ivoren toren. *Economisch Statistische Berichten*, 334–335.

Bak, H.-J. et al. (2015). Too much emphasis on research? an empirical examination of the relationship between research and teaching in multitasking environments. *Research in Higher Education 56*(8), 843–860.

Bastiaens, E. and J. Nijhuis (2012). From problem-based learning to undergraduate research: The experience of maastricht university in the netherlands. *Council on Undergraduate Research Quarterly 32*, 43.

Becker, W. E., W. Bosshardt, and M. Watts (2012). Revisiting how departments of economics evaluate teaching. Citeseer.

Becker, W. E. and M. Watts (1999). How departments of economics evaluate teaching. *American Economic Review*, 344–349.

Berk, R. A. (1988). Fifty reasons why student achievement gain does not mean teacher effectiveness. *Journal of Personnel Evaluation in Education 1*(4), 345–363.

Berk, R. A. (2014). Should student outcomes be used to evaluate teaching? *The Journal of Faculty Development 28*(2), 87.

Bettinger, E. and B. T. Long (2005). Help or hinder? adjunct professors and student outcomes. Technical report, What's happening to public higher education.

Bornmann, L. and H.-D. Daniel (2007). What do we know about the h index? *Journal of the American Society for Information Science and technology 58*(9), 1381–1385.

Bornmann, L., R. Mutz, and H.-D. Daniel (2008). Are there better indices for evaluation purposes than the h index? a comparison of nine different variants of the h index using data from biomedicine. *Journal of the American Society for Information Science and Technology 59*(5), 830–837.

Braga, M., M. Paccagnella, and M. Pellizzari (2014a). The academic and labor market returns of university professors. IZA Discussion Paper 7902, Institute for Labour Market Policy Evaluation.

Braga, M., M. Paccagnella, and M. Pellizzari (2014b). Evaluating students' evaluations of professors. *Economics of Education Review 41*, 71–88.

Brickley, J. and J. Zimmerman (2001). Changing incentives in a multitask environment: evidence from a top-tier business school. *Journal of Corporate Finance 7*, 367–396.

Burrell, Q. (2007). Hirsch index or hirsch rate? some thoughts arising from liang's data. *Scientometrics 73*(1), 19–28.

Cadez, S., V. Dimovski, and M. Zaman Groff (2015). Research, teaching and performance evaluation in academia: the salience of quality. *Studies in Higher Education*, 1–19.

Carrell, S. E. and J. E. West (2010). Does professor quality matter? evidence from random assignment of students to professors. *Journal of Political Economy 118*(3), 409–432.

Centra, J. A. (1983). Research productivity and teaching effectiveness. *Research in Higher education 18*(4), 379–389.

Cherastidtham, I., J. Sonnemann, and A. Norton (2013). The teaching-research nexus in higher education. (2013- 12).

Cretchley, P., S. Edwards, P. O'Shea, J. Sheard, J. Hurst, and W. Brookes (2014). Research and/or learning and teaching: a study of australian professors' priorities, beliefs and behaviours. *Higher Education Research and Development 33*(4), 649–669.

de Goede, M. and L. Hessels (2014). Drijfveren van onderzoekers. Technical report, Rathenau Institute.

De Philippis, M. (2015). *Multitask Agents and Incentives: The Case of Teaching and Research for University Professors*. CEP working paper 1386.

De Witte, K., N. Rogge, L. Cherchye, and T. Van Puyenbroeck (2013). Economies of scope in research and teaching: A non-parametric investigation. *Omega 41*, 305–314.

Egghe, L. (2006). How to improve the h-index. *The Scientist* (3), 14.

Elken, M. and S. Wollscheid (2016). The relationship between research and education: typologies and indicators. A literature review. , Nordic Institute for Studies in Innovation, Research and Education.

Emery, C. R., T. R. Kramer, and R. G. Tian (2003). Return to academic standards: A critique of student evaluations of teaching effectiveness. *Quality Assurance in Education 11*(1), 37–46.

Euwals, R. and M. E. Ward (2005). What matters most: teaching or research? empirical evidence on the remuneration of british academics. *Applied Economics 37*(14), 1655–1672.

Feld, J. and U. Zulitz (2016). Understanding peer effects: On the nature, estimation and channels of peer effects. *Journal of Labor Economics Forthcoming*.

Galbraith, C. S. and G. B. Merrill (2012). Faculty research productivity and standardized student learning outcomes in a university teaching environment: A bayesian analysis of relationships. *Studies in Higher Education 37*(4), 469–480.

Glanzel, W. (2006). On the opportunities and limitations of the h-index. *Science focus*.

Gottlieb, E. E. and B. Keith (1997). The academic research-teaching nexus in eight advanced-industrialized countries. *Higher Education 34*(3), 397–419.

Hattie, J. and H. W. Marsh (1996). The relationship between research and teaching: A meta-analysis. *Review of educational research 66*(4), 507–542.

Hattie, J. and H. W. Marsh (2004). One journey to unravel the relationship between research and teaching. In *Research and teaching: Closing the divide? An International Colloquium*, pp. 18–19.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences of the United States of America 102*(46), 16569–16572.

Hoffmann, F. and P. Oreopoulos (2009). Professor qualities and student achievement. *The Review of Economics and Statistics 91*(1), 83–92.

James, S. (2011). Does business school research have to be" excellent" to benefit business school teaching? *The international journal of management education 9*(2), 37–47.

Jenkins, A., M. Healey, and R. Zetter (2007). Linking teaching and research in disciplines and departments. Technical report, The Higher Education Academy.

Jewell, R. T., M. A. McPherson, and M. A. Tieslau (2013). Whose fault is it? assigning blame for grade inflation in higher education. *Applied economics 45*(9), 1185–1200.

Jin, B. (2006). H-index: an evaluation indicator proposed by scientist. *Science Focus 1*(1), 8–9.

Jin, B. (2007). The ar-index: complementing the h-index. *ISSI newsletter 3*(1), 6.

Johnson, V. E. (2003). *Grade inflation: A crisis in college education*. Springer Science & Business Media.

Krautmann, A. C. and W. Sander (1999). Grades and student evaluations of teachers. *Economics of Education Review 18*(1), 59–63.

Landry, R., M. Saïhi, N. Amara, and M. Ouimet (2010). Evidence on how academics manage their portfolio of knowledge transfer activities. *Research Policy 39*(10), 1387–1403.

Lanjouw, J. O. and M. Schankerman (2004). Patent quality and research productivity: Measuring innovation with multiple indicators*. *The Economic Journal 114*(495), 441–465.

McPherson, M. A., R. T. Jewell, and M. Kim (2009). What determines student evaluation scores? a random effects analysis of undergraduate economics classes. *Eastern Economic Journal 35*(1), 37–51.

Michalak Jr, S. J. and R. J. Friedrich (1981). Research productivity and teaching effectiveness at a small liberal arts college. *The Journal of Higher Education*, 578–597.

Moses, I. and P. Ramsden (1992). Academic values and academic practice in the new universities. *Higher Education Research and Development 11*(2), 101–118.

Noser, T. C., H. Manakyan, and J. R. Tanner (1996). Research productivity and perceived teaching effectiveness: A survey of economics faculty. *Research in Higher Education 37*(3), 199–221.

Rothman, A. and R. Preshaw (1975). Is scientific achievement a correlate of effective teaching performance? *Research in Higher Education 3*(1), 29–34.

Saha, S., S. Saint, and D. A. Christakis (2003). Impact factor: a valid measure of journal quality? *Journal of the Medical Library Association 91*(1), 42.

Salomons, A. and M. Goos (2014). Measuring teaching quality in higher education. *Tjalling C. Koopmans Research Institute Discussion Paper Series* (DP 14-16).

Sidiropoulos, A., D. Katsaros, and Y. Manolopoulos (2007). Generalized hirsch h-index for disclosing latent facts in citation networks. *Scientometrics 72*(2), 253–280.

Stack, S. (2003). Research productivity and student evaluation of teaching in social science classes: A research note. *Research in Higher Education 44*(5), 539–556.

uz Zaman, M. Q. (2004). *Review of the academic evidence on the relationship between teaching and research in Higher Education*. Department for Education and Skills London.

Walstad, W. B. and S. Allgood (2005). Views of teaching and research in economics and other disciplines. *American Economic Review*, 177–183.

Zamorski, B. (2002). Research-led teaching and learning in higher education. *Teaching in Higher Education 7*(4), 411–427.

Table 1: The relationship between research quality and teaching quality.

Authors / year	Journal	Measurement of research quality	Measurement of teaching quality	Level of analysis	Findings (-, 0, +)
Arnold (2006)	ESB, 2006	Being qualified in the research school	Student evaluations	Regression	(-) For bachelor year 1 and year 2 ; (+) For bachelor year 3 and masters
Braga et al. (2014b)	Economics of Education Review	H-index	Student evaluations	Random assignment of students to teachers, regression	(-)
Brickley and Zimmerman (2001)	Journal of Corporate Finance	Self-reported working papers and the number of new papers	Student evaluations	Regression	(-)
Cadez et al. (2015)	Studies in Higher Education	The proportion of papers published in high quality journals relative to all papers published by a particular researcher.	Student assessments	Regression	(+)
Centra (1983)	Research in Higher Education	Self-reported number of publications	Student ratings of instruction	Correlation	(0)

Table 1 continued.

De Witte et al. (2013)	Omega	University level multi- criteria Research Evaluation Score (RES-score)	Student evaluation forms	Regression, non-parametric approach	(+)
Euwals and Ward (2005)	Applied Economics	Number of publications	Self-reported teaching quality	Regression	(+)
Galbraith and Merrill (2012)	Studies in Higher Education	The number of indiscipline academic conference presentations, proceeding publications, journal publications and books.	Standardized learning outcome assessment test, developed individually for each course in each program by a committee of content experts in the subject area, with questions designed to assess each of the stated learning outcomes	Regressions, using the Bayesian classification technique and other methods	(+)
Gottlieb and Keith (1997)	Higher Education	Published articles and research preference of faculty members	Average weekly hours spent on teaching activities	Regression	(-)

Table 1 continued.

James (2011)	The international Journal of Management Education	Units of Assessment Panel made judgements about the quality of the research submission	Assessment by the Quality Assurance Agency	Correlation	(+)
Landry et al. (2010)	Research Policy	Self-reported publications	Self-reported diffusion of knowledge through teaching	Regression	(-)
Michalak Jr and Friedrich (1981)	The Journal of Higher Education	Number of publications, research papers in progress, systematic programs of study, and involvement in professional associations and conferences and exhibits.	Student evaluations, "exit interviews" with de- parting seniors, "grapevine" feedback from students, examination of course syllabi, and first-hand observation of the person's teaching.	Correlation	(+)
Noser et al. (1996)	Research in Higher Education	The number of papers nad books, paper presentations at national and regional meetings	Average teaching evaluation scores	Correlation	(+)
Moses and Ramsden (1992)	Higher Education	Self-reported number of publications, and number of research activities (weighted)	Self-reported commitment to teaching and student ratings of instruction	Correlation, multilevel analysis	(-)

Table 2: Student characteristics and sample selection.

	All students with no course evaluation					All students who provided evaluation					
	Obs.	Mean	St.Dev.	Min	Max	Obs.	Mean	St.Dev.	Min	Max	T-test
Grade	46651	6.55	1.80	0	10	27700	6.99	1.62	0	10	0.44**
Female	51739	0.35	0.48	0	1	28243	0.44	0.49	0	1	0.09**
Age	51739	20.85	2.35	16.23	44.56	28243	21.05	2.54	16.25	44.25	0.20**
Dutch	51739	0.31	0.46	0	1	28243	0.24	0.43	0	1	-0.07**
German	51739	0.45	0.50	0	1	28243	0.45	0.50	0	1	0.00
Other N.	51739	0.25	0.43	0	1	28243	0.31	0.46	0	1	0.06**
Bachelor 1	51739	0.31	0.46	0	1	28243	0.32	0.47	0	1	0.01
Bachelor 2&3	51739	0.48	0.50	0	1	28243	0.37	0.48	0	1	-0.11**
Master	51739	0.21	0.41	0	1	28243	0.31	0.46	0	1	0.10**
	All students who provided evaluation with no tutorial information					All students who provided evaluation with tutorial information					
	Obs.	Mean	St.Dev.	Min	Max	Obs.	Mean	St.Dev.	Min	Max	T-test
Grade	21913	6.95	1.64	0	10	5787	7.16	1.53	0	10	0.21
Female	22309	0.44	0.49	0	1	5934	0.44	0.49	0	1	0.00
Age	22309	20.81	2.45	16.25	44.25	5934	21.95	2.66	16.45	44.25	1.14**
Dutch	22309	0.23	0.42	0	1	5934	0.26	0.44	0	1	0.02
German	22309	0.46	0.50	0	1	5934	0.41	0.49	0	1	-0.06**
Other N.	22309	0.30	0.46	0	1	5934	0.34	0.47	0	1	0.03*
Bachelor 1	22309	0.37	0.48	0	1	5934	0.14	0.34	0	1	-0.23**
Bachelor 2&3	22309	0.38	0.49	0	1	5934	0.34	0.47	0	1	-0.03*
Master	22309	0.25	0.44	0	1	5934	0.52	0.50	0	1	0.27**

Table 3: Number of teachers and number of publications.

	A	B	C
0	48	24	23
1	15	16	17
2	8	8	8
3	6	13	5
4	3	11	8
5	2	2	3
6		3	5
7	1	3	4
8		2	2
9			4
10		1	1
11			2
12			1
At least 1	35	59	60
Total:	83	83	83

Table 4: Number of teachers and number of publications.

No publications	16
Only A	0
Only B	4
Only C	6
Only A & B	3
Only A & C	8
Only B & C	22
A, B and C	30

Table 5: Student characteristics in the analytical sample.

	Bachelor 1			Bachelor 2&3			Master		
	Obs.	Mean	St.Dev.	Obs.	Mean	St.Dev.	Obs.	Mean	St.Dev.
Grade	784	6.61	1.86	1970	7.19	1.56	3033	7.29	1.37
Female	810	0.42	0.49	2028	0.41	0.49	3096	0.48	0.49
Age	810	19.04	1.56	2028	20.83	1.77	3096	23.45	2.37
Dutch	810	0.26	0.44	2028	0.22	0.41	3096	0.28	0.45
German	810	0.47	0.50	2028	0.45	0.50	3096	0.36	0.48
Other N.	810	0.27	0.44	2028	0.33	0.47	3096	0.36	0.48

Table 6: Teacher characteristics in the analytical sample.

	Bachelor 1			Bachelor 2&3			Master		
	Obs.	Mean	St.Dev.	Obs.	Mean	St.Dev.	Obs.	Mean	St.Dev.
Evaluation score	810	8.01	1.26	2028	7.91	1.29	3094	7.93	1.14
Female	810	0.11	0.31	2028	0.08	0.28	3094	0.17	0.37
Dutch	810	0.48	0.5	2028	0.57	0.49	3094	0.59	0.49
Age	810	45.62	7.15	2028	46.44	8.62	3094	43.65	9.8
Professor 1	810	–	–	2028	0.05	0.21	3094	0.16	0.44
Professor 2	810	0.01	0.07	2028	0.09	0.29	3094	0.06	0.24
Assistant Prof.	810	0.11	0.31	2028	0.19	0.39	3094	0.35	0.49
Associate Prof.	810	0.13	0.34	2028	0.2	0.43	3094	0.22	0.47
Lecturer	810	0.76	0.43	2028	0.53	0.59	3094	0.27	0.44

Table 7: Descriptions of student and teacher specific variables.

Grade	Final exam grade of the student, from 1 to 10.
Evaluation score	Evaluation score of the teacher in a tutor group, from 1 to 10. "Evaluate the overall functioning of your tutor with a grade from 1 to 10. "
Female	1 if the student/tutor is female
Age	Age of the student/tutor in the academic year
Dutch	1 if the student/tutor is Dutch
German	1 if the student is German
Other N.	1 if the student/tutor is from another country
Professor 1/2	1 if the teacher is appointed as a professor
Assistant Prof.	1 if the teacher is appointed as an assistant professor
Associate Prof.	1 if the teacher is appointed as an associate professor
Lecturer	1 if the teacher is appointed as a lecturer

Table 8: The effect of student characteristics on survey response rate.

	(1) All students	(2) Bachelor 1	(3) Bachelor 2 & 3	(4) Master
Female	0.067***	0.080***	0.040***	0.093***
	(0.004)	(0.006)	(0.005)	(0.008)
Grade	0.032***	0.025***	0.038***	0.029***
	(0.002)	(0.003)	(0.002)	(0.003)
Age	-0.002	-0.008**	-0.008***	0.008***
	(0.001)	(0.003)	(0.002)	(0.002)
Dutch	-0.051***	-0.080***	-0.054***	-0.012
	(0.005)	(0.008)	(0.006)	(0.010)
Other N.	0.029***	0.019**	-0.015	0.088***
	(0.005)	(0.006)	(0.008)	(0.010)
Constant	0.330***	0.166*	0.257**	-0.198
	(0.062)	(0.074)	(0.078)	(0.159)
Course FE	Yes	Yes	Yes	Yes
Program FE	Yes	Yes	Yes	Yes
N	74351	23277	32654	18420

Robust standard errors clustered on course-year level are in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9: The effect of student characteristics on sample selection due to teacher information.

	(1) All students	(2) Bachelor 1	(3) Bachelor 2 & 3	(4) Master
Female	0.001	0.006	0.003	-0.004
	(0.004)	(0.005)	(0.007)	(0.008)
Grade	0.002	0.002	0.002	0.000
	(0.001)	(0.001)	(0.002)	(0.004)
Age	0.000	-0.001	0.002	-0.000
	(0.001)	(0.002)	(0.002)	(0.001)
Dutch	-0.005	-0.003	-0.004	-0.008
	(0.004)	(0.006)	(0.007)	(0.009)
Other N.	-0.010*	0.000	-0.017	-0.014
	(0.005)	(0.005)	(0.009)	(0.010)
Constant	0.090*	0.048	0.034	0.052
	(0.040)	(0.063)	(0.087)	(0.093)
Course FE	Yes	Yes	Yes	Yes
Program FE	Yes	Yes	Yes	Yes
N	27700	8861	10239	8600

Robust standard errors clustered on course-year level are in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10: Regression of publication variables on student characteristics.

	(1)	(2)	(3)	(4)
	Any publications	Any A publications	Any B publication	Any C publication
Female	0.004	0.004	0.023	0.024
	(0.006)	(0.011)	(0.014)	(0.017)
Dutch	0.001	-0.008	-0.001	0.005
	(0.008)	(0.009)	(0.023)	(0.014)
Other N.	-0.005	0.005	-0.014	-0.030
	(0.009)	(0.014)	(0.021)	(0.022)
Age	-0.001	-0.000	0.005	0.002
	(0.002)	(0.002)	(0.003)	(0.003)
Constant	1.021***	-0.001	-0.058	6.998***
	(0.028)	(0.029)	(0.058)	(0.065)
N	5934	5934	5934	5934

Robust standard errors clustered on course-year level are in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 11: Student grades estimations.

	(1) All students	(2) Bachelor 1	(3) Bachelor 2 & 3	(4) Master
Student: female	0.070	-0.067	0.042	0.121*
	(0.042)	(0.136)	(0.072)	(0.057)
Student: Dutch	-0.425***	-0.070	-0.519***	-0.496***
	(0.059)	(0.182)	(0.118)	(0.068)
Student: other nationality	-0.699***	-0.843***	-0.721***	-0.662***
	(0.063)	(0.200)	(0.135)	(0.071)
Student: age	-0.082***	-0.106*	-0.140***	-0.062***
	(0.012)	(0.049)	(0.028)	(0.012)
Tutor: female	0.111	0.246	0.178	-0.102
	(0.108)	(0.199)	(0.250)	(0.099)
Tutor: Dutch	-0.053	-0.134	0.242	-0.183
	(0.103)	(0.258)	(0.148)	(0.152)
Tutor: age	-0.012	0.001	-0.032	-0.002
	(0.007)	(0.020)	(0.018)	(0.010)
Any publications	0.072	0.164	0.110	0.105
	(0.103)	(0.546)	(0.232)	(0.115)
Constant	9.628***	9.766***	10.868***	8.656***
	(0.480)	(1.585)	(1.024)	(1.081)
Course/year fixed effects	Yes	Yes	Yes	Yes
Program fixed effects	Yes	Yes	Yes	Yes
N	5787	784	1970	3033

Robust standard errors clustered on course-year level are in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12: Alternative publication indicators.

	(1) All students	(2) Bachelor 1	(3) Bachelor 2 & 3	(4) Master
1. Total number of publications	0.014	-0.020	0.011	0.031
	(0.020)	(0.055)	(0.034)	(0.037)
2. At least an A publication	0.146	0.154	0.023	0.381**
	(0.165)	(0.503)	(0.491)	(0.138)
3. At least a B publication	0.111	0.164	0.064	0.185
	(0.109)	(0.546)	(0.222)	(0.123)
4. Number of A publications	0.099	0.067	0.200	0.197**
	(0.052)	(0.077)	(0.276)	(0.056)
N	5787	784	1970	3033

Only the coefficient estimates of the publication record variables are presented.

Other results remain very similar to the estimates presented in Table 11.

Robust standard errors clustered on course-year level are in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 13: Sensitivities: student grade estimations.

	(1) All students	(2) Bachelor 1	(3) Bachelor 2 & 3	(4) Master
1.At least an A publication	0.155	0.147	0.023	0.407**
	(0.164)	(0.503)	(0.491)	(0.141)
Tutor Female - Student Male	-0.077	-0.032	0.096	-0.335*
	(0.141)	(0.301)	(0.280)	(0.142)
Tutor Male - Student Female	0.019	-0.128	0.034	0.049
	(0.045)	(0.136)	(0.082)	(0.057)
Tutor Female - Student Female	0.326**	0.356	0.219	0.165
	(0.118)	(0.249)	(0.248)	(0.116)
2.At least an A publication	0.168	0.468	0.029	0.374**
	(0.165)	(0.512)	(0.536)	(0.141)
Peer age	0.012	-0.058	0.026	0.002
	(0.026)	(0.089)	(0.047)	(0.031)
Peer female	0.144	0.556	0.237	-0.133
	(0.123)	(0.328)	(0.197)	(0.188)
3.At least an A publication	-0.068	-0.041	-0.036	0.255
	(0.228)	(0.409)	(0.547)	(0.262)
Professor 1	0.250	–	0.842	0.080
	(0.162)	–	(0.431)	(0.189)
Professor 2	0.009	-1.313***	-0.083	0.163
	(0.219)	(0.361)	(0.393)	(0.195)
Associate Prof.	-0.019	-0.098	0.073	-0.081
	(0.128)	(0.143)	(0.296)	(0.126)
Assistant Prof.	0.210	0.988*	0.308	-0.098
	(0.190)	(0.415)	(0.312)	(0.166)
N	5787	784	1970	3033

Only the coefficient estimates of the publication record variables are presented.

Other results remain very similar to the estimates presented in Table 11.

Robust standard errors clustered on course-year level are in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 14: Teacher level estimations: evaluation scores.

	(1) All students	(2) Bachelor 1	(3) Bachelor 2 & 3	(4) Master
Tutor: female	-0.479	-0.242	-0.396	-0.653
	(0.456)	(0.723)	(0.569)	(0.839)
Tutor: Dutch	0.329	0.824	-0.206	0.475
	(0.287)	(0.507)	(0.416)	(0.580)
Tutor: age	-0.038	0.082**	-0.091	-0.048
	(0.025)	(0.028)	(0.061)	(0.027)
Share(Dutch)	0.123	0.365	-0.410	0.691
	(0.250)	(0.574)	(0.331)	(0.391)
Share(Other nationality)	0.243	-0.145	0.025	0.610
	(0.225)	(0.575)	(0.311)	(0.335)
Average(age)	0.063	0.158	-0.013	0.102
	(0.053)	(0.189)	(0.081)	(0.075)
Share(female)	-0.303	-0.639	-0.511	0.151
	(0.224)	(0.490)	(0.300)	(0.395)
Any publications	-0.264	1.908	-0.428	-0.150
	(0.261)	(1.096)	(0.486)	(0.307)
Constant	7.790***	-2.800	12.700***	7.360**
	(1.790)	(4.236)	(2.955)	(2.359)
Course/year fixed effects	Yes	Yes	Yes	Yes
N	1127	172	459	496

Robust standard errors clustered on course-year level are in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 15: Teacher level estimations: Alternative publication variables: evaluation scores.

	(1) All students	(2) Bachelor 1	(3) Bachelor 2 & 3	(4) Master
1.Total number of publications	-0.074	0.136	-0.120***	-0.064
	(0.039)	(0.111)	(0.030)	(0.099)
2.At least an A publications	-0.249	0.431	-1.071**	0.230
	(0.445)	(1.236)	(0.385)	(0.671)
3.At least an B publications	-0.130	1.730	-0.493	0.184
	(0.270)	(0.982)	(0.477)	(0.348)
4.Number of A publications	-0.052	0.216	-0.782***	0.286
	(0.151)	(0.284)	(0.104)	(0.241)
N	1127	172	459	496

Only the coefficient estimates of the publication record variables are presented. Other results remain the same.

Robust standard errors clustered on course-year level are in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

9 Appendix

Table 16: List of journals and classifications used by SBE.

	Journals			Journals	
1	Academy of Management Executive	B	36	Contemporary Accounting Research	B
2	Academy of Management Journal	A	37	Contemporary Economic Policy	C
3	Academy of Management Review	A	38	Corporate Governance	B
4	Accident Analysis and Prevention	B	39	Decision Sciences	B
5	Accounting and Business Research	B	40	Decision Support Systems	B
6	Accounting Review	A	41	Defence and Peace Economics	C
7	Accounting, Organisations and Society	A	42	Developing Economies	C
8	ACM Trans Information Systems	A	43	Eastern European Economics	C
9	Administrative Science Quarterly	A	44	Ecological Economics	B
10	Advances in Strategic Management	C	45	Econometric Reviews	C
11	African Development Review	C	46	Econometric Theory	A
12	Agricultural Economics	B	47	Econometrica	A
13	American Economic Review	A	48	Economic and Industrial Democracy	C
14	American Journal of Agricultural Economics	B	49	Economic Development & Cultural Change	B
15	American Journal of Economics and Sociology	C	50	Economic Development Quarterly	C
16	Annals of Operations Research	B	51	Economic Geography	B
17	Applied Economics	B	52	Economic History Review	B
18	Applied Economics Letters	C	53	Economic Inquiry	B
19	Applied Mathematical Modelling	C	54	Economic Journal	A
20	Applied Stochastic Models in Business and Industry	C	55	Economic Modelling	C
21	Asia-Pacific Journal of Operational Research	C	56	Economic Policy	C
22	Auditing: Journal of Practice and Theory	B	57	Economic Record	C
23	Australian Journal of Agricultural & Resource Economics	C	58	Economic Theory	B
24	Betriebswirtschaftliche Forschung und Praxis	C	59	Economica	B
25	British Journal of Industrial Relations	B	60	Economics and Philosophy	C
26	British Journal of Management	B	61	Economics Letters	B
27	Brookings Papers On Economic Activity	B	62	Economics of Education Review	B
28	Business History	C	63	Economics of Transition	C
29	Business History Review	C	64	Economist	C
30	California Management Review	B	65	Economy and Society	B
31	Cambridge Journal of Economics	C	66	Empirical Economics	C
32	Canadian Journal of Economics	B	67	Energy Economics	B
33	China Economic Review	C	68	Energy Journal	C
34	Computational Optimization and Applications	C	69	Engineering Optimization	C
35	Computers & Operations Research (Uk)	B	70	Entrepreneurship and Regional Development	B

Table 16 continued.

	Journals			Journals	
71	Environment & Planning A	B	106	Industrial Relations	B
72	Environment & Planning B	B	107	Infor	C
73	Environment & Planning C	C	108	Information and Management	B
74	Environment & Planning D	B	109	Information Economics and Policy	C
75	Environmental & Resource Economics	B	110	Information Processing & Management	B
76	European Accounting Review	C	111	Information Processing Letters	C
77	European Economic Review	A	112	Information Research	C
78	European Financial Management	C	113	Information Sciences	C
79	European Journal of Industrial Relations	C	114	Information Society	C
80	European Journal of Information Systems	C	115	Information Systems	B
81	European Journal of Operational Research	B	116	Information Systems Frontiers	C
82	European Journal of Work & Organizational Psychology	C	117	Information Systems Journal	C
83	European Review of Agricultural Economics	C	118	Information Systems Management	C
84	Europe-Asia Studies	C	119	Information Systems Research	A
85	Expert Systems With Applications	B	120	Informs Journal On Computing	B
86	Explorations in Economic History	C	121	Insurance, Mathematics & Economics	B
87	Feminist Economics	C	122	Interfaces	C
88	Financial Analysts Journal	C	123	International Economic Review	A
89	Financial Management	B	124	International Journal of Computer Integrated Manufacturing	C
90	Futures	C	125	International Journal of Cooperative Information Systems	C
91	Games and Economic Behavior	A	126	International Journal of Electronic Commerce	B
92	Geneva Papers On Risk & Insurance: Issues and Practice	C	127	International Journal of Finance and Economics	C
93	Group and Organization Management	C	128	International Journal of Flexible Manufacturing Systems	C
94	Group Decision and Negotiation	C	129	International Journal of Forecasting	B
95	Harvard Business Review	B	130	International Journal of Game Theory	B
96	Health Economics	B	131	International Journal of Industrial Organisation	B
97	Hitotsubashi Journal of Economics	C	132	International Journal of Information Management	C
98	Human Computer Interaction	B	133	International Journal of Management Reviews	B
99	Human Relations	B	134	International Journal of Manpower	C
100	Human Resource Management	A	135	International Journal of Market Research	C
101	IIE Transactions	B	136	International Journal of Operations & Production Management	C
102	IMF Staff Papers	B	137	International Journal of Phys Distr and Log Management	B
103	Industrial and Corporate Change	B	138	International Journal of Production Economics	C
104	Industrial and Labor Relations Review	B	139	International Journal of Production Research	B
105	Industrial Marketing Management	B	140	International Journal of Research in Marketing	B

Table 16 continued.

	Journals			Journals	
141	International Journal of Selection and Assessment	C	176	Journal of Consumer Psychology	C
142	International Journal of Service Industry Management	C	177	Journal of Consumer Research	A
143	International Journal of Technology Management	C	178	Journal of Corporate Finance	B
144	International Journal of Urban & Regional Research	B	179	Journal of Derivatives	B
145	International Labour Review	C	180	Journal of Development Economics	B
146	International Marketing Review	C	181	Journal of Development Studies	B
147	International Regional Science Review	C	182	Journal of Econometrics	A
148	International Review of Law & Economics	C	183	Journal of Economic Behaviour & Organization	B
149	International Small Business Journal	C	184	Journal of Economic Dynamics & Control	B
150	International Tax and Public Finance	C	185	Journal of Economic Education	C
151	Internet Research-Electronic Networking Appl and Policy	C	186	Journal of Economic Geography	C
152	Japan and the World Economy	C	187	Journal of Economic Growth	B
153	Journal of Accounting & Economics	A	188	Journal of Economic History	B
154	Journal of Accounting Research	A	189	Journal of Economic Issues	C
155	Journal of Accounting, Auditing and Finance	B	190	Journal of Economic Literature	A
156	Journal of Advanced Transportation	C	191	Journal of Economic Perspectives	A
157	Journal of Advertising	C	192	Journal of Economic Psychology	B
158	Journal of Advertising Research	C	193	Journal of Economic Theory	A
159	Journal of African Economies	C	194	Journal of Economics	C
160	Journal of Agricultural and Resource Economics	C	195	Journal of Economics & Management Strategy	C
161	Journal of Agricultural Economics	C	196	Journal of Empirical Finance	B
162	Journal of Air Transport Management	C	197	Journal of Environmental Economics & Management	B
163	Journal of Applied Econometrics	A	198	Journal of Evolutionary Economics	C
164	Journal of Banking & Finance	B	199	Journal of Finance	A
165	Journal of Business & Economic Statistics	A	200	Journal of Financial and Quantitative Analysis	A
166	Journal of Business and Psychology	C	201	Journal of Financial Economics	A
167	Journal of Business Ethics	C	202	Journal of Financial Intermediation	B
168	Journal of Business Finance & Accounting	B	203	Journal of Forecasting	B
169	Journal of Business Logistics	B	204	Journal of Futures Markets	C
170	Journal of Business Research	C	205	Journal of Global Optimization	C
171	Journal of Business Venturing	B	206	Journal of Health Economics	A
172	Journal of Common Market Studies	C	207	Journal of Housing Economics	C
173	Journal of Comparative Economics	B	208	Journal of Human Resources	A
174	Journal of Computer Information Systems	C	209	Journal of Industrial Economics	B
175	Journal of Consumer Affairs	C	210	Journal of Information Science	C

Table 16 continued.

	Journals			Journals	
211	Journal of Information Technology	B	246	Journal of Post Keynesian Economics	C
212	Journal of Institutional & Theoretical Economics	C	247	Journal of Product Innovation Management	B
213	Journal of Interactive Marketing	C	248	Journal of Productivity Analysis	B
214	Journal of International Business Studies	A	249	Journal of Public Economics	A
215	Journal of International Economics	A	250	Journal of Public Policy & Marketing	B
216	Journal of International Marketing	C	251	Journal of Real Estate Finance and Economics	C
217	Journal of International Money & Finance	B	252	Journal of Regional Science	B
218	Journal of Labor Economics	B	253	Journal of Regulatory Economics	C
219	Journal of Labor Research	C	254	Journal of Retailing	B
220	Journal of Law & Economics	A	255	Journal of Risk & Insurance	C
221	Journal of Law, Economics & Organization	B	256	Journal of Risk & Uncertainty	B
222	Journal of Macroeconomics	C	257	Journal of Rural Studies	B
223	Journal of Management	B	258	Journal of Safety Research	B
224	Journal of Management Accounting Research	B	259	Journal of Scheduling	C
225	Journal of Management Information Systems	A	260	Journal of Service Research	C
226	Journal of Management Inquiry	C	261	Journal of Small Business Management	C
227	Journal of Management Studies	B	262	Journal of Strategic Information Systems	C
228	Journal of Manufacturing Systems	C	263	Journal of the Academy of Marketing Science	B
229	Journal of Marketing	A	264	Journal of the ACM	B
230	Journal of Marketing Research	A	265	Journal of the Japanese and International Economies	C
231	Journal of Mathematical Economics	B	266	Journal of the Operational Research Society	B
232	Journal of Media Economics	C	267	Journal of the Operations Research Society of Japan	C
233	Journal of Monetary Economics	A	268	Journal of Transport Economics and Policy	C
234	Journal of Money, Credit & Banking	B	269	Journal of Urban Economics	B
235	Journal of Occupational and Organizational Psychology	B	270	Journal of World Business	C
236	Journal of Operations Management	A	271	Kyklos	C
237	Journal of Optimization Theory and Applications	B	272	Labour Economics	C
238	Journal of Organisational Behaviour Management	C	273	Land Economics	B
239	Journal of Organisational Change Management	C	274	Leadership Quarterly	B
240	Journal of Organizational Behavior	B	275	Long Range Planning	B
241	Journal of Organizational Computing and Ec	C	276	Macroeconomic Dynamics	C
242	Journal of Policy Modeling	C	277	Management Learning	B
243	Journal of Political Economy	A	278	Management Science	A
244	Journal of Population Economics	C	279	Manchester School	C
245	Journal of Portfolio Management	C	280	Marketing Letters	C

Table 16 continued.

	Journals			Journals	
281	Marketing Science	A	316	Probability in the Engineering and Informational Sciences	C
282	Mathematical Finance	B	317	Production and Operations Management	B
283	Mathematical Methods of Operations Research	C	318	Production Planning & Control	C
284	Mathematical Programming	A	319	Psychology & Marketing	B
285	Mathematical Social Sciences	B	320	Public Choice	B
286	Mathematics of Operations Research	A	321	Public Money and Management	C
287	MIS Quarterly	A	322	Quarterly Journal of Economics	A
288	MIT Sloan Management Review	B	323	Queueing Systems	C
289	National Tax Journal	B	324	R&D Management	C
290	Naval Research Logistics	B	325	Rairo-Operations Research	C
291	Negotiation Journal	C	326	Rand Journal of Economics	A
292	Networks	B	327	Real Estate Economics	B
293	New Technology, Work and Employment	C	328	Regional Science & Urban Economics	B
294	Omega - International Journal of Management Science	C	329	Regional Studies	B
295	Open Economies Review	C	330	Research Policy	B
296	Operations Research	A	331	Research-Technology Management	C
297	Operations Research Letters	B	332	Resource and Energy Economics	C
298	Optimization	C	333	Review of Accounting Studies	B
299	Optimization Methods & Software	C	334	Review of Economic Studies	A
300	OR Spectrum	C	335	Review of Economics & Statistics	A
301	Organization	B	336	Review of Financial Studies	A
302	Organization Science	A	337	Review of Income & Wealth	C
303	Organization Studies	B	338	Review of Industrial Organization	C
304	Organizational Behavior and Human Decision Processes	B	339	Scandinavian Journal of Economics	B
305	Organizational Dynamics	C	340	Scottish Journal of Political Economy	C
306	Organizational Research Methods	B	341	Service Industries Journal	C
307	Oxford Bulletin of Economics & Statistics (Uk)	B	342	Small Business Economics	B
308	Oxford Economic Papers	B	343	Social Choice & Welfare	C
309	Oxford Review of Economic Policy	C	344	South African Journal of Economics	C
310	Papers in Regional Science	C	345	Southern Economic Journal	B
311	Personnel Psychology	B	346	Strategic Management Journal	A
312	Policy Sciences	C	347	Studies in Comparative International Development	C
313	Portuguese Economic Journal	C	348	Studies in Nonlinear Dynamics and Econometrics	C
314	Post-Communist Economies	C	349	Sustainable Development	C
315	Post-Soviet Affairs	C	350	Technological Forecasting and Social Change	B

Table 16 continued.

	Journals	
351	Technology Analysis and Strategic Management	C
352	Technovation	C
353	Telecommunications Policy	C
354	Theory and Decision	C
355	Total Quality Management and business Excellence	C
356	Tourism Management	C
357	Transport Reviews	C
358	Transportation	B
359	Transportation Journal	B
360	Transportation Quarterly	C
361	Transportation Research Part A-Policy and Practice	B
362	Transportation Research Part B-Methodological	A
363	Transportation Research Part D - Transport and Environment	B
364	Transportation Research Part E-Logistics and Transportation Review	C
365	Transportation Research Record	C
366	Transportation Science	A
367	Trimestre Economico	C
368	Urban Studies	B
369	Weltwirtschaftliches Archiv-Review of World Economics	C
370	Wirtschaftsinformatik	C
371	World Bank Economic Review	B
372	World Bank Research Observer	C
373	World Development	B
374	World Economy	C

Publisher:

CPB Netherlands Bureau for Economic Policy Analysis
P.O. Box 80510 | 2508 GM The Hague
T (088) 984 60 00

March 2017 | ISBN 978-90-5833-769-6