RUHR
ECONOMIC PAPERS

Claudia Burgard
Barbara S. Grave

Does it Pay Off to Incentivize Universities?

Performance Funding in the German Higher Education System

RWI #457

Imprint

Ruhr Economic Papers

Published by

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI)
Hohenzollernstr. 1-3, 45128 Essen, Germany

Editors

Prof. Dr. Thomas K. Bauer
RUB, Department of Economics, Empirical Economics
Phone: +49 (0) 234/3 22 83 41, e-mail: thomas.bauer@rub.de

Prof. Dr. Wolfgang Leininger
Technische Universität Dortmund, Department of Economic and Social Sciences
Economics – Microeconomics
Phone: +49 (0) 231/7 55-3297, email: W.Leininger@wiso.uni-dortmund.de

Prof. Dr. Volker Clausen
University of Duisburg-Essen, Department of Economics
International Economics
Phone: +49 (0) 201/1 83-3655, e-mail: vclausen@vwl.uni-due.de

Prof. Dr. Christoph M. Schmidt
RWI, Phone: +49 (0) 201/81 49-227, e-mail: christoph.schmidt@rwi-essen.de

Editorial Office

Sabine Weiler
RWI, Phone: +49 (0) 201/81 49-213, e-mail: sabine.weiler@rwi-essen.de

Ruhr Economic Papers #457

Responsible Editor: Christoph M. Schmidt

ISSN 1864-4872 (online) – ISBN 978-3-86788-516-4

Ruhr Economic Papers #457

Claudia Burgard and Barbara S. Grave

Does it Pay Off to Incentivize Universities?

Performance Funding in the German Higher Education System

RWI

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über: *http://dnb.d-nb.de* abrufbar.

http://dx.doi.org/10.4419/86788516
ISSN 1864-4872 (online)
ISBN 978-3-86788-516-4

Claudia Burgard and Barbara S. Grave[1]

Does it Pay Off to Incentivize Universities?

Performance Funding in the German Higher Education System

Abstract

This paper analyzes the effect of the introduction of performance-related funding in the German university sector. Starting in the 1990's, the federal states of Germany introduced incentive-based funding systems in order to increase universities' performance and efficiency. We estimate the effects of this reform on four common outcome indicators of the funding models: The number of students and graduates, which are supposed to measure teaching performance, and the number of PhD graduates and the amount of third-party funds, which quantify research output. Using a difference-in-differences estimator, our results suggest that for increasing the outcomes in teaching, a weak incentive is sufficient while the research outputs are only affected if the incentive is strong enough. We further identify different responses by university types, which shows that the results are mainly driven by technical colleges. According to our findings, it is crucial to design the funding models carefully to provide the "right" incentives and hence to achieve the underlying goal of the reform.

JEL Classification: D04, I23, O31

Keywords: Higher education funding; financial incentives; policy evaluation; difference-in-differences

December 2013

1 Claudia Burgard, StepStone and RWI; Barbara Grave, Stifterverband and RWI. – The authors would like to thank Thomas Bauer, Daniel Baumgarten, Christoph Ehlert, Sandra Schaffner, Christoph M. Schmidt as well as participants at seminars at the RWI for helpful comments and discussions. All remaining errors are our own.– *All correspondence to: Claudia Burgard, StepStone Germany and RWI, Hohenzollernstr. 1-3, 45128 Essen, Germany, e-mail: claudia.burgard@gmx.de*

1 Introduction

The higher education sector is essentially characterized by increasing international competitiveness. To cope with resulting challenges in research and teaching and to demonstrate sustainability, universities have to improve their quality and performance in research and teaching. This can be encouraged by means of incentivizing universities through a more efficient allocation of financial resources. This concept of incentive-based funding in higher education has been increasingly used in many countries of the European Union (see, e.g., Eurydice European Unit, 2008). In Germany, where tertiary education is substantially publicly funded (71 % of a university's budget are state funds according to the Federal Statistical Office, 2011a), performance-related funding[1] has also been introduced starting in 1993. By now, an indicator-based funding (IBF) system at universities are implemented in all federal states, which are the responsible authorities for education policy in Germany.[2] This system is aiming at improving universities' quality and performance via incentives.

However, to our knowledge, empirical evidence on how this IBF scheme works out in the higher education sector are widely lacking. In order to judge its effectiveness and efficiency, and hence to set targeted incentives, such evaluations are needed, though. One exception is Bauer and Grave (2011), who investigate whether the funding reform led to grade inflation in the German higher education sector. Analyzing the reform introduction in North-Rhine Westphalia, they do not find evidence for such unintended reactions. Beyond that, there are several studies on the effects of performance-related pay systems in schools in other countries, which yield mixed evidence. While some studies find positive effects on student performance (Kingdon and Teal, 2007; Atkinson et al., 2009; Lavy, 2009), another study finds a negative effect in terms of a decline in student achievement and an increase in grade inflation (Martins, 2010). Further, there exists empirical evidence pointing to undesirable strategic reactions such as teaching to the rating (Burgess et al., 2005; Jacob, 2005; Reback, 2008) or cheating (Jacob and Levitt, 2003a,b).

One reason for the introduction of a performance-based allocation of public funds lied in the comparatively low performance and inefficiency of German universities compared to e.g. the US. Consequently, the former funding system, in that a university's budget relied on its capacity and on a negotiated amount of money (negotiated between the university and the federal government), has been reformed towards a formula-based allocation model including performance-related pay schemes. The new funding schemes aim at improving the efficiency and the accountability of public spending. Reallocating a fixed amount of public funds between universities according to a transparent allocation mechanism based on a set of specific performance indicators is supposed to trigger competition for these funds and thus to increase universities' performance. However, to achieve these aims, an IBF scheme needs to be well-

[1] In the following, "performance-related funding" and "indicator-based funding" are used interchangeably.

[2] The German higher education sector comprises different types of universities and colleges that mainly differ by the type of skills they provide to the students. The technical colleges (TCs, "Fachhochschulen") aim at providing a more practical education to the students while universities ("Universitäten") and technical universities (TUs, "Technische Universitäten") offer a more theoretical education. Some universities and technical colleges focus on some kind of specialized education, e.g. colleges of arts and music (CAMs, "Kunst- und Musikhochschulen"), colleges of education (CEs, "Pädagogische Hochschulen") or colleges of public administration ("Verwaltungsfachhochschulen"). If not mentioned differently, we refer to all types of universities and colleges as universities.

designed.

Filling the gap of lacking evidence on the effect of the existing IBF system, this paper investigates the following research questions. First, does indicator-based funding increase the performance of universities? That is, does the IBF system achieve its objectives? And second, is the intensity of the incentive, i.e. the relative weight of an indicator in the IBF model, important?[3] More precisely, do universities react to certain indicators only if the respective incentive is strong enough? Hence, the paper seeks to examine if and how universities react to financial incentives set by the federal states' government. In particular, we analyze the effect of the introduction of the funding reform on different indicators that are rewarded by the funding models. We selected the most commonly used indicators, which are the number of students, graduates, and PhD graduates as well as the amount of third-party funds. The paper contributes to the literature by providing first important empirical insights into impacts of reforms of public institutions' funding in Germany, which are highly relevant for policy decisions even beyond the field of higher education.

The remainder of the paper is as follows. The next section explains the institutional background and the funding reform in more detail, while in section 3, the data and the empirical strategy are described. Section 4 presents and discusses the regression results. Finally, section 5 concludes.

2 Institutional Background and Reform Incentives

The majority of German universities is publicly funded (about 63 % in 2011) with the federal states as their most important funding bodies. More precisely, in 2009, the share of public funds in the universities' budget was 71 %, while the remaining revenues consisted of third-party funds (20 %) and operating income (9 %) (without medical faculties, according to Federal Statistical Office, 2011a).[4] A special characteristic of the German higher education system is the states' competency to shape higher education policy autonomously.

In the beginning of the 1990's, the federal states became increasingly aware of the comparably low performance and inefficiency of German universities, and hence started to reform the higher education sector. By implementing the New Public Management (NPM) in the university sector, the federal states aimed at emulating a market-like environment based on managerial instruments such as the introduction of competition, emphasizing performance reporting and increasing autonomy of the universities.[5] One major part of these reforms consisted of reallocating the universities' funding. This reallocation constituted a substantial change, since public funds provided the most important source of income for universities.

[3] The weight of an indicator in the funding model expresses the relative amount of funds that is allocated based on the performance of this indicator.

[4] Third-party funds are acquired competitively from different sources, such as firms, the German Research Foundation (Deutsche Forschungsgemeinschaft (DFG)) or the government. They are mainly used to fund research projects.

[5] For a more detailed discussion on the reform see Bauer and Grave (2011) or de Boer et al. (2007).

Table 1: **Introduction of indicator-based funding at German universities**
Overall shares of indicator-based funds of total funds

State	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
BV	-	-	1.3	1.3	2.1	2.1	2.1	2.0	2.0	2.0	2.0	2.0
BW	-	-	-	20.0	20.0	20.0	2.0	20.0	20.0	20.0	20.0	20.0
LS	-	-	-	x	x	x	x	x	x	3.1	6.1	10.2
B	-	-	-	-	-	6.0	10.0	10.0	15.0	20.0	25.0	30.0
MW	-	-	-	-	-	1.6	1.5	2.5	2.5	4.1	4.1	6.1
BR	-	-	-	-	-	-	5.0	5.0	5.0	10.0	10.0	-
HS	-	-	-	-	-	-	100.0	100.0	100.0	100.0	100.0	100.0
TH	-	-	-	-	-	-	1.0	12.8	12.8	12.8	12.8	21.1
BB	-	-	-	-	-	-	-	94.9	94.9	94.9	94.9	94.9
SA	-	-	-	-	-	-	-	-	1.0	2.0	2.0	2.5
SH	-	-	-	-	-	-	-	-	-	5.0	5.0	5.0
SL	-	-	-	-	-	-	-	-	-	-	-	10.0
ST	-	-	-	-	-	-	-	-	-	-	-	-

Note: Numbers refer to universities and can differ for other university types. The federal states of North-Rhine Westphalia, Rhineland-Palatinate and Hamburg are excluded according to the sample restrictions described in section 3. The time period depicted conforms to the observation period in our data.
x: IBF already introduced at technical colleges. BV = Bavaria, BW = Baden-Wurttemberg, LS = Lower Saxony, B = Berlin, MW = Mecklenburg-Western Pomerania, BR = Bremen, HS = Hesse, TH = Thuringia, BB = Brandenburg, SA = Sachsen, SH = Schleswig-Holstein, SL = Saarland, ST = Saxony Anhalt.

Another driving force for the funding reform were growing complaints by universities concerning lacking transparency and accountability of the traditional funding system. The budget was negotiated between the university and the federal government and was mainly based on a university's capacity and on a historically derived budget that was overrolled annually. Trying to meet these complaints, the federal states implemented a new funding system – the IBF system – that is supposed to provide more transparency, to account for the universities' performance and to foster competition between universities. The idea of the indicator-based funding is to relate the public funds received by a university to its (relative) performance. Performance is measured by a set of indicators that aim to approximate the true performance of universities. These indicators were chosen to provide incentives to the universities such that they increase efficiency, quality and performance, and also to steer the universities' behavior. For example, in order to increase the general level of third-party funds in a federal state, the federal government can set a high weight on the corresponding indicator for third-party funds in the IBF model.

Due to the federal states' competencies, not only the time of implementation but also the specific characteristics of the IBF models vary between the federal states. The respective years of the reform introduction (within the observation period in this paper) and the shares that performance-related funds make up of total funds are depicted in Table 1. While North-Rhine Westphalia and Rhineland-Palatinate were the first states to introduce an IBF system in 1993 and 1994, respectively, most federal states introduced it at the beginning of or in the mid-2000's. Saxony-Anhalt introduced the new funding system as the last federal state only recently in 2011. The share of indicator-based allocated funds varies between less than 5 % in e.g. Saxony and Mecklenburg-Western Pomerania and more than 90 % in e.g. Hesse and Brandenburg as shown in Table 1. The competition for public funds can take place either between all universities in a federal state or only between universities of the same kind (i.e., for instance, only between

universities or only between technical colleges (TC)).

The indicators used to measure the performance are related to different dimensions, mainly focusing on teaching and research. The most commonly used indicators regarding teaching are the number of students and the number of graduates, while concerning research, the number of PhD graduates and the amount of third-party funds are the prevailing criteria. Further areas covered by the indicators are the size of the staff at universities, success rates, length of study, or gender equality, which are, however, of minor importance in most states. The type and number of indicators used as well as their intensity, i.e., their relative weight to which they enter the model, vary a lot between the federal states and over time. For reduction of complexity, Table 2 lists averages of these weights over time for each state in our sample. Nine indicators are used on average, ranging from four indicators in Schlewig-Holstein and 24 indicators in Hesse. The share to that funds are allocated according to these indicators is relatively low in most states, however. Over time, the federal states revised the IBF models also regarding this share of indicator-based allocated funds, which mostly tends to increase. The different types of funding models all set financial incentives which are supposed to influence the universities' performance and to steer their behavior.

In general, it is well known from personnel economics, that the behavior of economic agents is strongly affected by incentives. In particular, incentives are expected to induce positive effects since they can, for instance, work as productivity-enhancing mechanisms. However, it can result in a balancing act to set incentives "right", i.e., to direct them such that they contribute to achieving the intended reactions. Moreover, agents exhibit heterogeneous reactions that partly might even have counter-productive effects.

In the following, the incentive mechanisms in the case of the higher education funding reform in Germany are discussed. Here, the economic agents are represented by the universities. The incentives provoked by the reform vary across federal states depending on several factors. First, the incentives are influenced by the composition of the indicators in the IBF model, since the indicators might address diverging objectives. For example, the IBF model may include a set of indicators that rewards the university for investing more time into teaching and into research at the same time which results in conflicting incentives. Second, the strength of an incentive set by an individual indicator or a group of indicators generally depends on the overall share of the indicator-based funds of total public funds. In general, the higher the IBF share of public funds, the stronger is the incentive (ceteris paribus). Additionally, the strength of the incentive is directed through the individual intensity, i.e. the weight, of each indicator. Furthermore, how a single university decides to follow which indicators, will in particular depend on its individual strategic policy or its relative strengths compared to other universities. That is, each university will decide, given the size of the incentives, whether and until which point it is worth to follow a certain indicator.

The decision of a university to follow more than one indicator can cause trade-offs, i.e. the universities might concentrate on one indicator at the cost of another. That is, incentives ad-

Table 2: **IBF Indicators - Relative weights (Average over all years)**

State	Univ. type	Start IBF	Share IBF	Stud.	Grad.	PhD grad.	TPF	Staff	Succ. rate	L. of study	Gend. equ.	Oth.[1]
BV	U	1999	1.9	0.3	0.3	0.1	0.4	0.4	0.1	-	0.1	0.1
BV	TC	1999	0.4	0.1	0.1	-	-	0.1	-	-	-	-
BW	U	2000	18.0	2.4	2.7	0.9	5.6	-	-	-	2.6	3.8
BW	TC	2000	19.0	4.3	4.3	-	2.0	-	-	-	2.1	6.4
BW	CE	2000	23.0	4.7	4.7	0.4	3.2	-	-	-	2.6	7.4
LS	U	2006	7.4	-	2.6	0.9	2.6	-	-	-	0.4	0.9
LS	TC	2000	19.1	3.2	5.0	-	0.3	-	-	-	0.3	10.2
B	U	2002	18.3	-	-	1.7	5.8	-	4.5	2.7	0.9	2.6
B	TC	2002	20.8	-	-	-	1.9	-	8.3	5.0	1.0	4.6
B	CAM	2005	24.0	-	-	-	1.8	-	9.6	5.8	1.2	5.6
MW	U	2002	3.8	0.5	0.7	0.2	1.6	0.2	-	-	0.3	0.3
MW	TC	2002	3.8	1.0	1.3	-	0.3	0.2	-	-	0.5	0.5
BR	U	2003	7.0	-	2.3	1.2	2.3	-	-	0.6	-	0.6
BR	Arts	2003	7.0	-	1.8	-	1.8	-	-	1.8	-	1.8
BR	TC	2003	7.0	-	1.6	-	1.2	-	-	2.0	-	2.3
HS	U	2003	100.0	18.4	0.2	1.4	-	-	-	0.2	18.4	61.5
HS	TC	2003	100.0	3.8	0.3	1.6	-	-	-	0.2	21.7	72.5
HS	CAM	2003	100.0	5.2	0.3	1.6	-	-	-	0.2	21.3	71.4
TH	U	2003	14.5	3.1	2.8	1.7	4.1	0.5	-	-	0.7	1.5
TH	TC	2003	14.4	4.8	3.7	-	3.1	0.5	-	-	0.7	1.5
TH	CAM	2004	16.7	4.5	3.6	2.0	3.5	0.6	-	-	0.9	1.7
BB	U	2004	94.9	56.2	8.0	2.0	6.0	18.7	-	-	2.0	2.0
BB	TC	2004	95.0	63.8	8.0	2.0	6.0	11.3	-	-	2.0	2.0
SA	U	2005	1.0	-	-	0.1	0.2	-	0.1	0.2	0.2	0.2
SA	TC	2005	0.8	-	-	0.1	0.1	-	0.1	0.3	-	0.2
SA	CAM	2005	0.8	-	-	-	-	-	0.1	0.3	-	0.4
SH	U	2006	5.0	-	-	0.5	2.0	-	2.0	-	0.5	-
SH	TC	2006	.0	-	-	-	2.0	-	2.0	0.5	0.5	-
SL	U	2008	10.0	-	-	-	4.0	-	-	-	-	6.0
SL	TC	2008	10.0	-	-	-	4.0	-	-	-	-	6.0

Note: [1]Others include e.g. patents, prices, publications, further education.
In some federal states, the set of indicators changed over time: Graduates at Bavarian universities were introduced in 2001, TPF at TCs in Baden-Wurttemberg in 2005, TPF at TCs in Lower Saxony in 2006, PhD graduates at universities in Mecklenburg-Western Pomerania in 2005, PhD graduates and TPF in Saxony in 2006; Students at TCs in Lower Saxony in 2006, PhD graduates and TPF were abolished in Saxony in 2008. IBF = Indicator-based funding. Stud. = Students, Grad. = Graduates, TPF = Third-party funds. Succ. rate = Success rate, L. of study = Length of study, Gend. equ. = Gender equality. U = University, TC = Technical college, CE = College of Education, CAM = College of Arts and Music. State abbreviations see Table 1.

dressing a certain outcome can imply an indirect, negative incentive regarding another outcome. In particular, it can be difficult to follow the two main objectives of teaching and research, but as well to improve different indicators within one dimension. The following section describes the mechanisms that might be at work with regard to the outcomes considered in this study (assuming that capacities of scientific personnel are relatively fixed in the short and medium run).

First, to increase student numbers, a university can improve the quality of studies and/or expand the number of university places and the variety in fields of study. The former requires personnel capacities resulting in a trade-off between teaching and research. If the scientific personnel spends more time on teaching improvement, less time will be available for research projects, be it third-party fund acquisition or dissertation projects. The latter requires financial resources, i.e. it constitutes an investment which is profitable if the expected additional funds that will be received for additional students exceed the respecting costs for the expansion. Second, if a university aims at increasing the number of graduates, it can e.g. enable more students to pass exams, or enable them to complete their studies more efficiently. The probability to pass an exam can be increased either if better grades are awarded by the university or if students are better prepared. Grade inflation is an example for a negative incentive that can result from an original incentive of the funding model. Better exam preparations in the form of more or higher-quality tutorials, or better course guidance in general, again require personnel capacities which can lead to a trade-off as described above.

Third, if a university decides to raise the number of PhD graduates, it can achieve this goal if doctoral students are charged less obligations in administrative tasks, in teaching, or in projects their dissertation does not benefit from. This can again induce trade-offs, either between teaching and research (if less time is dedicated to teaching), or as well within research objectives (if less time is dedicated to third-party fund projects and fundraising). Further, more dissertations will be completed, if the (time-consuming) supervision by professors is improved, or if requirements are reduced. This implies again trade-offs or negative incentives, respectively. Fourth, the amount of third-party funds can be increased if more fundraising proposals are written or if their quality is improved. Both results in a higher work load for scientific personnel, which can only be managed if time for other activities like teaching, administrative tasks, or research is cut down.

The above discussion makes clear which mechanisms can provoke trade-offs between teaching and research, or within the research dimension. This mainly occurs because the available working time of the scientific personnel is limited. Further, the incentives provided in the funding models can induce negative incentives on other indicators which can result in an unintended deterioration of unadressed or (subjectively perceived) less important indicators.

The fact that the expected reaction of the universities is not clear a priori emphasizes the need of empirical insights into this question. In the following, we will shed some light on first, the reform effect on the outcome indicators, second, whether different funding intensity matters,

and third, we examine differences between university types.

3 Data and Empirical Strategy

The empirical analysis is based on the *Statistic of Students and Examinations*, which is administrative data provided half-yearly by the German Federal Statistical Office ("Statistisches Bundesamt"). The data of this full census of students enrolled at German universities are reported by the university administrations and examination offices. We aggregate the individual data at the university level to extract information on e.g. the number of students, first-year students, graduates, and PhD graduates per university. We extend the set of variables by merging information on the university level from other sources, that is information on third-party funds (source: Federal Statistical Office, 2011a), on universities' founding years, and on university types (source: German Rectors' Conference, 2012). On the federal state level, we add the age-specific share of high school graduates with university entrance qualification ("allgemeine bzw. Fach-Hochschulreife") (source: Federal Statistical Office, 2011b). Since the additional data are not available per semester, we aggregate the whole dataset on a yearly basis, i.e., our observation units in the regression analysis are universities per year. The time period covered by the data ranges from 1997 to 2008.

The sample that we analyze is subject to the following restrictions. First of all, we exclude medicine faculties because different types of funding models were introduced for them which were mostly independently developed from the IBF models analyzed here. This is because the financial conditions in medicine-related fields generally differ strongly from those in other fields. Next, we drop the federal states of North-Rhine Westphalia and Rhineland-Palatinate due to their treatment prior to the first year of observation in the data. The city state of Hamburg (6 universities) is excluded due to incomplete data since the universities did not report them to the statistical office. Further, we drop university types that are not affected by the funding reform, which are colleges of public administration ("Verwaltungsfachhochschulen") and private universities. Under these restrictions, 173 universities in 13 federal states are available for the empirical analysis.

The aim of our investigation is twofold. First, we are interested in identifying the overall average effect of the introduction of the indicator-based funding on selected outcomes. Second, we examine heterogeneous effects for different funding intensities employed in the federal states and heterogeneous effects by university types. We implement the former by splitting the sample into universities/states with low and high funding incentives respecting each indicator. Low and high funding incentives are defined by the weight of each specific indicator in the funding model of the respective state. As thresholds to define "high" and "low", we use the 50th percentile of each indicator by university types. For example, when analyzing the treatment effect on the number of students, the respective treatment intensity is defined by the weight that is attributed to this indicator in the funding model. A university then belongs to the low intensity (sub-)sample, if it is among those universities with a student indicator weight below

the median. Accordingly, the high intensity (sub-)sample consists of those universities with a student indicator weight equal to or higher than the median.

To identify the causal effects of the funding reform on the outcome variables, we rely on a difference-in-differences (DiD) approach, which exploits the fact that the IBF scheme was not introduced simultaneously in all federal states.[6] The idea of the DiD strategy is to compare the development of an outcome variable over time between a treatment group and a well-defined control group. This comparison can be used to remove any bias due to changes over time that are common to both groups. In our case, the treatment and control group are defined by the time of the reform's implementation in each federal states. I.e., the observations after the introduction in the respective state belong to the treatment group, while observations before the introduction accordingly serve as a control group. For example, in 2000, Bavaria, Baden-Wurttemberg and TCs in Lower Saxony form the treatment group while all other states and other university types in Lower Saxony are in the control group. One state, Saxony-Anhalt, is part of the control group during the whole observation period since the reform in that state took place after the time period we observe in the data (compare Table 1). In order to account for unobserved factors that might have influenced the treatment and the outcomes, such as a university's negotiating power, we include fixed effects on the university level in our estimation framework.

The crucial identification assumption of a classical DiD, the so-called "common trends assumption", requires that in absence of the treatment, i.e. without the funding reform, the difference in the outcomes between treatment and control group (year-state-observations) would have been constant over time. This means that there are no time-varying factors influencing the outcome and affecting treatment and control observations differently. That implies that there were no other policy changes or regional economic shocks that coincide with the timing of the reform and that affect the two groups differently.

In fact, there were higher education policies that were implemented during the time period under investigation. We argue, however, that they are unproblematic in our framework. They either do not affect the relative development of the outcomes or we are able to control for their potentially distortionary effect. First, the Higher Education Pact 2020 ("Hochschulpakt 2020"), that started in 2007 and is aiming at increasing the number of university places, might, if at all, affect the estimated treatment effect on the number of students. Robustness checks do not provide evidence on that, however (compare section 4). Second, in order to rule out potential problems arising from the Excellence Initiative ("Exzellenzinitiative") or from the introduction of tuition fees in some federal states, we include dummy variables for being a winner university of the Excellence Initiative and for the presence of tuition fees.[7] Further, the advancement of the Bologna reform is taken into account by means of including university-specific shares of first-year students starting a Bachelor or Master degree rather than a former "Diplom" or the

[6] A general description of the DiD strategy can e.g. be found in Ashenfelter and Card (1985).

[7] The Excellence Initiative of the German Federal Ministry of Education and Research and the German Research Foundation is a funding program that promotes and rewards outstanding research. It consists of three lines of funding that include overall future concepts, so-called excellence clusters and graduate schools.

like. We control for that because the number of students or graduates might be affected by the Bologna process since former degrees were split into two consecutive degrees, the Bachelor and the Master degree.

When analyzing the effect of the funding reform, we take account of the multi-dimensionality of university's performance by focusing on the two main objectives of universities, i.e. research and teaching activity, as described in section 2. We model this by implementing a regression framework containing four equations, one for each of the above described outcome variables that reflect dimensions of research and teaching. PhD graduates are analyzed only for universities ("Universitäten") and TUs since the other four types, i.e. TCs, educational colleges and colleges of arts and music, are either not entitled to award PhD degrees or do only have very few PhD graduates.

Formally, the DiD approach is implemented by estimating four single regression equations (for $k = 1, \ldots, 4$) separately by using pooled OLS:

$$log(Y_{ijt}^k) = \beta_0^k + X_{ijt}^{k\prime}\beta^k + \delta^k Treat_{jt}^k + \gamma^k ibf_{jt}^k + \lambda_t^k + \alpha_i^k + \varepsilon_{ijt}^k \quad (1)$$

where $\log(\mathbf{Y}_{ijt}^k)$ is the logarithm of the respective outcome variable and $\mathbf{X}_{ijt}$ is a vector of covariates for university i in state j at time t that includes information on the university and on the state level. In particular, $\mathbf{X}_{ijt}$ comprises the age-specific share of high school graduates with university entrance qualification within a state, the share of students in natural sciences or engineering, the share of first-year students in a "new degree" (Bachelor or Master), and dummy variables indicating the state-level introduction of tuition fees and being a winner university of the Excellence Initiative, respectively.[8] The share of students in natural sciences or engineering is supposed to account for fundamental differences in the acquisition of third-party funds or in the share of PhD students in those fields of study. The age-specific share of high school graduates with university entrance qualification within a state controls for differences in the potential number of first-year students across states, which might affect student and graduate numbers. The set of control variables $\mathbf{X}_{ijt}$ varies between the equations, depending on whether the outcome corresponds to teaching or to research.[9] λ_t is a vector of year fixed effects ($t = 1997, \ldots, 2008$) and α_i is a vector containing university fixed effects, while ε_{ijt} is an idiosyncratic error term. In order to avoid problems arising from a correlation of residuals over time and/or across states, standard errors are clustered at the state-year level in all estimations.

$Treat_{jt}$ is the treatment variable, which equals one if the reform is introduced in state j at time t, and zero otherwise. The corresponding coefficient, δ^k, represents the coefficient of interest measuring the effect of the treatment, i.e. of the funding reform, on the respective outcome. ibf_{jt} represents the overall share of indicator-based funding of total funds. We include this state-specific level of performance funding in order to account for the large variation in the intensity of indicator-based funding across states (compare Table 1).

[8] See Table A.1 in the Appendix for a definition of the variables.

[9] The age-specific share of high school graduates with university entrance qualification within a state and the share of first-year students in a "new degree" are only included in the two teaching equations while the share of students in natural sciences or engineering is only included in the two research equations.

Not only the time of introduction and the set of indicators used to allocate the public funds differs between the state-specific models, but also the intensity by that each indicator is involved differs largely (compare Section 2). Therefore, in a second step, we examine the heterogeneous effects for different intensities of the indicators in each state, i.e., we analyze the treatment effects for states with high and low indicator weights separately.

Since the different types of universities also differ by the importance they ascribe to teaching and research activities, we further allow for heterogenous effects by university type. This is implemented by an interaction of the treatment variable with a dummy variable indicating the university type (TC or university). Lastly, we allow for a time lag in the treatment effect by splitting the treatment effect into three categories (indicated as $\delta_{<2yrs.}$, $\delta_{2-3yrs.}$, and $\delta_{>3yrs.}$ in result tables). By doing so, the effect in the year of the reform introduction and one year after, the effect 2-3 years after introduction, and the effect more than 3 years after introduction are measured separately.

Descriptive statistics of the change in the four outcome variables between the pre- and post-reform periods are presented in Table 3. Although universities of arts or music and colleges of education are in our estimation sample as well, here we concentrate on the prevalent university types in Germany, which are universities and TCs (26 % and 50 % of all universities in 2011, amounting 65 % and 31 % of the total number of students, respectively). Table 3 reveals that, except for the number of PhD graduates, there is an overall upward trend in the outcomes over time in most federal states when comparing pre- with post-reform numbers. Significant increases in the number of students, graduates, and third-party funds can be observed more often at TCs than at universities. Whether these positive development in the outcomes can be ascribed to a causal reform effect, will be uncovered by the regression analysis in the following section.

4 Results

This section presents the results from DiD estimations of the effect of the funding reform on four outcomes, namely the (log of the) number of students, graduates, and PhD graduates, as well as the (log of the) amount of third-party funds. Unconditional and conditional treatment effects are shown in Table 4 (on the left- and right-hand side, respectively).[10] Unconditionally, the effect of the reform on all four outcomes is positive and highly significant. We find the reaction to be strongest for third-party funds and less pronounced for PhD graduates. In the lower part, we split up the effect according to different post treatment periods, i.e., we allow the effect to differ over time after the treatment. The outcomes we consider can naturally only adjust after some time lag since the attraction of students, the graduation of (PhD) students and the acquisition of third-party funds certainly takes some time. In line with this argumentation, the results suggest that the unconditional effects are increasing with more years in treatment

[10] The full results can be found in the Appendix Table A.3.

Table 3: **Comparison of means: Pre- and post-reform values of outcome variables**

	Universities			TCs		
State	**Pre**	**Post**	**Δ**	**Pre**	**Post**	**Δ**
Students						
BV	16,741	15,203	-1,537	3,422	4,033	611
BW	13,912	14,778	866	2,756	3,544	788***
LS	13,312	13,618	306	4,601	6,925	2,323***
B	35,698	35,307	-391	5,022	6,397	1,375*
MW	7,912	11,124	3,212***	2,450	3,121	671**
BR	18,852	22,513	3,662***	4,155	5,341	1,186
HS	21,507	22,206	699	6,706	8,776	2,071***
TH	9,351	12,041	2,691	3,069	4,172	1,103***
BB	7,581	10,289	2,708*	1,938	2,711	773***
SA	11,521	13,440	1,919	4,193	5,198	1,005***
SH	7,699	8,783	1,085	2,863	3,452	589
SL	15,629	14,596	-1,033	3,262	4,112	850
Graduates						
BV	1,592	1,713	121	675	585	-90
BW	1,477	1,501	23	423	592	169***1
LS	1,295	1,940	646***	667	1,148	481***
B	2,574	3,301	727***	743	1,130	387***
MW	590	1,010	420***	258	478	219***
BR	1,473	1,633	160	540	613	73
HS	1,897	2,128	231*	769	1,091	322***
TH	710	1,150	440***	347	581	234***
BB	589	976	387***	229	400	172***
SA	962	1,350	389	579	852	273***
SH	723	889	166	314	433	119*
SL	1,432	1,637	205	450	737	287
PhD graduates						
BV	331	359	28	-	-	--
BW	457	412	-44	-	-	--
LS	183	205	22	-	-	--
B	841	610	-231***	-	-	--
MW	180	181	1	-	-	--
BR	265	263	-2	-	-	--
HS	413	391	-22	-	-	--
TH	183	230	47	-	-	--
BB	59	102	44***	-	-	--
SA	416	287	-129**	-	-	--
SH	211	223	12	-	-	--
SL	297	286	-11	-	-	--
Third party funds						
BV	24,707	25,137	430	686	1,257	571***
BW	34,768	45,640	10,872*	758	1,511	754***
LS	18,688	28,078	9,390**	4,824	10,293	5,470***8
B	36,549	56,039	19,490***	998	2,148	1,150***
MW	9,759	18,240	8,481***	648	1,438	790***
BR	52,753	70,968	18,216***	2,517	2,772	254
HS	24,584	36,849	12,264***	1,623	2,940	1,316***
TH	12,856	21,592	8,736**	551	1,325	774***
BB	10,498	15,326	4,827**	1,356	2,653	1,297***
SA	11,906	16,682	4,777	2,935	4,375	1,441***
SH	11,919	15,325	3,406	2,021	2,062	41
SL	23,811	31,186	7,374	677	1,447	770

Notes: Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ (based on t-test for comparison of means).

(lower part of Table 4). The corresponding conditional effects presented on the right-hand side of the Table are substantially lower. Hence, the control variables and year effects explain a considerable part of the observed effects. Still, the effect on university graduates remains positive significant and increases over time. Further, after some years in treatment, the number of PhD graduates decreases significantly (see lower right part of Table 4). From these results, the teaching dimension seems to react positively and the research dimension negatively to the funding reform.

Table 4: **Effects of indicator-based funding - DiD estimations, full sample**

	Unconditional effects				**Conditional effects**			
	Stud.	**Grad.**	**PhD gr.**	**TPF**	**Stud.**	**Grad.**	**PhD gr.**	**TPF**
δ	0.208***	0.399***	0.156***	0.674***	-0.018	0.112*	-0.060	0.004
	(0.019)	(0.051)	(0.042)	(0.046)	(0.023)	(0.059)	(0.060)	(0.068)
Controls	No	No	No	No	Yes	Yes	Yes	Yes
Year effects	No	No	No	No	Yes	Yes	Yes	Yes
R^2	0.165	0.137	0.032	0.123	0.369	0.269	0.098	0.259
$\delta_{<2yrs.}$	0.123***	0.234***	0.120**	0.465***	-0.016	0.108**	-0.100	0.036
	(0.025)	(0.045)	(0.050)	(0.050)	(0.022)	(0.054)	(0.063)	(0.069)
$\delta_{2-3yrs.}$	0.195***	0.360***	0.198***	0.636***	-0.019	0.129*	-0.095	-0.077
	(0.021)	(0.052)	(0.053)	(0.057)	(0.026)	(0.073)	(0.068)	(0.081)
$\delta_{>3yrs.}$	0.307***	0.603***	0.159***	0.920***	0.019	0.190*	-0.312***	-0.126
	(0.022)	(0.066)	(0.049)	(0.057)	(0.034)	(0.098)	(0.093)	(0.097)
Controls	No	No	No	No	Yes	Yes	Yes	Yes
Year effects	No	No	No	No	Yes	Yes	Yes	Yes
R^2	0.228	0.196	0.036	0.150	0.372	0.270	0.120	0.261
Obs.	1,942	1,942	668	2,026	1,942	1,942	668	2,026

Notes: Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01. Standard errors in parentheses.
The included control variables are contained in the full Table A.3 in the Appendix.
Stud. = Students, Grad. = graduates, PhD grad. = PhD graduates, TPF = third-party funds.

In a next step, we split the sample by the intensity of the reform in terms of the funding intensity concerning each indicator. That is, we split the sample by the median weight of each outcome indicator in the funding model.[11] Table 5 displays the corresponding estimation results, where the left part of the Table comprises observations that exhibit a low indicator intensity and the right part comprises those with a high indicator intensity.[12] The previously observed positive effect of the reform on the number of graduates turns out to be driven by universities in states that set a relatively low weight on this indicator (see upper left part of Table 5). As the lower left part of the Table shows, the effect is again growing over time.[13]

For the other outcome indicators in the low intensity sample, significant reform effects only become visible when allowing for heterogenous effects over time. In particular, a positive effect on students becomes most pronounced after a four-year time lag, while, for third-party funds, a strong negative effect becomes apparent after four and more years after the reform's introduction. Finally, the coefficient for the number of PhD graduates is significantly positive affected by the reform two or three years after introduction if the incentive is only weak. Taken together, with low indicator incentives, teaching-related outcomes react positively to the reform

[11] We also estimated the corresponding fully interacted model, which we do not present here. However, our discussion of the significance of differences between coefficients from the low and high intensity sample rely on these results.

[12] The full results can be found in the Appendix Table A.4.

[13] Appendix Table A.5 displays the number of universities that are constituting the treatment group in each subsample when splitting up the effects according to time in treatment.

while within the research dimension, third-party funds react negatively.

For the high intensity sample, a different picture emerges (see right part of Table 5). In contrast to the effect for universities with a low incentive on graduates, we now do not find a significant reform effect on graduate numbers at universities with a high incentive on that indicator. The corresponding coefficients are significantly different between the high and low intensity sample. Further, while the effect on third-party funds is significantly positive now, it is significantly negative on student numbers. Splitting the reform effect according to time in treatment shows that both effects increase over time (lower right part of Table 5). Moreover, a negative effect on PhD graduates becomes visible.

Hence, the distinction between high and low funding incentives reveals opposing effects between the two subsamples regarding third-party funds and student numbers. Recalling the results from the pooled sample (Table 4), it can be seen that these opposing effects had neutralized each other resulting in before insignificant effects on the two outcomes. Further, the distinction shows that enhancing efforts in third-party fundraising seem to require a higher incentive, while at the same time, student numbers and the number of PhD graduates are neglected when the respective incentives are high.

Thus, with high indicator incentives, there are effects working into opposite directions. First, in the teaching dimension, student numbers drop, while in the research dimension, third-party funds increase. Second, within the research dimension, the number of PhD graduates decreases in turn. This hints at the hypothesis, that not all outcomes can be increased at the same time and certain outcomes can constrain each other.

Table 5: **Effects of indicator-based funding - DiD estimations, by intensity**

	Low indicator share				**High indicator share**			
	Stud.	**Grad.**	**PhD gr.**	**TPF**	**Stud.**	**Grad.**	**PhD gr.**	**TPF**
δ	0.031	0.233***	0.146	0.044	-0.167***	-0.047	-0.151*	0.119*
	(0.025)	(0.077)	(0.095)	(0.136)	(0.037)	(0.050)	(0.089)	(0.061)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.408	0.229	0.162	0.225	0.385	0.620	0.138	0.410
$\delta_{<2yrs.}$	0.040*	0.215***	0.065	0.205	-0.155***	-0.037	-0.184**	0.181***
	(0.021)	(0.065)	(0.082)	(0.135)	(0.039)	(0.047)	(0.076)	(0.052)
$\delta_{2-3yrs.}$	0.046*	0.274***	0.262**	-0.021	-0.169***	-0.053	-0.325***	0.271***
	(0.024)	(0.080)	(0.127)	(0.137)	(0.043)	(0.068)	(0.081)	(0.068)
$\delta_{>3yrs.}$	0.150***	0.389***	0.077	-0.367**	-0.213***	-0.017	-0.663***	0.563***
	(0.033)	(0.111)	(0.124)	(0.148)	(0.046)	(0.083)	(0.115)	(0.094)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.424	0.235	0.180	0.240	0.386	0.620	0.191	0.432
Obs.	1,295	1,280	285	945	647	662	383	1,081

Notes: Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.
The included control variables are contained the full Table A.4 in the Appendix.
Stud. = Students, Grad. = graduates, PhD grad. = PhD graduates, TPF = third-party funds.

Finally, we replicate Table 5 by explicitly allowing for differences between university types, focusing on differences between universities and TCs since they constitute the majority of all

German universities and are accordingly the most relevant ones.[14] The results presented in Table 6 show that there are no significant differences between the two groups in the low intensity sample (left part of the Table), but only in the sample with a high indicator intensity (right part of the Table).[15,16]

The estimations reveal, that the negative effect on student numbers within the high intensity sample is entirely driven by TCs, while it is insignificant for universities. The before insignificant effect on graduate numbers (see Table 5) now splits up into a significant negative effect for TCs and an insignificant effect for universities. Respecting third-party funds, it can be observed that the increase of this outcome is also driven by TCs and turns out to be insignificant for universities.

These findings show that, if the incentive on the respecting indicators is high, TCs mainly respond to the funding reform if compared to universities. Interestingly, it is observed that they decrease student and graduate numbers while they focus on third-party fund acquisition. Further, both university types react by increasing the number of graduates even if the funding share is relatively low.

Table 6: **Effects of indicator-based funding - DiD estimations, by intensity: Differences between university types**

	Low indicator share			**High indicator share**		
	Students	**Grad.**	**TPF**	**Students**	**Grad.**	**TPF**
δ	0.011	0.360***	-0.099	0.008	-0.049	0.033
	(0.044)	(0.086)	(0.072)	(0.060)	(0.053)	(0.026)
δ*Type: TC	0.002	-0.071	0.006	-0.399***	-0.181*	0.110
	(0.070)	(0.178)	(0.128)	(0.100)	(0.097)	(0.098)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.484	0.296	0.271	0.462	0.682	0.454
$\delta_{<2yrs.}$	0.028	0.382***	-0.083	0.032	-0.029	0.051*
	(0.044)	(0.087)	(0.073)	(0.063)	(0.052)	(0.029)
$\delta_{2-3yrs.}$	-0.030	0.299***	-0.113	-0.016	-0.086	0.005
	(0.055)	(0.098)	(0.084)	(0.064)	(0.063)	(0.033)
$\delta_{>3yrs.}$	-0.017	0.258**	-0.212**	-0.031	-0.077	0.045
	(0.071)	(0.107)	(0.087)	(0.074)	(0.077)	(0.052)
$\delta_{<2yrs.}$*Type: TC	-0.010	-0.097	0.062	-0.413***	-0.195**	0.159*
	(0.066)	(0.158)	(0.129)	(0.102)	(0.091)	(0.090)
$\delta_{2-3yrs.}$*Type: TC	0.086	0.118	-0.113	-0.401***	-0.189	0.320**
	(0.080)	(0.193)	(0.136)	(0.100)	(0.123)	(0.133)
$\delta_{>3yrs.}$*Type: TC	0.155	0.418	0.136	-0.380***	-0.210	0.609***
	(0.117)	(0.270)	(0.163)	(0.120)	(0.134)	(0.183)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.492	0.309	0.289	0.463	0.683	0.483
Obs.	1,295	1,280	945	647	662	1,081

Notes: Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. The included control variables are contained in the full Table A.6 in the Appendix. Grad. = graduates, TPF = third-party funds.

Concluding from our findings, it seems important to take the funding intensity of the indica-

[14] PhD graduates are not contained in this analysis since TCs are not entitled to award PhD degrees.

[15] The full results can be found in the Appendix Table A.6.

[16] The estimations were additionally carried out separately by indicator intensity and by university type instead of this interacted version. Statements regarding coefficients for TCs (whose significance cannot be derived from Table 6) are based on those results, which are not presented here, however.

tors as well as differences between university types into account. This can, for instance, clearly be seen in the example of graduate numbers. The overall treatment effect shown in Table 4 is significantly positive, which only persists for low funding intensities (see Table 5). Analyzing differences between university types reveals, that the treatment effect is even negative for TCs if the financial incentive for this outcome is high. Further, an increase and decrease in third-party funds and in student numbers, respectively, were only visible after splitting the sample by indicator intensities. These effects turned out to be driven by TCs, if the incentive is high enough.

An intended positive response of the reform by setting high funding weights only seems to be realized in research activity as third-party funds are the only indicator reacting positively to higher incentives within the time period analyzed. There is evidence that this effect is significantly stronger for TCs than for universities. At the same time, however, a cut-back of teaching efforts is observed. In the case of graduates, low and high indicator weights induce opposite effects for TCs. That is, already a low weight leads to an increase in graduate numbers, while a high weight even causes a decrease. Thus, we find evidence that for addressing the teaching dimension, low indicator weights have clear effects, while for addressing the research dimension, high incentives show more pronounced effects. However, the reactions partly work into an unintended direction, i.e., some of the addressed outcomes deteriorate after the reform introduction, which could hint at trade-off reactions between the target dimensions.

In order to support the robustness of our results, we conducted some sensitivity checks. First, dropping the observations for Bremen in 2008, when the reform was abolished again, does not affect the results. Further, we do not consider the Ashenfelter's dip problem (Ashenfelter, 1978) likely to distort our results since the outcomes analyzed are not possible to be influenced within a short time period like between the announcement and introduction of the reform. It can further be seen in the reported tables that almost all observed effects are visible not earlier than two years after treatment. Moreover, it can be shown that our results do not depend on the pattern of control group observations that results from the different timing of the introduction of the reform. Re-running the estimations cutting the sample in 2004 or 2005, i.e. dropping the last three or four years from the regression sample, leaves our results mainly unchanged.

5 Discussion and Conclusion

This paper examines the effects of a funding reform in the German higher education system on teaching and research output. The reform has started in the 1990's and aims at improving the efficiency and performance of universities by setting incentives through a performance-related allocation of public funds. The federal states of Germany, which are in charge of the education policy, introduced such funding models at different points in time and designed them individually. In the scope of the funding reform, different dimensions of universities' performance are measured by means of defined indicators. Teaching performance is e.g. quantified by the number of students and graduates, amongst others. Indicators such as the number of PhD

graduates and the amount of third-party funds are supposed to measure research output. The relative weights of these indicators as well as the overall share of indicator-based funding differs largely across states.

Exploiting the variation in the timing of the state-level introduction of the reform, we apply a difference-in-differences strategy to investigate the effects of the reform on the above four outcomes. In a second step, we differentiate by the intensity to which these indicators enter the funding models. We further explore differences between university types, focusing on universities and technical colleges (TCs). Allowing for these heterogeneous effects yields important insights into the reactions to different forms of funding models.

To affect the research dimension in a positive way, i.e. to reach an increase in the amount of third-party funds, incentives need to be stronger, while the teaching dimension responds positively to weaker incentives. An explanation could be that universities regard it worthwhile to spend more ressources on third-party fund acquisition. Hence, universities only increase efforts in writing proposals if the amount of public funds they receive is high enough, i.e. the indicator share in the funding model is relatively high. By doing so, they might have to cut back their efforts in improving the teaching dimension because of limited ressources.

While the latter effect is observed at universities and technical colleges alike, respecting high indicator incentives, differences between the university types appear. More precisely, for TCs, unintended effects are observed as student and graduate numbers decrease when the respective indicator weight is high. Such unintended effects can be induced by "wrong" incentives. One explanation for that phenomenon is, that the existing funding models are build on a purely quantitative basis without taking any qualitative aspects into account. Hence, it is possible to increase the amount of a certain indicator at the cost of its quality or the quality of another indicator. One solution could be to incorporate criteria into the funding models that ensure a certain quality level.

To sum up, our results underline that it is crucial to design indicator-based funding models carefully by providing incentives such that the intended reactions are achieved best possible. Thereby, differences with regard to the target dimensions, such as research and teaching, should be taken into account. Our findings additionally provide evidence that different university types deserve careful attention in the design of such funding models.

References

Ashenfelter, O. (1978). Estimating the Effects of Training Programs on Earnings. *The Review of Economics and Statistics 6*(1), 47–57.

Ashenfelter, O. and D. Card (1985). Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs. *The Review of Economics and Statistics 67*(4), 648–660.

Atkinson, A., S. Burgess, B. Croxson, P. Gregg, C. Propper, H. Slater, and D. Wilson (2009). Evaluating the Impact of Performance-related Pay for Teachers in England. *Labour Economics 16*(3), 251–261.

Bauer, T. K. and B. S. Grave (2011). Performance-related Funding of Universities - Does more Competition Lead to Grade Inflation? Ruhr Economic Papers No. 288, RWI, RUB.

Burgess, S., C. Propper, H. Slater, and D. Wilson (2005). Who Wins and Who Loses From School Accountability? The Distribution of Educational Gain in English Secondary Schools. The Centre for Market and Public Organisation 05/128, Department of Economics, University of Bristol, UK.

de Boer, H. F., J. Enders, and U. Schimank (2007). On the Way Towards New Public Management? The Governance of University Systems in England, the Netherlands, Austria and Germany. In D. Jansen (Ed.), *New Forms of Governance in Research Organizations - Disciplinary Approaches, Interfaces and Integration*, pp. 135–152. Dordrecht: Springer.

Eurydice European Unit (2008). Higher Education Governance in Europe. Policies, Structures, Funding and Academic staff. Education and culture dg, European Commission.

Federal Statistical Office (2011a). Statistisches Bundesamt: Bildung und Kultur. Monetäre hochschulstatistische Kennzahlen. Fachserie 11 Reihe 4.3.2, lfd. Jahrgänge, Wiesbaden.

Federal Statistical Office (2011b). Statistisches Bundesamt: Hochschulstandort Deutschland 2009. Ergebnisse aus der Hochschulstatistik. Wiesbaden, Gruppe VI B, "Bildung, Forschung und Entwicklung, Kultur".

German Rectors' Conference (2012). Hochschulrektorenkonferenz: Hochschulkompass. http://www.hochschulkompass.de/hochschulen/download.html.

Jacob, B. A. (2005). Accountability, Incentives and Behavior: The Impact of High-stakes Testing in the Chicago Public Schools. *Journal of Public Economics 89*(5-6), 761–796.

Jacob, B. A. and S. D. Levitt (2003a). Catching Cheating Teachers: The Results of an Unusual Experiment in Implementing Theory. In W. G. Gale and J. R. Pack (Eds.), *Brookings-Wharton Papers on Urban Affairs*, pp. 185–209. Washington, DC: Brookings Institution Press.

Jacob, B. A. and S. D. Levitt (2003b). Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating. *Quarterly Journal of Economics 118*(3), 843–877.

Kingdon, G. G. and F. Teal (2007). Does Performance Related Pay for Teachers Improve Student Performance? Some Evidence from India. *Economics of Education Review 26*(4), 473–486.

Lavy, V. (2009). Performance Pay and Teachers' Effort, Productivity and Grading Ethics. *American Economic Review 99*(5), 1979–2011.

Martins, P. (2010). Individual Teacher Incentives, Student Achievement and Grade Inflation. CEE Discussion Papers 0112, Centre for the Economics of Education, LSE.

Reback, R. (2008). Teaching to the Rating: School Accountability and the Distribution of Student Achievement. *Journal of Public Economics 92*(5-6), 1394–1415.

A Appendix - Tables

Table A.1: **Definition of Variables**

Variable	Definition
a. Dependent variables	
Log(Students)	Logarithm of the number of students currently studying (without medicine students)
Log(Graduates)	Logarithm of the number of graduates without PhD graduates (without medicine graduates)
Log(PhD graduates)	Logarithm of the number of PhD graduates (without medicine PhD graduates)
Log(Third-party funds)	Logarithm of third-party funds in current year in Euro
b. Explanatory variables	
δ^k	0/1-variable; 1 if the reform is introduced in state j at time t
$\delta^k_{<2yrs.}$	0/1-variable; compare δ^k, 1 in year of introduction and one year after reform in state j
$\delta^k_{2-3yrs.}$	0/1-variable; compare δ^k, 1 in year two and three after reform in state j
$\delta^k_{>3yrs.}$	0/1-variable; compare δ^k, 1 in year four and more after reform in state j
Share of ibf funding	Share of indicator-based funds of a university's total funds
Age-spec. pop.	Age-specific share of high school graduates with university entrance qualification
Share of natural sc./engin.	Share of students in natural sciences or engineering
New degree	Share of first-year students in a Bachelor or Master degree
Tuition fees	0/1-variable; 1 if state j charges tuition fees at time t
Elite univ.	0/1-variable; 1 if university is selected as a elite university by the policy program Excellence Initiative
Type: TC	0/1-variable; 1 if university type is technical college ("Fachhochschule"), 0 if university type is university ("Universität")

Table A.2: **Summary statistics by university type**

	Universities		**TU**		**TC**		**CAM**		**CE**	
Variable	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
a. Dependent variables										
Students	14,886	9,849	15,056	10,262	4,126	2,557	713	662	3,259	1,169
Graduates	1,448	946	1,396	1,002	604	399	105	110	544	194
PhD graduates	312	274	234	187	.	.	3	2	5	4
Third party funds	27,576	24,414	52,818	46,492	2,099	2,540	256	394	581	571
b. Explanatory variables										
Share ibf fnd. (in %)	53.8	49.9	64.2	48.0	60.5	48.9	52.8	50.3	55.6	49
Age-spec. pop. (in %)	37.5	6.0	37.4	4.4	37.3	6.1	38.3	5.8	40.5	4.6
Share of nat. (in %)	47.1	22.7	76.2	15.2	89.8	15.0	1.8	4.3	5.1	3.6
New degree (in %)	21.1	28.2	21.5	29.9	27.5	36.3	5.3	15.7	2.6	5.4
Tuition fees (in %)	10.5	30.7	8.3	27.8	9.0	28.6	9.3	29.1	16.7	37.5
Elite univ. (in %)	1.6	12.6	1.9	13.5	0.0	0.0	0.0	0.0	0.0	0.0

Notes: TU: Technical universities, TC: Technical colleges, CAM: Colleges of arts and music, CE: Colleges of education. Share ibf fnd. = Share of ibf funding, Age-spec. pop. = Age-specific population, Share of nat. = Share of natural sciences/engineering.

Table A.3: **Effects of indicator-based funding - DiD estimations, full sample, full results**

	Students	**Grad.**	**PhD gr.**	**TPF**
δ	-0.018	0.112*	-0.060	0.004
	(0.023)	(0.059)	(0.060)	(0.068)
Share of ibf funding	0.000	-0.002*	0.001	-0.001
	(0.000)	(0.001)	(0.001)	(0.001)
Age-spec. pop.	-0.005**	-0.009***		
	(0.002)	(0.003)		
Share new degree	0.180***	0.283***		
	(0.035)	(0.072)		
Tuition fees	-0.073**	-0.055	-0.036	-0.102
	(0.029)	(0.049)	(0.093)	(0.079)
Elite univ.	-0.090***	-0.051	-0.112	-0.204***
	(0.032)	(0.062)	(0.108)	(0.075)
Share of natural sc./engin.			-1.093*	0.724**
			(0.610)	(0.341)
Constant	8.100***	6.010***	5.302***	6.314***
	(0.077)	(0.154)	(0.325)	(0.233)
Year effects	Yes	Yes	Yes	Yes
R^2	0.369	0.269	0.098	0.259
F-value	1,942	1,942	668	2,026

Notes: Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.
Standard errors in parentheses.
Grad. = graduates, PhD grad. = PhD graduates, TPF = third-party funds.

Table A.4: **Effects of indicator-based funding - DiD estimations, by intensity, full results**

	Low indicator share				**High indicator share**			
	Students	**Grad.**	**PhD gr.**	**TPF**	**Students**	**Grad.**	**PhD gr.**	**TPF**
δ	0.031	0.233***	0.146	0.044	-0.167***	-0.047	-0.151*	0.119*
	(0.025)	(0.077)	(0.095)	(0.136)	(0.037)	(0.050)	(0.089)	(0.061)
Share of ibf funding	-0.010***	-0.004***	-0.075**	-0.003*	0.001*	0.002***	0.001	0.001
	(0.002)	(0.001)	(0.034)	(0.002)	(0.000)	(0.001)	(0.001)	(0.001)
Age-spec. pop.	-0.005***	-0.003			-0.013	-0.012*		
	(0.002)	(0.003)			(0.010)	(0.006)		
Share new degree	0.231***	0.339***			0.210***	0.198***		
	(0.046)	(0.114)			(0.059)	(0.064)		
Tuition fees	-0.111***	0.054	-0.165	-0.258	-0.052	-0.077*	0.046	0.014
	(0.032)	(0.067)	(0.150)	(0.171)	(0.056)	(0.040)	(0.104)	(0.110)
Elite univ.	-0.096**	-0.103	0.057	-0.466***	-0.027	-0.056	-0.258**	-0.031
	(0.042)	(0.071)	(0.193)	(0.164)	(0.041)	(0.043)	(0.108)	(0.075)
Share of natural sc./engin.			0.900	1.196*			-2.506***	0.627
			(0.635)	(0.663)			(0.865)	(0.394)
Constant	7.946***	5.484***	3.870***	5.384***	8.738***	6.745***	6.401***	6.992***
	(0.066)	(0.181)	(0.338)	(0.420)	(0.352)	(0.213)	(0.448)	(0.265)
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.408	0.229	0.162	0.225	0.385	0.620	0.138	0.410
F-value	1,295	1,280	285	945	647	662	383	1,081

Notes: Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01. Standard errors in parentheses.
Grad. = graduates, PhD grad. = PhD graduates, TPF = third-party funds.

Table A.5: **Number of universities constituting the respective treatment group**

	Low indicator share				High indicator share			
	Stud.	**Grad.**	**PhD gr.**	**TPF**	**Stud.**	**Grad.**	**PhD gr.**	**TPF**
$\delta_{<2yrs.}$	77	76	21	50	57	58	32	91
$\delta_{2-3yrs.}$	75	74	20	55	57	58	32	89
$\delta_{>3yrs.}$	54	60	12	40	56	50	24	69

Notes: The table shows the number of universities in the low and high intensity sample, respectively, being part of the splitted treatment groups (splitted according to time in treatment).

Table A.6: **Effects of indicator-based funding - DiD estimations, by intensity: Differences between university types, full results**

	Low indicator share			High indicator share		
	Students	**Grad.**	**TPF**	**Students**	**Grad.**	**TPF**
δ	0.011	0.360***	-0.099	0.008	-0.049	0.033
	(0.044)	(0.086)	(0.072)	(0.060)	(0.053)	(0.026)
δ*Type: TC	0.002	-0.071	0.006	-0.399***	-0.181*	0.110
	(0.070)	(0.178)	(0.128)	(0.100)	(0.097)	(0.098)
Share of ibf funding	-0.015***	-0.004***	0.001*	-0.002*	0.003***	-0.002***
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share of ibf funding*Type: TC	0.010**	0.000	-0.001	0.006***	-0.000	0.005***
	(0.004)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Tuition fees	-0.197***	-0.101	-0.108	-0.083	-0.162**	-0.172***
	(0.045)	(0.110)	(0.072)	(0.070)	(0.067)	(0.043)
Tuition fees*Type: TC	0.241***	0.334**	0.443***	0.037	0.183**	0.279**
	(0.056)	(0.169)	(0.123)	(0.060)	(0.071)	(0.126)
Age-spec. pop.	-0.003	-0.002		-0.026**	-0.012*	
	(0.003)	(0.005)		(0.013)	(0.006)	
Age-spec. pop.*Type: TC	-0.001	0.004		0.029***	-0.003	
	(0.003)	(0.007)		(0.009)	(0.010)	
Share new degree	0.094	-0.117		0.313***	0.288***	
	(0.097)	(0.336)		(0.090)	(0.082)	
New degree*Type: TC	0.031	0.306		-0.308**	-0.402***	
	(0.117)	(0.372)		(0.125)	(0.132)	
Elite univ.	-0.059	-0.154*	0.219***	0.039	-0.032	0.253***
	(0.057)	(0.084)	(0.066)	(0.041)	(0.070)	(0.080)
Share of natural sc./engin.			0.139			0.492
			(0.506)			(0.510)
Share of natural sc./engin.*Type: TC			-1.521***			0.536
			(0.581)			(0.905)
Constant	7.932***	5.489***	5.825***	8.393***	6.890***	6.862***
	(0.058)	(0.150)	(0.348)	(0.297)	(0.211)	(0.314)
Year effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.484	0.296	0.271	0.462	0.682	0.454
F-value	1,295	1,280	945	647	662	1,081

Notes: Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01. Standard errors in parentheses. Grad. = graduates, TPF = third-party funds.