The effect of department size on quality of research in science: Evidence from the UK research assessment exercise

Otto Toivanen and Michael Waterson

THE EFFECT OF DEPARTMENT SIZE ON QUALITY OF RESEARCH IN SCIENCE: EVIDENCE FROM THE UK RESEARCH ASSESSMENT EXERCISE

Otto Toivanen and Michael Waterson*

February 2013

Abstract

We examine whether the organisation of individual researchers into departments has an impact on the average research quality, beyond the impact of the individuals themselves. Using data on science disciplines generated from the UK Research Assessment Exercises in 1996, 2001 and 2008, we find that department size has an independent positive effect on research quality.

Keywords: economies of scale, RAE, research quality, department size, sciences.

JEL numbers: O31, L88, L52, D83

* Otto Toivanen: Department of Managerial Economics, Strategy and Innovation, Katholieke Universiteit Leuven, and CEPR. Email: otto.toivanen@econ.kuleuven.be. Michael Waterson: Economics department, University of Warwick, CV4 7AL Coventry UK. (Corresponding author): Email: Michael.waterson@warwick.ac.uk. We thank Geert Dhaene, Giles Carden and Andrew Oswald for discussions. Toivanen's research has been supported by GOA/12/003.

1. Introduction

A key question in the organization of scientific research is resource allocation, specifically the question of whether or not resources should be concentrated or dispersed. Significant university funding in most countries comes from the government either directly as part of base funding and/or through competitive research funding. Baldly put, is this best directed at individuals or departments? To inform this debate, we study the question: Does an increase in department size lead to an increase in measured quality of research? If it does, funding systems that encourage concentration of resources are to be preferred over systems that lead to (too much) fragmentation.

We address this question using data on the sciences in the UK. The UK has run one of the largest policy experiments in science policy: the Research Assessment Exercise (RAE). The RAE has introduced competition for scarce resources in a government funded university system by rewarding departments for the quality of their research. Franzoni, Scellato and Stephan (2011) provide evidence that country-level incentives lead to more *Science* submissions, but effects differ between individualized and departmental incentives. The RAE leads to departments containing better scholars getting better funding. Does size brings additional benefits on top of the incentive effects built into the RAE? Our results show that department size has an independent positive impact on research quality. This constitutes a second effect through which the RAE affects quality of research, the first one being the incentive effect.

The RAE provides a unique window on the impact of scale on research quality by providing measures of quality that are designed to be consistent across different scientific disciplines (“fields” in what follows) and which are directly tied to resource allocation. While the RAE has

been studied widely, to our knowledge most existing research has concentrated on the RAE mechanism rather than the effects of RAE on (research) outcomes. As examples, Doyle and Arthurs (1998) study grade inflation in business studies between RAE 1992 and 1996 without finding any; Moed (2008) is a bibliometric study of UK scientists between RAE 1996 and RAE 2001 that finds that UK authors increased the number of articles in high-impact journals, and that departments increased the number of research active staff; Li, Thelwall, Musgrove and Wilkinson (2003) study the relationship between web link based metrics of research quality and the RAE 2001 ratings; and Clerides, Pashardes and Polycarpou (2009) study possible bias in economics ratings in RAE 1996 and RAE 2001 by comparing the RAE ratings to other metrics of research quality. An early paper of somewhat more relevance to us is Johnes, Taylor and Francis (1993; see also Johnes and Taylor, 1992), who find that the size of a cost centre (which is essentially the same as a department; see section 2) is consistently associated with the research selectivity score across most sciences in the 1988/89 research assessment exercise.¹

There is relevant prior research, across a wide variety of different samples, with relatively little consensus: Waldinger (2010, 2012) utilizes the dismissal of Jewish and other scientists from German universities in the 1930s to study how the number and quality of peers (= colleagues in the same department) affect research productivity. Using individual level data in mathematics, physics and chemistry he finds that PhD students are affected by faculty quality (Waldinger 2010), but no evidence of peer effects (2012). Dubois, Rochet and Schlenker (2010) find no evidence of peer effects using worldwide data on mathematicians. Kim, Morse and Zingales (2009) find positive peer effects in economics and finance in the 70s and 80s, and negative peer

¹ However, this early research evaluation is not as clearly associated with the performance of individuals as are more recent exercises; the criteria are related to the group (for example, the median individual within the group).

effects in the 90s. In contrast to Waldinger, these authors cannot control for critical choice of location by the scientist. Azoulay, Zivin and Wang (2010) using U.S. data on life scientists and the death of “star” scientists, find that losing a high quality peer has a negative impact on the co-authors, but the effect is not dependent on losing a (non-) local peer. Borjas and Doran (2012) find that the emigration of ex-Soviet mathematicians had negative effects on the publication records of U.S mathematicians whose research overlapped with the Soviets. In contrast to these papers, evidence (Henderson and Cockburn 1996) suggests that larger research efforts are more productive in the pharmaceutical industry, and that the size of the research portfolio has a positive impact on research productivity.

We contribute to the literature by utilizing the variation caused by a major funding experiment, using data across several science fields, and in concentrating on the effects on department quality. Instead of studying the effect of the quality of individual peers on the research output of individual researchers, we study the effects of the number of peers on the quality of research, measured at the department level, while controlling for past quality of research. We thereby provide evidence that speaks directly to the attractiveness of the existing UK policy.

The rest of the paper is structured as follows: In the next section we detail the data. In section three we provide a descriptive analysis of the data. Section four is devoted to the econometric analysis, and section five concludes.

2. The Data

We collected data on the RAEs carried out in 1996, 2001 and 2008 from the RAE website. These data contain information on the institution (university), unit of assessment (department), field

of science, and the RAE quality assessment outcome. As the RAE data on faculty is not directly informative of department size because departments can choose who to submit, we matched data on faculty numbers by “cost centre”, collected by the Higher Education Statistics Agency (HESA). We took the total number of independent researchers, lecturers and professors as our measure of department size (the list of job categories can be found in the Appendix). Several of the HESA cost centres can be matched to RAE units of assessment:² Psychology, Biology, Chemistry, Physics, Mathematics (where we sum together pure and applied mathematics and statistics), Computer Science, General Engineering, Chemical Engineering, Civil Engineering, Electrical and Electronic Engineering, Mechanical Aeronautical and Manufacturing Engineering, and Metallurgy and Materials.³ These constitute the fields of science we have in our data.

Our unit of analysis is a department in a given year. We observe 334 departments for all three waves of RAE. These department-year observations constitute our sample. Below in section three we compare these departments to those departments for which we have less than three years’ observations.⁴

Whereas in years 1996 and 2001, departments were given a grade on a seven point scale of 1 to 5*, in 2008 RAE assessed “quality profiles” of departments on a five point scale (0 to 4*). These can be disaggregated into “Research outputs”, “Research environment”, and “Esteem indicators”. Our measure of the research quality of a department in the 2008 RAE is the proportion of staff, in per cent, allocated to the highest quality category (4*) according to

² For more information on how we matched the HESA cost centres to the RAE units, see Table A1 in the Appendix.

³ A major omission is Medicine, where we were unable to achieve a good match across successive RAEs.

⁴ It turned out that we cannot include departments from merged universities, most notably the University of Manchester. The reason for this is that e.g. in the case of University of Manchester, both merging universities had cost centres in (most of) the fields that we consider. This means that the post- and pre-merger departments are not comparable.

“Research outputs”.⁵ We decided to exclude the “environment” and esteem “categories” because they are less closely associated with individuals and their (current) research output, and they might themselves potentially relate to departmental size, without any necessarily causal explanation.

We rescale the quality of research measures from 1996 and 2001 to the 2008 measures.⁶ To do this, we first rank the 334 departments in each year. We then assign each department in year 1996 and 2001 that 2008-style grade which it would have achieved in 2008 if it had the same rank in 2008 as it had in 1996 or 2001. As the 1996 and 2001 grades were categorical, there are multiple departments with the same grade. For these departments we give the average 2008-style grade of the departments in the same ranks in 2008. As an example, departments that achieved grade 5* in 2001 were ranked 280-334.⁷ We assign each of them the average grade of the departments ranked 280 – 334 in 2008. This naturally introduces some measurement error into our dependent variable. That is a concern only because we will estimate a dynamic panel data model, meaning that the lagged dependent variable also suffers from measurement error. We deal with this by using an appropriate estimator.

⁵ The measures take the following values (see e.g. <http://www.rae.ac.uk/results/intro.aspx>):

4* Quality that is world-leading in terms of originality, significance and rigour.

3* Quality that is internationally excellent in terms of originality, significance and rigour but which nonetheless falls short of the highest standards of excellence.

2* Quality that is recognised internationally in terms of originality, significance and rigour.

1* Quality that is recognised nationally in terms of originality, significance and rigour.

1 Quality below that is unclassified.

⁶ See the Appendix for details.

⁷ Note that the rank is increasing in the grade.

3. Descriptive Analysis

We present descriptive statistics of department quality in Table 1 where we report the fraction of faculty given grade 4 (“Quality that is world-leading in terms of originality, significance and rigour), and the fraction of faculty given either grade 4 or 3 (“Quality that is internationally excellent in terms of originality, significance and rigour but which nonetheless falls short of the highest standards of excellence”) in the 2008 RAE, and our equivalent measures of department quality for 1996 and 2001. Our method of converting the 1996 and 2001 grades results in the means being equal across the years. However, one can observe some variation in the quartiles: The median varies from 13 in 1996 and 2008 to 15.5 in 2001. The 25th percentile increases over time, whereas the 75th percentile fluctuates.

[TABLE 1 HERE]

In Table 2 we report descriptive statistics of changes in department quality. These are defined as $(\text{quality in RAE year } t - \text{quality in RAE year } t-T) / \text{quality in RAE year } t-T$. The mean department has improved its quality over each RAE cycle, but there is wide variation. Departments at the 25th percentile have worsened in quality, the median department has more or less maintained its quality, and the departments at the 75th percentile have significantly improved.

[TABLE 2 HERE]

As can be seen from Table 3, the 2008 grade is highly correlated with the 1996 and 2001 grades, with correlation coefficients of 0.57 (p-value < 0.000) and 0.71 (p-value < 0.000). The

correlation between the grade in 1996 and change in grade between 1996 and 2008 is -0.31 (p-value < 0.000), consistent with Galtonian regression to the mean.

[TABLE 3 HERE]

We have collected descriptive statistics on department size into Table 4. The average department size in our estimation sample has changed from 36.5 in 1996 to 40.9 in 2008 with little change in the standard deviation. There is a clear shift in the distribution, with all quartiles in 2008 (25%: 23.3, 50%: 33.7, 75%: 49.1), higher than in 1996 (25%: 19.7, 50%: 30.1, 75%: 44).⁸ The median department has thus more than 3 FTE faculty extra in 2008 than in 1996. Looking at change in FTE equivalents (middle panel) we see that the mean department has grown by 2 (2.5) faculty members between 1996 and 2001 (2001 and 2008). The departments at the 25th percentile of the growth distribution have lost between 2.5 and 5 FTE faculty whereas those at the 75th percentile have gained 7-8 FTE faculty members. In proportional terms, the difference between mean and median, as well as the quantiles, testify that the distribution of proportional growth is very skewed, with the median department growing at a moderate pace, but some departments growing very rapidly. This partly reflects that fact that there are some very small departments in the data: in them, one or two new faculty members lead to very high growth rates, skewing the distribution.

[TABLE 4 HERE]

To study this change further, we show in Figure 1 the distributions of “normalized” department sizes where we display the distribution of the residuals of an estimation where we regressed

⁸ These numbers are slightly lower in the whole sample (see Appendix). Our estimator controls for selection by size.

department size in years 1996, 2001 and 2008 onto field dummies, thus cleaning out differences in department size across fields of science. The mean of this normalized department size is zero by construction over the three years. We used our estimation sample of 334 departments observed in all three RAEs. As can be seen, the distribution has shifted to the right over time, meaning that departments across the whole distribution have become larger over time.

[FIGURE 1 HERE]

Just like research quality, department size is correlated over time (Table 5). The correlations are quite high at round 0.8.

[TABLE 5 HERE]

It is conceivable that it is not only the size of the department that matters for research quality, but the size of the science community at the university. To study whether the size of the science faculty, measured at the university level, has an effect on research quality we aggregated our department level measures of faculty to a university level measure of faculty.⁹ These are reported in Table 6. There we see that the science faculty of the universities has grown from an average of 270 to 305 over 12 years. The growth seems to have affected all quartiles, with the 25th percentile growing from 160 to 174 between 1996 and 2008, the median from 252 to 320, and the 75th percentile from 400 to 447.

[TABLE 6 HERE]

We then turn to the correlation between department size and research quality in Table 7. These are contemporaneously correlated with correlation coefficients of 0.20 (p-value < 0.000)

⁹ There are obvious measurement issues here as 1) we were unable to match some RAE science units to HESA cost centres 2) we only aggregate over those departments which we observe over the three RAEs.

in 1996, 0.29 (p-value < 0.000) in 2001 and 0.15 (p-value < 0.000) in 2008. While positive, these correlations are lower than those over time for both quality and for size. Nonetheless, the descriptive analysis shows that department size and research quality are positively correlated.

As a final issue in this section, we turn to the possible differences in size between the departments included in our estimation sample, and other science departments. Recall that we need to observe a department in all three waves of the RAE to be able to include it in our estimation sample. We have collected statistics on those departments that do not fulfil this criterion into Table 8. Looking first at department size, we see that the excluded departments are smaller in size in each RAE; these differences are statistically significant. We also see that they obtain lower quality scores in 1996 and 2001, but higher quality scores in 2008, than the 334 departments in our estimation sample. One explanation for this might be university mergers between the 2001 and 2008 RAEs, since we had to exclude cost centres from merging universities because their sizes before and after the merger are not comparable. Most notably, several of University of Manchester's departments scored very well in the 2008 RAE. There is therefore some reason to worry about selection into our sample. While the fact that we first difference and thereby purge any effects from department (and thereby university) specific time-invariant effects partly ameliorates this effect, this feature of our data may dampen the external validity of our study.

[TABLES 7 AND 8 HERE]

4. The Econometric Analysis

4.1. The econometric model

When estimating the independent effect of department size on department research quality, possible impacts of past quality on current quality need to be controlled: There are numerous departments that have stayed at the top of research quality rankings for many years. We use the widely employed dynamic panel data estimator (Anderson and Hsiao 1982, Arellano and bond 1991, Baltagi 2008).¹⁰ This estimator allows us to control for department (and hence also university-) specific unobserved differences that are invariant over time, and for the endogeneity of the lagged quality and contemporaneous department size.¹¹ These properties of the estimator allow a causal interpretation of our size coefficient.

The equation we start from is

$$Quality_{ijt} = \alpha Quality_{ijt-1} + \beta Faculty_{ijt} + \mu_i + \epsilon_{ijt}. \quad (1)$$

In (1) $Quality_{ijt}$ is the quality of research in department i (i.e., field i) in university j at time t , $Faculty_{ijt}$ is the size of department i (field i) in university j in time t , μ_i is a department (and university) specific unobserved, time-invariant, quality shock, ϵ_{ijt} is a department (and university) and period-specific shock to department quality, and α and β are parameters to be estimated. Thus, μ_i captures for example permanent (i.e., time-invariant) differences at department or university level on management practices or financial and other endowments that affect research quality, whereas ϵ_{ijt} captures the period-specific shocks to department quality. After taking first differences this can be written as:

$$\Delta Quality_{ijt} = \alpha \Delta Quality_{ijt-1} + \beta \Delta Faculty_{ijt} + \Delta \epsilon_{ijt}, \quad (2)$$

¹⁰ For an accessible survey, see Bond (2002).

¹¹ See the Appendix.

where Δ is the difference operator.

The identification assumptions we make are: First, the initial conditions are assumed to be uncorrelated with the subsequent period-specific shocks ϵ_{ijt} , i.e., that the initial conditions are predetermined (e.g. Bond 2002, pp. 145). This means that the dependent variable $Quality_{ijt}$ may be correlated with unobserved, time-invariant, department-specific effects, μ_i . Second, the period-specific shocks ϵ_{ijt} are assumed to be serially uncorrelated. This renders the twice lagged level of the dependent variable $Quality_{ijt-2}$ (department research quality in 1996 in our case) a valid instrument for the differenced lagged dependent variable $\Delta Quality_{ijt-1}$ (the difference between department research quality in 2001 and 1996 in our case). Third, we assume that the department size variable $Faculty_{ijt}$ is predetermined, i.e., that department size in RAE for year t is uncorrelated with the year- t period-specific shock ϵ_{ijt} . Department size is allowed to be correlated with department-specific time-invariant unobservable factors μ_i , and earlier period-specific shocks. As an example, we allow the department size in 2008 to be correlated with the shocks to research quality in 2001, but the 2008 RAE is an independent exercise. Notice that we do test for the implication of over-identification that the assumption of department size being predetermined yields (against the alternative of department size being endogenous).

It is worthwhile noting that we take what one could call “long” differences, as the RAEs we use were executed at intervals of five and seven years.¹² We do not utilize the level-moments introduced by Blundell and Bond (1998). We experimented with using them, but over-

¹² See Griliches and Hausman (1986) and Hahn, Hausman and Kuersteiner (2007).

identification tests always suggested that we reject the null hypothesis of being able to use those moments.

4.2 The econometric analysis

We present the estimation results in Table 9. Using OLS (column 1) we find a coefficient of 0.6 for the lagged dependent variable, and a coefficient of 0.025 for the department size variable (“faculty”). As is well known from the literature on dynamic panel data estimators, the OLS estimator of the coefficient of the lagged dependent variable is upward, the fixed effects estimator downward biased. We observe this pattern in our data, as the fixed effects estimate of the coefficient of the lagged dependent variable is -0.58 and statistically significant. The coefficient of department size is however again positive and significant. It is somewhat higher than the OLS estimate. Using our preferred dynamic panel data estimator that treats department size as predetermined, we find that the effect of lagged department quality is positive but very imprecisely measured, and that department size has a positive causal effect on research quality, consistent with our main hypothesis. Our estimator passes the standard specification tests for dynamic panel data estimators. Statistical tests also suggest that it is prudent to treat department size as predetermined: We could not reject the null hypothesis that department size is predetermined instead of exogenous, but rejected the null that it is endogenous instead of predetermined.

In terms of economic significance, the coefficient of 0.244 on department size implies that raising department size by one individual increases the proportion of faculty whose research is “world-leading in terms of originality, significance and rigour” by 0.24 percentage points. Thus, a department that increased size by one standard deviation of the department size distribution

(26.64) would be predicted to improve research quality by 6.46 percentage points, half a standard deviation of the research quality distribution. The effects that this would have on resource allocation depend on the specific field (and assessment round).

[TABLE 8 HERE]

To gauge the robustness of the result, we estimated a number of alternative specifications which are reported in Table 10. In the first column we use as the dependent variable the proportion of faculty in the two highest quality categories of RAE 2008 (instead of only the highest). The coefficient of department size is more than twice that in main equation (0.665 versus 0.244) and statistically significant at the 10% level. In the second column we exclude the smallest and largest departments, measured by size in 1996. Columns 3 and 4 report results when we exclude mathematics (these being “mergers” between mathematics, applied mathematics, and statistics departments, due to the accounting convention of HESA). Excluding mathematics has no effect on the coefficient of department size, but excluding psychology leads to a smaller coefficient of 0.142 which is still statistically significant. Adding university level science faculty (column 5) has no effect, while adding the squared department size (column 6) leads to a higher point estimate for linear department size and a small (in absolute value) and statistically insignificant coefficient for squared department size. A different (broader) definition of faculty to include Researcher A and B on the PCEF scale and Researcher on the CSCFC lead to no change in the coefficient of department size.

Finally, we re-estimated the model using two-step GMM standard errors with the so-called Windmeijer correction, and allowing for clustering at the university level. In both cases, the estimated standard errors were smaller than the one-step robust standard errors reported.

Using the two-step Windmeijer corrected covariance matrix we obtained a coefficient 0.241 for the department size variable ($Faculty\ in\ field_{ijt}$) with standard error 0.105. To be able to use standard errors clustered at the university level we used the standard two stage least squares estimator. The department size coefficient was 0.244 with standard error 0.103. These compare to the coefficient of 0.244 and standard error of 0.111 obtained using the robust one-step estimator that are reported in Table 8.

[TABLE 9 HERE]

5. Conclusions

In the UK sciences, during 1996 – 2008, department size has had an independent positive and, we suggest, causal effect on the quality of research. This is positive news for the RAE (and its successor), which provides incentives for departments to improve the quality of their research, and therefore implicitly builds on the premise that the possibly resulting concentration of resources is beneficial overall to the UK scientific endeavour. Our result suggests that such a robust reinforcing factor exists and is non-negligible.

The mixed results in previous work caution against extrapolating our result. As is the case with any econometric analysis, our result is dependent on the institutional environment in which the data we use has been generated. Our data were generated in an environment in which incentives are provided for departmental reorganization. This is not the case in all countries. One should therefore take care when drawing (wrong) policy conclusions of the benefits of increasing faculty size by mechanical, top-down approaches, e.g., by forcing departmental

mergers across scientific disciplines. In an environment without RAE style incentives for quality of research, concentration may not lead to the hoped-for outcome.

References

- Anderson, T.W. and C. Hsiao (1982) Formulation and estimation of dynamic models using panel data, *Journal of Econometrics*, 18, 47-82.
- Arellano M. and S. Bond (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *The Review of Economic Studies*, 58, 277-297.
- Azoulay, P., J. S. Graff Zivin and J. Wang (2010) Superstar extinction, *Quarterly Journal of Economics*, 125, 549-589.
- Baltagi B. H., 2008, *Econometric analysis of panel data*, John Wiley & Sons Ltd, UK.
- Borjas G. and K. Doran (2012), The collapse of the Soviet Union and the productivity of American mathematicians, *The Quarterly Journal of Economics*, 1143-1203.
- Clerides, S., P. Pashardes and A. Polycarpou (2009) Peer review vs metric-based assessment: Testing for bias in the RAE ratings of UK economics departments, *Economica*, pp. 1-20.
- Doyle, J. and A. Arthurs (1998) Grade inflation in the UK's 1996 Research Assessment Exercise?, *Omega*, 26, pp. 461-465.
- Dubois P., J.-C. Rochet and J.-M. Schlenker (2012) Productivity and mobility in academic researcher: Evidence from mathematicians, mimeo, U. of Toulouse.
- Franzoni C., G. Scellato and P. Stephan (2011) Changing incentives to publish, *Science*, 333, 702-703.
- Griliches, Z., and J. Hausman (1986) Errors in variables in panel data, *Journal of Econometrics*, 31, pp. 93-118.
- Hahn, J., J. Hausman and G. Kuersteiner (2007) Long difference instrumental variables estimation for dynamic panel models with fixed effects, *Journal of Econometrics*, 140, pp. 574-617.
- Henderson, R. and I. Cockburn (1996) Scale, Scope and Spillovers: The Determinants of Research Productivity in Drug Discovery, *RAND Journal of Economics*, 27, 1, pp. 32-59.
- Johnes, J. and Taylor, J. (1992) The 1989 Research Selectivity Exercise: A Statistical Analysis of Differences in Research Rating between Universities at the Cost Centre Level *Higher Education Quarterly* 46, 1, pp 67-87.
- Johnes, J., Taylor, J. and Francis, B. (1993) The research performance of UK universities: a statistical analysis of the results of the 1989 research selectivity exercise. *Journal of the Royal Statistical Society A* 156, 2, pp. 271-286.

Kim E. H., A. Morse and L. Zingales (2009) Are elite universities losing their competitive edge? *Journal of Financial Economics*, 93, 353-381.

Li, X., M. Thelwall, P. Musgrove, D. Wilkinson (2003) The relationship between WIFs of inlinks of computer science departments in UK and their RAE ratings or research productivities in 2001, *Scientometrics* 57, pp. 239-255.

Moed, H. (2008) UK Research Assessment Exercise: Informed judgement on research quality or quantity? *Scientometrics*, 74, pp. 153-161.

Waldinger F. (2010) Quality matters: The expulsion of professors and the consequences for PhD student outcomes in Nazi Germany, *Journal of Political Economy*, 118, 787-831.

Waldinger F. (2012) Peer effects in science – evidence from the dismissal of scientist in Nazi Germany, *The Review of Economic Studies*, forthcoming.

Table 1. Department research quality						
	1996		2001		2008	
Statistics	class 4	class 3 and 4	class 4	class 3 and 4	class 4	class 3 and 4
Mean fraction of faculty	13.33	54.85	13.33	54.85	13.33	54.85
Sd	8.13	18.21	7.94	17.95	8.37	18.73
p25	6.81	44.03	7.35	45.48	7.50	44.80
p50	13.01	57.34	15.48	61.95	13.05	56.40
p75	20.35	70.47	15.48	61.95	18.30	67.50
N	334	334	334	334	334	334

Table 2. Changes in department quality			
Statistics	2008/2001	2001/1996	2008/1996
mean	0.26	0.28	0.46
sd	1.34	1.24	2.49
p25	-0.37	-0.24	-0.38
p50	-0.05	0.08	-0.04
p75	0.39	0.34	0.52
N	324	332	332

Table 3. Correlations of research quality		
	1996	2001
1996	1.00	
2001	0.71 (0.00)	1.00
2008	0.57 (0.00)	0.51 (0.00)
Note: reported numbers are correlation coefficient and (p-value).		

Table 4. Full-time equivalent (FTE) department level faculty numbers			
FTE faculty			
Statistics	1996	2001	2008
mean	40.81	38.78	36.18
sd	26.16	27.92	25.12
p25	23.30	21.45	19.30
p50	33.70	30.80	30.01
p75	49.00	45.92	44.90
N	334	334	334
Change in FTE faculty			
Statistics	2008/2001	2001/1996	2008/1996
mean	2.03	2.60	4.63
sd	18.12	15.92	17.90
p25	-4.42	-2.45	-4.44
p50	2.20	2.32	4.06
p75	8.39	7.31	12.41
N	334	334	334
Proportional change			
Statistics	2008/2001	2001/1996	2008/1996
mean	0.49	1.54	1.68
sd	5.11	12.01	11.22
p25	-0.13	-0.08	-0.12
p50	0.07	0.08	0.16
p75	0.31	0.29	0.47
N	334	334	334

Table 5. Correlations of department size		
	1996	2001
1996	1.00	
2001	0.82 (0.00)	1.00
2008	0.76 (0.00)	0.78 (0.00)
Note: reported numbers are correlation coefficient and (p-value).		

Table 6. University level science FTE faculty			
stats	1996	2001	2008
mean	268.72	288.00	304.29
sd	148.55	165.66	152.94
p25	152.55	150.45	173.50
p50	252.93	283.42	319.90
p75	396.17	414.15	446.90
N	334	334	334

Table 7. Correlations of department size and quality in different RAEs						
	1996		2001		2008	
	size	4 star	size	4 star	Size	4 star
Size	1.00		1.00		1.00	
4 star quality	0.21 (0.00)	1.00	0.30 (0.00)	1.00	0.15 (0.00)	1.00
3 and 4 star	0.20 (0.00)	0.96 (0.00)	0.29 (0.00)	0.96 (0.00)	0.07 (0.23)	0.79 (0.00)
Note: reported numbers are correlation coefficient and (p-value).						

Table 8. Differences in mean department size and quality between departments in the sample and not in the sample						
Department size						
	1996		2001		2008	
	not in sample	in sample	not in sample	in sample	not in sample	in sample
Mean	26.71	36.18	23.86	38.78	32.37	40.81
N	150	334	143	334	114	334
p-value		0.00		0.00		0.00
Department quality						
	1996		2001		2008	
	not in sample	in sample	not in sample	in sample	not in sample	in sample
Mean	3.38	4.03	3.83	4.57	18.46	13.33
N	150	334	143	334	114	334
p-value		0.00		0.00		0.00
Note: p-value is for a two-sided t-test on the differences between the means. In 1996 and 2001, the quality measure is an index running from 1 to 5.5 (a seven point scale, in effect, because category 3 was subdivided). In 2008, it is the value as used in the regressions.						

Table 9. Estimation results			
	OLS	FE	AB
Quality_1	0.596*** (0.032)	-0.581*** (0.067)	0.430 (0.393)
Faculty	0.025** (0.010)	0.046*** (0.018)	0.244** (0.111)
Overidentification and exogeneity tests, p-value			
Sargan overid	-	-	0.930
Hansen overid	-	-	0.941
Diff-in-Hansen	-	-	0.941
F-tests of instruments in 1st stage			
Quality_1	-	-	39.816
Faculty	-	-	12.660
Observations	668	668	334
R-squared	0.376	0.210	-
<p>Notes: reported numbers are coefficient and (robust one-step s.e.).</p> <p>Quality_1 = quality measured at the previous RAE.</p> <p>Faculty = number of FTE faculty as reported in HESA.</p> <p>OLS = ordinary least squares.</p> <p>FE = fixed department effects.</p> <p>AB = Arellano-Bond.</p> <p>In AB, Quality_1 treated as endogenous, Faculty as predetermined.</p> <p>Sargan overid = p-value of instrument overidentification test.</p> <p>Hansen overid = p-value of instrument overidentification test.</p> <p>Diff-in-Hansen = p-value of difference-in-Hansen test exogeneity of instruments.</p> <p>F-test Quality_1 = F-test of the t-2 lagged level of department quality in the 1st stage.</p> <p>F-test Faculty = F-test of t-1 and t-2 lagged levels of department size in the 1st stage.</p> <p>*** p<0.01, ** p<0.05, * p<0.1</p>			

Table 10.
Robustness tests

	use class3 and class 4	exclude smallest and largest dept.	exclude mathematics	exclude psychology	include univ. science faculty	include faculty squared	extended faculty definition
Quality _{t-1}	0.838 (0.894)	0.334 (0.335)	0.433 (0.411)	-0.124 (0.236)	0.370 (0.515)	0.069 (0.302)	0.424 (0.389)
Faculty in field	0.680* (0.348)	0.206** (0.087)	0.240** (0.102)	0.144** (0.061)	0.237** (0.096)	0.326* (0.178)	0.238** (0.108)
Univ. faculty		-	-	-	-0.012 (0.028)	-	0.424
Faculty in field sq.		-	-	-	-	-0.001 (0.001)	-
	334	332	302	279	334	334	334

Notes: reported numbers are coefficient and (robust one-step s.e.). All equations estimated using Arellano-Bond, with Quality_{t-1} treated as endogenous and Faculty in field and University faculty treated as predetermined. *** p<0.01, ** p<0.05, * p<0.1.

Figures 1: Distribution of normalized department size

Appendix

In this Appendix we report 1) the construction and matching of the data and 2) the rescaling of the 1996 and 2001 RAE scores to match the 2008 RAE scale.

1. Construction and matching of the data

Our sources of RAE data are various pages on the RAE www-site <http://www.rae.ac.uk/>. We downloaded the following data:

1. Data on staff numbers for RAE 2008 from <http://www.rae.ac.uk/Submissions/default.aspx>
2. Data on FTE staff numbers from <http://www.rae.ac.uk/Results/>
3. [Data on the 2008 RAE http://www.rae.ac.uk/Submissions/default.aspx](http://www.rae.ac.uk/Submissions/default.aspx)
4. All RAE 2001 data from <http://www.rae.ac.uk/2001/results/>
5. All RAE 1996 data from <http://www.rae.ac.uk/1996/database/index.html>

Our source for the HESA data is Higher Education Statistics Agency (HESA), from where we obtained the files 32826_Data.xlsx and 32826_Definitions.doc. These contained information at the institution/cost centre/year level on the number of employees with different job titles (see vii. below).

All data apart from the quality profiles were available as zipped excel files. The quality profiles were manually copy pasted into an excel file.

We then did the following to produce a matched data set:

- i. We organized the 1996 RAE sub-profile data (for some units of assessment, there were multiple submissions) into one observation / unit of assessment and institution.
- ii. We then created unit of assessment numbers to match the RAE 2008.

- iii. We cleaned the RAE 2001 data of multiple submission / unit of assessment and institution. RAE 2001 grades were calculated as staff-weighted averages in case of a multiple submission. As staff weights we used the full-time equivalent of the number of staff who were submitted to the RAE.
- iv. The 1996 and 2001 data were matched by higher education institution identifier ("hesainst") and unit of assessment number ("uoa").
- v. The RAE 2008 data was cleaned of multiple submissions (see above) into one observation / unit of assessment and institution. RAE 2008 grades were calculated as staff-weighted averages in case of a multiple submission. As staff weights we used the full-time equivalent of the number of staff who were submitted to the RAE.
- vi. The 1996 and 2001 RAE data were matched to the 2008 RAE data by higher education institution identifier ("hesainst") and unit of assessment number ("uoa").
- vii. The HESA data was aggregated to cost centres by institution (=university). We included the following job categories into our faculty count:

- (01) Lecturer (PCEF scale)
- (02) Senior lecturer (PCEF scale)
- (03) Principal lecturer (PCEF scale)
- (04) Head of department (PCEF scale)

- (74) Locally determined scale - Researcher
- (73) Locally determined scale - Lecturer
- (72) Locally determined scale - Senior/Principal lecturer
- (71) Locally determined scale - Professor

- (37) Research grade II (UAP scale)
- (38) Research grade III (UAP scale)
- (39) Research grade IV (UAP scale)
- (31) Lecturer A (UAP scale)
- (32) Lecturer B (UAP scale)
- (33) Senior lecturer (UAP scale)
- (34) Professor (UAP minimum)

- (40) Clinical lecturer
- (41) Clinical senior lecturer
- (42) Clinical professor

- (61) Lecturer (CSCFC scale)
- (62) Senior lecturer (CSCFC scale)
- (63) Professor/Head of department (CSCFC scale)

We created separate variables as follows: Researcher (all included researchers and research grades), Lecturer (all lecturers), Senior lecturer (all senior lecturers and principal lecturers), Professor (all professors and heads of department). These were then summed together to form our key explanatory variable “ $Faculty_{ijt}$ ”, measuring in full-time equivalents the number of employees that are research-active in department i in institution j at time t .

- viii. We checked that the institution code and name in HESA corresponds to that in the RAE.
 - ix. We then matched 2008 RAE unit of assessment identifier to HESA cost centre identifier as shown in Table A1:
- [TABLE A1 HERE]
- x. The HESA and the RAE 1996 – 2008 data were matched using institution code and RAE 2008 unit of assessment identifier.

2. Re-scaling of the 1996 and 2001 RAE scores to match the 2008 RAE scale

The procedure we followed was as follows:

1. We ranked all 334 departments in our estimation sample in each of the three years 1996, 2001 and 2008 by their RAE grade (i.e., these are departments that were in all 3 RAEs).
2. For the 1996 and 2001 ranking, we did the following:
 - a. Starting from the top, we would assign the department ranked n^{th} in 1996 (2001) the grade that the n^{th} ranked department obtained in 2008.

- b. If several departments obtained the same grade in 1996 (2001), say departments ranked n^{th} to m^{th} , all these departments were awarded the average grade of the departments ranked n^{th} to m^{th} in 2008.

To give an example, in 2001 departments ranked 143 – 278 (where rank 1 indicates the lowest grade) all were awarded grade 5. These departments were assigned the grade average of the departments ranked 143 – 278 in 2008.

Notice that this rescaling introduces some measurement error into our *dependent* variable that is classical or almost so, and thus is not an issue otherwise, but is one further reason why the lagged dependent variable needs to be instrumented.

This procedure yielded our dependent variable “ Quality_{ijt} ” which measures the quality of department i in institution j at time t .

Table A1: Matching RAE unit of assessments and H ESA cost centres

RAE 2008 unit of assessment		HESA cost centre
10	Dentistry	(02) Clinical dentistry
14	Biological Sciences	(10) Biosciences
18	Chemistry	(11) Chemistry
19	Physics	(12) Physics
20	Pure Mathematics	(24) Mathematics
21	Applied Mathematics	(24) Mathematics
22	Statistics and Operational Research	(24) Mathematics
23	Computer Science and Informatics	(25) IT and Systems Sciences, Computer Software Engineering and for previous years, (39) (amalgamated 2008.
24	Electrical and Electronic Engineering	(20) Electrical, electronic & computer engineering
25	General Engineering	(16) General engineering
26	Chemical Engineering	(17) Chemical engineering
27	Civil Engineering	(19) Civil engineering
28	Mechanical, Aeronautical and Manufacturing Engineering	(21) Mechanical, aero & production engineering
29	Metallurgy and Materials	(18) Mineral, metallurgy & materials engineering
44	Psychology	(07) Psychology and Behavioural Sciences

FACULTY OF ECONOMICS AND BUSINESS
DEPARTMENT OF MANAGERIAL ECONOMICS, STRATEGY AND INNOVATION

Naamsestraat 69 bus 3500
3000 LEUVEN, BELGIË
tel. + 32 16 32 67 00
fax + 32 16 32 67 32
info@econ.kuleuven.be
www.econ.kuleuven.be/MSI