The impact of overeducated and undereducated workers on firm-level productivity - First evidence for Germany

Philipp Grunau[*]
Institute for Employment Research

Preliminary, May 2014

Please do not circulate

Abstract

Many literature contributions deal with the productivity effects of excess and deficit educational attainment of workers on the individual level. Since it is conceivable that there are spillover effects from those mismatched to their well-matched colleagues, from an employer's vantage point it is highly important to know the net effect of educationally mismatched employees on productivity on the firm level. Providing first representative evidence for Germany, this study analyses the impact of overeducated and undereducated workers among a firm's workforce on its productivity.

Using linked employer-employee data I estimate dynamic panel production functions by means of a system GMM estimator. I find that undereducated workers among a firm's workforce impair its (firm-level) productivity, implying that establishment's HR management should avoid the recruitment of undereducated workers. The effect for overeducated employees turns out negative, too, albeit small and insignificant.

JEL classification: J21, J24, J82, M51

Keywords: educational mismatch, overeduation, undereducation, productivity, system GMM, linked panel data.

[*] Institute for Employment Research, Department Establishments and Employment, Regensburger Str. 104, 90478 Nuremberg, Germany. E-mail: philipp.grunau@iab.de, phone: +49-911-179-7943.

Introduction

The process of educational expansion is prevalent in most countries and supported by politics. For instance, an official goal of the European Union states the enhancement of participation in (higher) education by a substantially degree. So far, the target rate of successful completion of tertiary education (for the population aged 30 to 34) within the EU28 of 40% in 2020 has been converged to up to 35,7% in 2012 (European Commission, 2013a). If this development is not met or preceded by changes in the demand related to requested education, overeducation arises (McGuinness, 2006). For the European Union, this mismatch of demand and supply side has substantially increased, especially since the outbreak of the Great Recession in 2008 (European Commission, 2013b). In fact, overeducation applies to 21% of individuals with a tertiary education degree within the EU28 in 2012 (European Commission, 2013a). Additionally, in case of horizontal mismatches due to deficit supply of highly educated workers within certain fields of education (e.g. STEM), firms may be forced to employ undereducated workers from this field.

When it comes to an assessment of the importance of educational mismatch, one may stress the cost and benefit dimension of educational mismatch (e.g. McGuinness, 2006). Both on the individual and the societal level, the occurrence of overeducation (undereducation) is usually accompanied by direct and opportunity costs (benefits). However, the role of firms should not be neglected, as the occurrence of educational mismatch is clearly conditional on these labour market agents' decision to allocate a certain applicant to a certain job in the first place. Be the reasons underlying this decision of forced (skilled worker shortage) or of deliberate (expected advantages, e.g. productivity-related) nature, it is highly relevant for establishments to know which effects can be expected in order to adjust their personnel management strategy accordingly. As Hartog (2000) states: "It would obviously be highly informative if we knew the effect of over- and undereducation on productivity, rather than on wages." (ibid., p. 139)

Numerous studies so far have dealt with this question, albeit facing two substantial restrictions. First, they link overeducation and undereducation, respectively, with individual performance through several related proxies like job satisfaction, absenteeism, shirking or – most often – through wages, which at least triggers some doubts due to the significant but restricted link between these proxies and productivity. And second, these analyses address the correlation at the individual rather than on the firm level. Allowing for the possibility of (indirect) spillover effects from mismatched workers on co-workers' productivity aside from direct effects would require estimating a net effect on the estab-

lishment level. Furthermore, for firm-level productivity we have direct and much more convincing measures as opposed to the proxy approach for the estimation on the individual level.
However, there have barely been efforts addressing this issue on the establishment level. In the pioneering paper, Kampelmann and Rycx (2012) using linked employer-employee panel data for Belgium detected an enhancing effect of overeducation and an impairing effect of undereducation on firm-level productivity. For Germany, though, apart from a case study by Haugrund (1990) with restricted external validity, to my best knowledge no comparable analyses have been performed as yet.

I exploit a large-scale linked employer-employee dataset for Germany, comprising information from both detailed survey data on firms and administrative data on these firms' employees. For the panel period 2004 to 20010/11 a full set of information for about 23,000 establishment-year observations is available. By means of a system GMM estimator I account for potential endogeneity caused by either time-invariant (e.g. long-term HR management strategy), time-varying unobserved firm heterogeneity (e.g. concerning the focal regressors of this analysis, the shares of overeducated an undereducated workers, respectively), or autocorrelated errors due to the neglected (proper) consideration of a dynamic production process.

The remainder of this paper is structured as follows: Section 2 provides a brief literature review and a discussion of potentially conceivable results. It is followed by a description of the data (Section 3) and the econometric strategy (Sector 4). The results of the analysis are presented in Section 5 including a brief discussion of their sensitivity to changes in the analysis approach. Section 6 concludes and derives implications for research, firms, and policy makers.

Literature review and theoretical considerations

As yet, there has been a broad range of studies analysing the impact of over- and undereducation on wages on the individual level, most of which address the issue of discrepancies in the returns to different parts of education. These analyses agree on the finding that overeducated (undereducated) workers in terms of wages get rewarded (punished) for their surplus (deficit) education when compared to their properly matched colleagues with higher (lower) educational attainment. A comprehensive review of respective literature contributions is provided by McGuinness (2006) or Leuven and Oosterbeek (2011) in their meta-analyses. However, in order to be able to draw conclusions from these findings on the impact on productivity would require the crucial assumptions derived from Human Capital Theory (Becker, 1964) that (i) additional educational attainment enhances an individual's productivity and that (ii) productivity differences directly translate into earnings variations. Considering the positive yet imperfect correlation between education, productivity and wages (Dearden, Reed, and Van Reenen, 2006), using wages as sole proxy and thus neglecting other crucial factors can result in misleading conclusions and implications and have to be taken with caution.

Aside from wages there have been other attempts to proxy individual productivity. Tsang et al. (1991) find a negative correlation between overeducation and job satisfaction. Using the detour over job satisfaction, Tsang (1987) by use of two equations estimates that beyond the negative effect on job satisfaction there is also a positive correlation with firm output, implying a negative correlation between overeducation and firm productivity. Buechel (2002), in contrast, does not find a correlation of overeducation with job satisfaction for Germany. Furthermore, he addresses the impact of other potential correlates like health status, training participation and firm tenure. Against his initial intuition he concludes that overeducated individuals (in low-skill jobs) tend to be more productive[1] than their properly allocated peers (of lower education). For undereducated workers, Allen and van der Velden (2001) detect no impact on the satisfaction with their current job. Nevertheless, given the restricted correlation between job satisfaction and productivity (Judge et al., 2001), those two concepts should not be treated as equal.

There are only a few analyses dealing with the direct impact of over- and undereducation on firm-level productivity instead of its impact on correlates like job satisfaction or quit rates, among others. To my best knowledge, there is only one paper that pursues this research question. In this pioneering paper based on establishment data from Belgium, translating the attained education of a firm's staff into its required, surplus and deficit components, Kampelmann and Rycx (2012) find evidence that years of overeducation entail a premium, years of undereducation (at least among young workers) a penalty in productivity for the employing firm. Although focusing on mismatch related to skills rather than to education, Jones et al. (2009) find some evidence for both under-skilled and under-skilled workers to enhance financial performance of firms for Great Britain. The only analysis for Germany focusing on direct productivity effects of educational mismatch from the firm's viewpoint is a case study by Haugrund (1990), although due to its design its external validity can be considered as limited[2]. Therefore, my analysis depicts first representative evidence for Germany using a large-scale dataset.

[1] In this case, the term „more productive" is derived from an empirically revealed superior health status, a higher possibility of participating in on-the-job training measures, and longer tenure. The findings on job satisfaction are inconclusive.

[2] Haugrund's analysis is restricted to a single company as well as to technicians and engineers.

An ex-ante assessment of the potential effects of educational mismatch on productivity strongly depends on the theoretical background as well as on the focus on either the individual itself or on cross-effects with its colleagues.

Following Human Capital Theory (Becker 1964), each educational investment enhances an individual's productivity, no matter whether the additional part is required by the job pursued. Hence, overeducated (undereducated) workers should be more (less) productive than their adequately matched colleagues of lower (higher) education, but equally productive compared to those of the same educational attainment whose job requirements match their endowments. However, the signalling hypothesis of Spence (1973) implies that investment in formal education has no direct impact on an individual's productivity. Instead, it entails an allocating function by assigning the most productive individuals to the highest-paying jobs. Therefore, although more education does not directly translate into higher productivity, a positive correlation between those two things is implied by the theory, hence resulting in the same implications for this paper as the Human Capital Theory.

A different conclusion can be derived when following explanatory approaches based on job satisfaction. As literature shows (e.g. Leuven and Oosterbeek, 2011), while overeducated (undereducated) workers tend to earn more (less) than their well-allocated colleagues of a similar job, they face wage penalties (premiums) compared to their well-matched peers of the same educational level. This under-utilization (over-utilization) of parts of their attained education is not compensated by higher (lower) wages, leaving a gap between the actual returns to education and those expected in the prior decision of investment in education. Following that line of argumentation, one would expect overeducated (undereducated) workers to have a lower (higher) job satisfaction than their properly allocated colleagues of less (more) education. On the other hand, there may be frustration effects accompanying odereducation and undereducation caused by over-challenging or under-challenging, respectively. In sum, these considerations point to a decreased job satisfaction of overeducated workers, whereas the effect for undereducation is inconclusive. While empirical research has indeed confirmed both effects, e.g. for the U.S. (e.g. Tsang et al., 1991), the UK (e.g. Battu et al., 2000) or Belgium (e.g. Verhaest and Omey, 2006), this is not true for Germany (Buechel, 2002). Since this study investigated the German case, productivity effects caused by differences between overeducated and adequately educated employees with respect to job satisfaction are not expected, at least not in the mere consideration of isolated effects on the individual level.

However, when considering the firms' viewpoint, i.e. productivity effects at the firm-level – as is the aim of this paper – one cannot simply neglect the possibility of spillover effects from the presence of educationally mismatched workers on the individual productivity of the remaining staff when assessing the effects of mismatched workers on the establishment level. To give an illustrative example: Suppose an undereducated worker is hired for the same job as you do. Acknowledging the stylized fact that the deficit education entails a wage premium compared to the well-matched colleagues, i.e. with more education (which is you), this may impose pressure on the existing staff. In your opinion, this recruitment can signal a new cost-reducing strategy of HR management in turn imposing a perceived job threat to the well-matched yet more expensive employees. You and your colleagues concerned could respond to this increased pressure by either boosting their own work effort – thus enhancing productivity – in order to reduce their dispensability or by showing defecting behaviour – particularly if they consider themselves quite dispensable. The latter, for instance, could lead to shirking or reduced cooperation with the new mismatched colleague – in line with the insider-outsider theory – or even an employer change and the associated loss of human capital for the firm. Then again, as mentioned by Battu et al. (2003), overeducated workers could improve firm-level productivity through beneficial skill spillovers.

Considering these partly opposing effects of overeducated and undereducated workers among a firm's workforce on its (firm-level) productivity, in particular with respect to the spillover effects, the pending question on the net effect for Germany remains to be answered empirically, which the following chapters will be dedicated to.

Data

In order to be able to pursue the potential impact of educationally mismatched workers among a firm' staffs on its performance/productivity, I have to gather information from different data sources, involving both administrative and survey data.

At first, for the information on the required level of education (per occupation) I use a large-scale sample from the Employment History of the Institute for Employment Research (IAB) of the (German) Federal Employment Agency (BA). The original dataset covers all employees liable to social security contributions in Germany. From a large 10 per cent sample of this basic population I draw the mode of attained education per occupation on a three-digit level, based on the classification of occupations KldB 1988. Since there is no information on the time spent in the educational system, the analysis is based on successful educational attainment in terms of obtained degrees/credentials. It should be noted that the present analysis focusses on the vertical educational mismatch, i.e. only discrepancies between attained and required education at the quantitative margin. Deviations concerning the qualitative dimension (horizontal mismatch), meaning that someone's field of education differs from the occupational field – e.g. a learned carpenter working as a haulage driver – are not taken into account. Table A1 in the appendix contains descriptive information on both the occurrence of modes of education (=required education) across highest educational degrees and the distribution of individuals from the estimation sample across (highest) educational attainment.

Subsequently, I use the cross-sectional model 2 of the linked employer-employee dataset of the IAB (LIAB) for the period for the years 2004 to 2011, which provides a link between representative survey data on firms in Germany from the IAB Establishment Panel with administrative individual data from the IAB stemming from those mentioned notifications of employers in Germany about their employees liable to social security contributions (Alda et al., 2005). By merging the information on required educational attainment (the mode of acquired education) per occupation from the first dataset to the LIAB data, I am able to calculate the shares of overeducated, undereducated and adequately educated workers among the staff of each firm covered by the IAB Establishment Panel survey, where the latter is required for the productivity assessment.

Since the IAB Establishment Panel does not contain direct information on each surveyed firm's capital stock, it has to be approximated. I follow the modified perpetual inventory approach proposed by Mueller (2008), which is supposed to give more reliable values of capital stock than to proxy by overall or replacement investment.

As is usual with establishment surveys, information about each firm's output, investment behaviour and exports is gathered for the preceding business year. To avoid an implausible assignment of events in terms of their sequence, these information have to be transferred into the previous panel year, what results in a final panel for the years 2004 to 2010. Additionally, to assure comparability between firms' business behaviour, i.e. profit maximization, the sample is adjusted by excluding establishments from the public sector and non-profit sector[3]. And last but not least, all price-based components of the analysis, i.e. sales, intermediate inputs and investment, are depreciated based on sector-specific producer price indices from the German Federal Statistical Office.

A summary of the means and standard deviations of the variables incorporated in the regressions in this paper is available in the appendix (Table A2).

[3] To be more precise, I exclude education, health and social work, arts, entertainment and recreation, non-profit organizations, and public administration.

Estimation strategy

There are several potentially detrimental econometric issues that have to be addressed when trying to obtain consistent estimates within a production function. The most important concern the existence of both time-invariant and time-variant unobserved heterogeneity across establishments as well as the dynamics in the adjustment process of firm-level productivity.

Time-invariant unobserved heterogeneity between firms may arise from management characteristics and personnel policy, only to mention a few. Following this argument, it is conceivable that the management's degree of professionalism or the time horizon of its actions as well as the (long-term, i.e. time-constant) recruitment strategies affect both productivity and the potential recruitment of workers with excess or deficit educational attainment for the respective job. I therefore apply a fixed effects estimator controlling for both firm and year fixed effects to estimate

$$ln\mathrm{VA}_{it} = \mathrm{shOE}_{it}\alpha_1 + \mathrm{shUE}_{it}\alpha_2 + \mathrm{x}'_{it}\beta + \gamma_i + \delta_t + \varepsilon_{it}$$

with $ln\mathrm{VA}_{it}$ the log value added of firm i in year t[4], firm fixed effects in γ_i and year fixed effects in δ_t. At the heart of the regression there are two variables, shOE_{it} and shUE_{it}, which convey the information on the share of overeducated and undereducated workers among a firm i's staff. There is also a vector x'_{it} of variables including potentially important characteristics which are accounted for to avoid omitted variable bias.

However, beyond the scope of unobserved time-invariant firm characteristics it is conceivable that there is time-variant unobserved heterogeneity across establishments. For instance, there is a potential correlation between the unobserved shocks in the productivity of firms and their input levels, meaning that firms who in contrast to the researcher are aware of these shocks might adjust their input factors accordingly. This would lead to biased OLS estimates of both inputs and productivity itself, and in case of time-variance to biased FE as well. This issue has first been addressed by Marschak and Andrews (1944) and is usually referred to as simultaneity bias.

So far, several solutions to this problem have been developed, which I will only deal with briefly. Fixed effects estimation can only help to expunge the simultaneity bias under the strong assumption that the part of total factor productivity (TFP) the firm is aware of before choosing its input levels is time-invariant, which is unlikely. It further entails the restriction that other time-invariant information cannot be used in the estimation.

[4] Value added is defined by sales minus intermediate inputs.

Olley and Pakes (1996) came up with a semi-parametric two-step estimator that exploits the firm's investment decision to proxy unobserved productivity shocks. This, however, requires a strictly monotonous relationship between investment and output. Hence, observations with zero investment have to be spared, which in most firm-level dataset – like the one used for the analysis of my paper – applies to a substantial share of firm-year observations, leading to a strong reduction of the sample. Therefore, Levinsohn and Petrin (2003) proposed an estimation strategy similar to the one of Olley and Pakes (1996). In contrast, their semi-parametric two-step estimator uses the information on firms' intermediate inputs instead of investment to proxy for the unobserved part of firm-specific productivity. Since this information is available on a level much more reliable than with investment, this estimator received substantial attention, especially when its rather complex implementation in statistical packages became very easy, when Petrin et al. (2004) developed a Stata command. Wooldridge (2009) refined the approach of Levinsohn and Petrin (2003) by implementing the moment conditions in a GMM framework, writing these conditions in two equations with the same dependent variable but with differing sets of instruments. The resulting one-step GMM estimator has several advantages over the two-step versions of both Olley and Pakes (1996) and Levinsohn and Petrin (2003). First, it avoids potential problems concerning identification in the first stage (Ackerberg et al., 2006), and second, it exploits the cross-equation correlation and leads to standard errors both robust with respect to serial correlation and heterskedasticity (Wooldridge, 2009).

Although covering for the simultaneity bias in an adequate manner and producing efficient results, these econometric approaches fail to account for the endogeneity of other regressors included in the productivity regression which has not already been purged by the fixed effects and which is not due to correlation with the part of the idiosyncratic error term the firm is aware before choosing inputs. It is easily conceivable that there might be both a simultaneity bias and time-varying unobserved firm characteristics both affecting the shares of over- and undereducated workers and the employing firm's productivity, causing an omitted variables bias. Regarding the wage level, for instance, on the one hand, more productive firms are believed to pay higher wages. On the other hand, research has detected wage effects for both overeducation and undereducation, respectively, implying a direct impact on the employing establishment's wage level. In order to circumvent the endogeneity issues caused by these biases some kind of instrumentation strategy is required.

To overcome these restrictions, my estimation strategy is constituted as follows: I employ a system GMM estimator, which entails two equations: Like in the cognate difference GMM estimator, the differences equation expunges the firm fixed effects

$$ln\, VA_{it} = \Delta\mathrm{shOE}_{it}\alpha_1 + \Delta\mathrm{shUE}_{it}\alpha_2 + \Delta\mathrm{x}'_{it}\beta + \Delta\delta_t + \Delta\varepsilon_{it}$$

and the endogenous (differenced) regressors are instrumented by lagged levels of these covariates, using the assumption that the latter are uncorrelated with the differenced error term $\Delta\varepsilon_{it}$ (Anderson and Hsiao, 1981 and 1982; advanced by Arellano and Bond, 1991).

Furthermore, within the second (levels) equation the levels of endogenous regressors are instrumented with lagged differences of these variables, as proposed by Arellano and Bover (1995). This introduction of additional moment conditions can improve efficiency markedly (Blundell and Bond, 1998). Moreover, this estimation approach not only allows for the inclusion of time-invariant regressors[5] – which would vanish in difference GMM –, but also for the dynamics of the production process by including the lagged dependent variable as additional regressor (Roodman, 2009a). The pursuit of dynamical completeness of the model is important as its negligence causes autocorrelated errors if the adjustment process of firm-level productivity is indeed dynamic (Wooldridge, 2010).

A further, albeit infrequently addressed econometric issue first mentioned by Wedervang (1965) concerns the potential attrition bias caused by plants dropping out of the sample during the observation period of a panel dataset. When assuming that these drop-outs are selected non-randomly, because exits are correlated to initial productivity – what has been shown both theoretically (Jovanovic, 1982) and empirically (Fariñas and Ruano, 2005) –, one should try to respond to that. Since a sole restriction to a balanced panel may result in biased estimates, Olley and Pakes (2003) implemented a correction for this attrition bias in their method. However, since within their analyses they find hardly any difference in the estimates between the correction and the simple use of an unbalanced panel. Therefore, I forego to account for this attrition bias.

In order to obtain standard errors robust to heteroskedasticity and serial correlation, a two-step estimator is applied (Roodman, 2009a). Responding to the prevalent issue of downward biased estimates of standard errors in two-step system GMM estimation, the Windmeijer (2005) correction is used.

The resulting unbalanced panel contains a full set of required information for the respective time period 2004-2010 for 23,127 observations from 6,094 establishments. Compared to OLS or Fixed Effects estimation, the number of observations depending on the chosen set of instruments is further restricted due to its identification approach which exploits lagged levels and lagged differences of the included regressors as instruments.

[5] However, since the inclusion of time-invariant variables asymptotically does not exert influence on the other coefficients (Roodman, 2009a), it would only make sense if one was interested in this time-invariant regressor's own impact on the dependent variable.

Results

Throughout this section I will assess the research question of this paper, videlicet whether overeducated and undereducated staff members affect firm-level productivity, by estimating productivity regressions exploiting the econometric procedure described in the previous chapter. The estimations will be based on the Cobb-Douglas functional form, meaning that the input factors labour and capital enter the production function linearly. A test for robustness of the results with regard to this choice does not change the conclusions of this paper[6].

In the final sample of 23,127 firm-year observations the mean values for the share of overeducated workers among a firm's workforce is 6.1 per cent, whereas for undereducated workers it is substantially larger at 10.1 per cent. The comparably low figures can be explained by the method of identification. First, educational attainment in terms of acquired degrees entails far less manifestations than using years of education. Second, the use of realized matches is known to lead to lower mismatch figures than self-assessment. Still, both the very low shares and the fact that the proportion of undereducated workers is higher than that of overeducated employees seem unusual. To some extent, these descriptive results may of course be driven by the circumstance that analyses on the individual level may deviate from those on the firm level, and as there are no comparable (firm-level) analyses for Germany, I have no obvious benchmark to compare to.

Turning to unconditional correlations, the observed combinations of share of overeducated workers and firm-level productivity are plotted in Figure 1, whereas those for undereducation are displayed in Figure 2. Hence, without recognition of potentially detrimental endogeneity issues and influential control variables there does not seem to be an obvious correlation between neither overeducation nor undereducation on the one hand and firm-level productivity on the other hand.

[6] I also applied a production function of the much more flexible trans-log form (Christensen et al., 1973) but the results remained fairly stable (cf. specification (3) of Table A4a in the appendix).

Figure 1: Scatterplot for share of overeducated workers among a firm's staff and its productivity

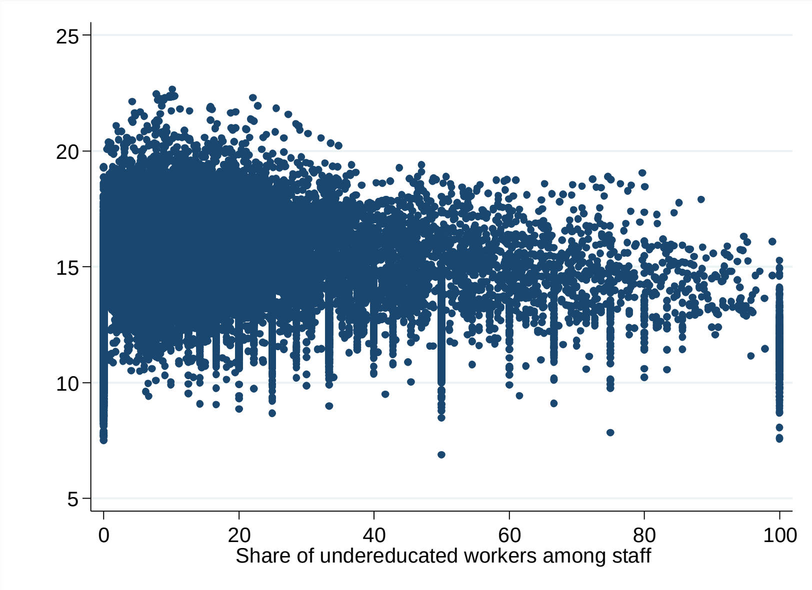
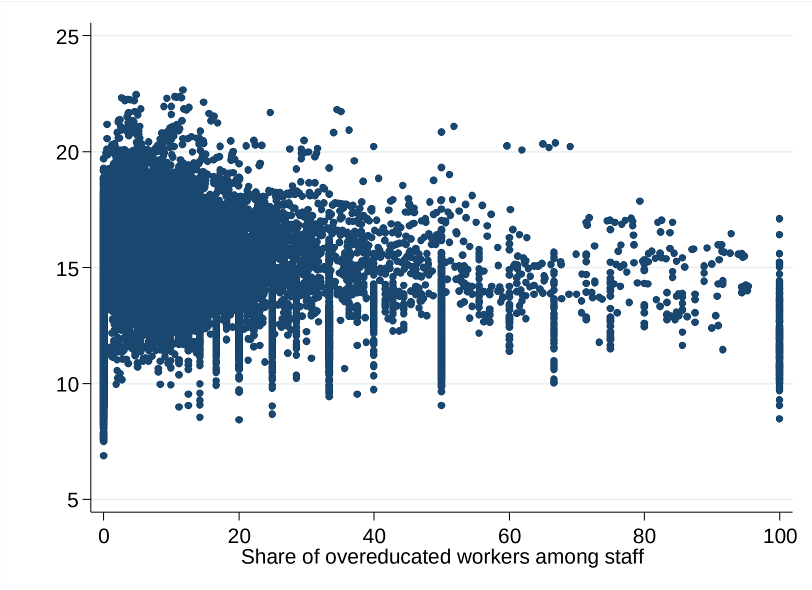


Figure 2: Scatterplot for share of undereducated workers among a firm's staff and its productivity



Approaching conditional correlations, apart from the production inputs labour and capital – and lagged value added for the system GMM estimation – I also include a set of control variables that can be considered relevant within this setting. First, the employing firm's staff structure is taken into account, because educational mismatch is supposed to be accompanied by other structural characteristics of employees. For example, overeducated workers are more likely to be female, part-time employed, and of young age (e.g. Leuven and Oosterbeek, 2011). Therefore, I include the percentages of these groups of workers in the model specification as well as those of marginal employed workers. Second, in order to avoid capturing a mere education effect in the estimates the proportion of employees with vocational and university degree, respectively, are introduced to the estimation equation. Third, literature has shown – at least for Germany – that the industrial relations affect productivity both positively (works councils, e.g. Mueller, 2011) and negatively (collective agreement, e.g. Huebler and Jirjahn, 2003). Thus, since it is conceivable that both works councils and collective agreements restrict the recruitment of overeducated and undereducated job applicants, I add two variables indicating the existence of either of these two common institutions of industrial relation. Fourth, as Jirjahn and Mueller (2014) recently unveiled the latter effect's dependence upon the ownership of the establishment, an indicator for foreign ownership is included in the model, too. Fifth, I control for single plants and incorporated firms, because these characteristics should have a substantial impact on an establishment's recruitment options and success as well as affect its productivity. Sixth, exporting firms are known to have higher productivity (e.g. Greenaway and Kneller, 2008), although this seems to a substantial extent to be due to selection (Aw et al., 2009). Therefore, these employers can afford to attract the preferred personnel through superior job characteristics, which is why a variable indicating if parts of a firm's profits are realized abroad is also introduced to the model specification. Seventh and last, it has become a stylized fact that overeducated (undereducated) workers tend to earn more (less) than their colleagues who have just the education required for the occupation (Leuven and Oosterbeek, 2011). Combined with the reasoning that more productive firms can afford higher wages and in turn can pick the best workers, it seems important to control for different wage levels across firms. This is accomplished with the inclusion of the logarithmized firm-year-specific wage median of full-time employees[7].

Accounting for time-invariant unobserved heterogeneity across establishments, the simultaneity bias, further endogenous variables among both the focal regressors and the control variables, as well as for the dynamics of the production process avoiding the "dynamic panel bias" (Nickell, 1981), a (dynamic) system GMM estimator is applied. In order to avoid endogeneity caused by overfitted instrumented variables, an issue that according to Roodman (2009b) frequently emerges with system GMM estimators but usually is not addressed properly, I reduce the set of instruments. Following Roodman, there are two ways to circumvent this instrument proliferation: First, by restricting the depth of lags to be used as instruments, and second, by collapsing the instruments matrix. The coefficients displayed below (Table 1) are therefore obtained by using a collapsed matrix of the first two lags of levels and differences used of instruments[8].

[7] The wage median is drawn from the Establishment History Panel (BHP) of the Institute of Employment Research (IAB), which depicts a full sample of establishments in Germany that at least employ one employee liable to social security contributions on the reference date (June 30th of each year). For further information see Spengler (2008).
[8] To test for sensitivity of results to this decision, I also applied system GMM estimations using extended but still restricted sets of instruments. This did not affect the estimated coefficients substantially (cf. specifications (1) and (2) of Table A4a in the appendix).

From the coefficients displayed in Table 1 the impact of educationally mismatched workers within a firm's workforce on its productivity turns out to be negative. While this effect is rather small and insignificant for overeducated workers, it turns out substantial and significant for undereducated employees. To return the precise estimates, an increase in the share of undereducated workers by one per cent impairs firm-level productivity by about 0.45 per cent. Hence, this finding is not only statistically significant but economically significant, too.

Table 1: Estimation results from dynamic system GMM regressions

	Dependent variable: Gross value added (log)[+]
Regressors	GMM-SYS
Share of overeducated workers in workforce	-0.0012
	(0.0028)
Share of undereducated workers in workforce	-0.0045**
	(0.0021)
Lagged gross value added (log)	0.4643***
	(0.0768)
Capital (log)	-0.0298
	(0.0279)
Labour (log)	0.4737***
	(0.0852)
Share of workers with vocational degree in workforce	0.0004
	(0.0017)
Share of workers with university degree in workforce	0.0029
	(0.0029)
Share of females in workforce	-0.0053*
	(0.0027)
Share of part-time workers in workforce	-0.0035*
	(0.0018)
Share of workers not older than 25 in workforce	0.0020*
	(0.0012)
Share of workers aged 50 or older in workforce	-0.0001
	(0.0007)
Share of marginal workers in workforce	-0.0022
	(0.0048)
Export business (d)	0.0679
	(0.0813)
Single plant (d)	-0.0224
	(0.0546)
Incorporated firm (d)	0.1642**
	(0.0731)
Works council (d)	-0.0106
	(0.1176)
Collective agreement (d)	-0.0544
	(0.0833)
Excellent technical state of capital stock (d)	0.0857
	(0.0959)
Good technical state of capital stock (d)	*reference category*
Fair technical state of capital stock (d)	0.0077
	(0.0690)
Poor technical state of capital stock (d)	-0.1484

	(0.2115)
Wage level (median)	0.0082***
	(0.0023)
Number of observations	23,127
Number of firms	6,094
Arellano-Bond test for AR(2)	0.446
Hansen test of overid. restrictions	0.532
Number of instruments	92

*Source: IAB Establishment Panel 2004-2010. Two-step estimator applied using the Windmeijer (2005) correction, leading to standard errors robust to heteroskedasticity and cross correlation (in parentheses). Full set of year dummies and constant included, but not displayed. Apart from the year dummies, all variables are treated as being endogenous. Significance denoted as: *** p<0.01, ** p<0.05, * p<0.1.*
[+] *in prices of 2006.*

For the system GMM estimator to be regarded as appropriate two tests have to be passed. Both the null hypotheses of the Arellano-Bond test for second-order autocorrelation in the differenced errors, which is crucial for the completeness of the model, and of the Hansen test of overidentifying restrictions cannot be rejected at fairly high p-values (bottom of Table 1). Based on these test results, the applied (dynamic) system GMM estimator is assumed to depict a valid choice.

Comparing the results gained within the present analysis with related work, especially with the paper of Kampelmann and Rycx (2012), reveals some differences. While both analyses agree on the negative impact of undereducation on firm-level productivity, the latter authors find a positive impact for overeducation, whereas my analysis unearths a negative yet small and insignificant impact. This discrepancy could either be explained by differences in the methodological approach[9] or can be ascribed to heterogeneous country specifics between Belgium and Germany.

Sensitivity analyses

Although the issue of reverse causality imposing endogeneity and biasing estimates is technically accounted for by applying the system GMM estimator – conditional on the correctness of the underlying assumptions concerning the moment conditions – one may still argue that there is a fundamental difference in expectations about the impact of miseducation on firm-level productivity with respect to the degree of voluntariness of the recruitment decision. The rationale behind this is that more productive firms are less afflicted by (skilled) labour shortages and therefore have a larger pool of (better) job applicants. Thus, if these firms *choose* to employ an educationally mismatched appli-

[9] For instance, Kampelmann and Rycx (2012) regress on years of education instead of workforce shares. Furthermore, the set of regressors varies between the two studies (e.g. concerning the capital stock). Moreover, the studies use different lengths and locations of the time period (1999-2006 vs. 2004-2010).

cant, it may be due to expected beneficial characteristics (e.g. extraordinary talent). Less productive firms, on the other hand, may *have to* employ an overeducated or undereducated individual due to a lack of alternatives, leaving them with low-performing workers. To account for this potential root for endogeneity, I re-estimated the main regression exploiting information on each firm's success with filling vacant job positions. If an establishment reports that it was not able to fill all available positions in a certain year, this is regarded as a proxy for experienced labour shortages. This information is introduced to my system GMM production estimation in several ways, both as additional regressor (in several lags, as well) and as filter variable to restrict analysis to firms who did not experience recruitment issues. Furthermore, I allowed for interaction effects between the experienced labour shortage and the recruitment of overeducated and undereducated workers, respectively. However, as Table A3 (in the appendix) demonstrates, none of these variations changed the results substantially, pointing to their robustness.

To further test for the robustness of the results, I apply several additional changes to my estimation strategy (displayed in Tables A4a and A4b): First, addressing the criticism against system GMM estimation concerning the sole reliance on internal instruments, I make use of external instruments to check if results are influenced. In order to instrument the focal variables of my analysis, the share of overeducated and undereducated workers, respectively, I choose the proportion of the respective group within each sector-year-cell (comprising of 43 distinct sectors and 7 years). While the consequently estimated coefficient for the share of overeducated workers remains (highly) insignificant, the coefficient for undereducated workers substantially gains in size (from 0.45 to 1.14 per cent) and retains its significance (confer specification (1) in Table A4a).

Furthermore, neither changing the functional form of the productivity regression from Cobb-Douglas to Trans-log (specification (2)) nor varying the sample period by restricting it to the period 2005-2010 (3) changed the results substantially. Same holds true for the expansion of the instrument count by allowing more available lags to enter the instrument matrix (specifications (4) and (5)). Even using a corrected version of the education variable[10] in the LIAB data (6) that should reduce the amount of false information yielded no different results: The signs of the focal coefficients always remained negative and the impact of undereducation retained statistical significance. The coefficients of these sensitivity analyses for the focal variables are displayed in Tables A4a and A4b.

[10] Unfortunately, I only have cross-sectional data on the individuals employed in the sampled firms. Therefore, some kind of correction of the education variable is only possible for those cases in which an individual is employed for more than one year within the same firm or across firms covered by the sample. Hence, I do not refer to the correction of Fitzenberger et al. (2005) which has been proposed for the administrative longitudinal data of the IAB.

Conclusions

From a theoretical point of view, the impact of overeducated and undereducated workers on the employing firm's productivity is ambiguous, especially when allowing for – both beneficial and detrimental – (motivational) spillover effects from mismatched employees on their colleagues' productivity. Therefore, the question about the net effect remains to be answered by empirical analysis. This study addresses this issue with linked employer-employee panel data from Germany for the period 2004-2010/11 and provides first evidence for this country. Applying a dynamic system GMM estimator, thus taking into account several potential causes for endogeneity, I show that undereducated workers among a firm's workforce impair its (firm-level) productivity. In contrast, overeducation exerts no significant impact on productivity.

Last but not least the important issue of measurement error in the reported information on workers' educational attainment, which is focal to the analysis, should be briefly still duly paid attention to. Assuming that this measurement error is random it should introduce randomly distributed uncertainty to the calculated proportions of overeducated and undereducated workers. This again would result in a flattening of the slope of the regression line, hence biasing the coefficients towards zero, meaning insignificance. Therefore, the estimates for the impact of overeducated and undereducated workers on firm-level productivity should depict lower bounds with respect to effect size.

The findings of this analysis entail several implications for research, establishments and policy makers. First, future research should try to identify the composition of the estimated net effects, hence shedding light on to which degree they are driven by the direct effect and to which by the spillover effect. Second, firms should avoid the recruitment of workers for jobs they are undereducated for, at least if possible. Considering overeducated workers, due to the insignificant coefficient firms can at least not be dissuaded from recruiting respective workers. Third and last, the detected effects for undereducation pose a severe problem for the integration of (mostly low-educated) long-term unemployed individuals into the labour market. This stresses the importance of both state funded training programmes to bridge the productivity gap and hiring subsidies as incentives for firms to hire undereducated workers nonetheless.

Literature

Ackerberg, Daniel, Kevin Caves, and Garth Frazer (2006): Structural identification of production functions. UCLA Department of Economics. *mimeo*.

Alda, Holger, Stefan Bender, and Hermann Gartner (2005): The linked employer-employee dataset created from the IAB-Establishment Panel and the process-produced data of the IAB (LIAB). *Schmollers Jahrbuch* 125, 327-336.

Anderson, Theodore W. and Cheng Hsiao (1981): Estimation of dynamic models with error components. *Journal of the American Statistical Association* 76 (375), 598-606.

Anderson, Theodore W. and Cheng Hsiao (1982): Formulation and estimation of dynamic models using panel data. *Journal of Econometrics* 18 (1), 47-82.

Arellano, Manuel and Stephen Bond (1991): Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58, 277-297.

Arellano, Manuel and Olympia Bover (1995): Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68 (1), 29-51.

Aw, Bee Y., Mark J. Roberts, and Daniel Yi Xu (2009): R&D Investment, Exporting, and Productivity Dynamics. *NBER Working Paper* No. 14670.

Battu, Harminder, Clive R. Belfield, and Peter J. Sloane (2000): How Well Can We Measure Graduate Over-Education and Its Effects? *National Institute Economic Review* 171 (1), 82-93.

Battu, Harminder, Clive R. Belfield, and Peter J. Sloane (2003): Human Capital Spillovers within the Workplace: Evidence for Great Britain. *Oxford Bulletin of Economics and Statistics* 65 (5), 575-594.

Becker, Gary S. (1964): Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education. New York: Columbia University Press.

Blundell, Richard und Stephen Bond (1998): Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87 (1), 115-143.

Buechel, Felix (2002): The effects of overeducation on productivity in Germany – the firms' viewpoint. *Economics of Education Review* 21 (3), 263-275.

Christensen, Laurits R., Dale W. Jorgenson, and Lawrence J. Lau (1973): Transcendental Logarithmic Production Frontiers. *The Review of Economics and Statistics* 55 (1), 28-45.

Dearden, Lorraine, Howard Reed, and John Van Reenen (2006): The Impact of Training on Productivity and Wages: Evidence from British Panel Data. *Oxford Bulletin of Economics and Statistics* 68 (4), 397-421.

European Commission (2013a): *Education and Training Monitor 2013*. Luxembourg: Publications Office of the European Union.

European Commission (2013b): Employment and Social Developments in Europe 2012. *Commission Staff Working Document* Volume VIII/IX.

Fariñas, Jose and Sonia Ruano (2005): Firm productivity, heterogeneity, sunk costs and market selection. *International Journal of Industrial Organization* 23 (7-8), 505-534.

Fitzenberger, Bernd, Aderonke Osikominu, and Robert Völter (2005): Imputation Rules to Improve the Education Variable in the IAB Employment Subsample. *ZEW Discussion Paper No. 05-10*. Mannheim.

Greenaway, David and Richard Kneller (2008): Exporting, productivity and agglomeration. *European Economic Review* 52 (5), 919-939.

Hartog, Joop (2000): Over-education and earnings: where are we, where should we go? *Economics of Education Review* 19 (2), 131-147.

Haugrund, Stefan (1990): Qualifikationsreserven in der industriellen Forschung und Entwicklung: Eine empirische Untersuchung zur Häufigkeit und zu ausgewählten Folgen einer arbeitsplatzbezogenen, individuellen Überqualifikation. Stuttgart: Poeschel.

Huebler, Olaf and Uwe Jirjahn (2003): Works councils and collective bargaining in Germany: the impact on productivity and wages. *Scottish Journal of Political Economy* 50 (1), 1-21.

Jirjahn, Uwe and Steffen Mueller (2014): Non-union worker representation, foreign owners, and the performance of establishments. *Oxford Economic Papers* 66 (1), 140-163.

Jones, Melanie, Richard Jones, Paul Latreille, and Peter Sloane (2009): Training, Job Satisfaction, and Workplace Performance in Britain. *LABOUR* 23 (s1), 139-175.

Jovanovic, Boyan (1982): Selection and the evolution of industry. *Econometrica* 50 (3), 649-670.

Judge, Timothy, Carl Thoresen, Joyce Bono, and Gregory Patton (2001): The job satisfaction-job performance relationship: A qualitative and quantitative review. *Psychological Bulletin* 127 (3), 376-407.

Kampelmann, Stephan and François Rycx (2012): The Impact of Educational Mismatch on Firm Productivity: Evidence from Linked Panel Data. *Economics of Education Review* 31 (6), 918-931.

Levinsohn, James and Amil Petrin (2003): Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies* 70 (2), 317-341.

Leuven, Edwin and Hessel Oosterbeek (2011): Overeducation and Mismatch in the Labour Market. Hanushek, Eric A., Stephen Machin and Ludger Woessmann (eds.): *Handbook of the Economics of Education, Volume 4*. Amsterdam: Elsevier.

Marschak, Jacob and William H. Andrews, Jr. (1944): Random Simultaneous Equations and the Theory of Production. *Econometrica* 12 (3/4), 143-205.

McGuinness, Séamus (2006): Overeducation in the Labour Market. *Journal of Economic Surveys* 20 (3), 387-418.

Mueller, Steffen (2008): Capital stock approximation using firm level panel data: a modified perpetual inventory approach. *Journal of Economics and Statistics* 228 (4), 357-371.

Mueller, Steffen (2011): Works councils and firm profits revisited. *British Journal of Industrial Relations* 49 (s1), s27-43.

Nickell, Stephen (1981): Biases in dynamic models with fixed effects. *Econometrica* 49 (6), 1417-1426.

Olley, S and A Pakes (1996): The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica* 64 (6), 1263-1298.

Petrin, Amil, Brian P. Poi, and James Levinsohn (2004): Production function estimation in Stata using inputs to control for unobservables. *The Stata Journal* 4 (2), 113-123.

Roodman, David (2009a): How to do xtabond2: an introduction to difference and system GMM in Stata. *The Stata Journal* 9 (1), 86-136.

Roodman, David (2009b): A Note on the Theme of Too Many Instruments. *Oxford Bulletin of Economics and Statistics* 71 (1), 135-158.

Spence, Michael (1973): Job Market Signaling. *The Quarterly Journal of Economics* 87 (3), 355-374.

Spengler, Anja (2008): The Establishment History Panel. *Schmollers Jahrbuch* 128 (3), 501-509.

Tsang, Mun C. (1987): The Impact of Underutilization of Education on Productivity: A Case Study of the U.S. Bell Companies. *Economics of Education Review* 6 (3), 239-254.

Tsang, Mun C., Russell W. Rumberger, and Henry M. Levin (1991): The Impact of Surplus Schooling on Worker Productivity. *Industrial Relations* 30 (2), 167-315.

Verhaest, Dieter and Eddy Omey (2006): The Impact of Overeducation and its Measurement. *Social Indicators Research* 77 (3), 419-448.

Wedervang, Frøystein (1965): Development of a Population of Industrial Firms. Oslo: Scandinavian University Books.

Windmeijer, Frank (2005): A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics* 126 (1), 26-51.

Wooldridge, Jeffrey M. (2009): On estimating firm-level production functions using proxy variables to control for unobservables. *Economic Letters* 104 (3), 112-114.

Wooldridge, Jeffrey M. (2010): Econometric Analysis of Cross-Section and Panel Data. Second Edition. Cambridge: MIT Press.

Appendix

Table A1: Required and attained education

Highest educational degree		Mode per occupation	Individuals in sample
Secondary school certificate		16	10.96 %
University entrance diploma ("Abitur")		0	2.29 %
Completed vocational training		283	70.88 %
University (of applied sciences) degree		33	15.86 %
	∑	332	99.99 %

Sources: IAB Employment History (10% sample; left column) and Linked Employer-Employee Data from the IAB (LIAB; right column). The numbers displayed are based on cross sections of both datasets for the year 2008.

Table A2: Summery Statistics of regression sample

Variables	Mean	SD
Gross value added (log)	14.09	1.98
Share of overeducated workers in workforce	6.05	12.53
Share of undereducated workers in workforce	10.15	16.82
Capital (log)	13.74	2.41
Labour (log)	3.43	1.57
Share* of workers with vocational degree in workforce	61.89	24.24
Share of workers with university degree in workforce	8.18	15.09
Share of females in workforce	32.09	25.74
Share of part-time workers in workforce	14.97	19.70
Share of workers not older than 25 in workforce	7.80	12.67
Share of workers aged 50 or older in workforce	30.41	20.76
Share of marginal workers in workforce	2.73	8.73
Exporting business (d)	0.327	0.469
Single plant (d)	0.797	0.402
Incorporated firm (d)	0.681	0.466
Works council (d)	0.276	0.447
Collective agreement (d)	0.461	0.498
Wage level (median)	69.52	27.45
Excellent technical state of capital stock (d)	0.179	0.383
Good technical state of capital stock (d)	0.515	0.500
Fair technical state of capital stock (d)	0.279	0.449
Poor technical state of capital stock (d)	0.027	0.162
Number of observations	25,458	
Number of firms	6,339	

*Source: IAB Establishment Panel 2004-2010. * All share variables range from 0 to 100.*

Table A3: Results from system GMM estimations for sensitivity analyses

Dependent variable: Gross value added (log)+				
	Sample restrictions: None		Sample restrictions: Firms without experienced labour shortages in t (5) or in both t and t-1 (6)	
Selected regressors	(2)	(3)	(4)	(5)
Share of overeducated workers in workforce	-0.0003	-0.0004	-0.0007	-0.0006
	(0.0029)	(0.0026)	(0.0027)	(0.0030)
Share of undereducated workers in workforce	-0.0048**	-0.0045**	-0.0034*	-0.0046**
	(0.0020)	(0.0020)	(0.0020)	(0.0020)
Experienced labour shortage (t)	-0.0328	-	-	-
	(0.1199)			
Interaction term 1	-	-0.0040	-	-
(overeducated*labour $shortage_t$)		(0.0046)		
Interaction term 2	-	0.0048	-	-
(undereducated*labour $shortage_t$)		(0.0060)		
Number of observations	23,022	23,022	20,347	18,817
Number of firms	6,089	6,089	5,892	5,622
Arellano-Bond test for AR(2)	0.242	0.283	0.173	0.361
Hansen test of overid. restrictions	0.431	0.328	0.702	0.500
Number of instruments	96	104	92	92

*Source: IAB Establishment Panel 2004-2010. Two-step estimator applied using the Windmeijer (2005) correction, leading to standard errors robust to heteroskedasticity and cross correlation (in parentheses). Full set of year dummies included. Significance denoted as: *** p<0.01, ** p<0.05, * p<0.1.* + *in prices of 2006.*

Table A4a: Further sensitivity analyses (1)

Dependent variable: Gross value added (log)+			
	External instruments for focal regressors	Trans-log (instead of Cobb-Douglas)	Restricted panel period (2005-2010)
Selected regressors	(1)	(2)	(3)
Share of overeducated workers in workforce	0.0023	-0.0015	-0.0017
	(0.0031)	(0.0028)	(0.0034)
Share of undereducated workers in workforce	-0.0139***	-0.0049**	-0.0047**
	(0.0053)	(0.0020)	(0.0022)
Number of observations	23,127	23,127	19,554
Number of firms	6,094	6,094	5,956
Arellano-Bond test for AR(2)	0.684	0.460	0.522
Hansen test of overid. restrictions	0.174	0.152	0.390
Number of instruments	132	104	91

*Source: IAB Establishment Panel 2004-2010. Two-step estimator applied using the Windmeijer (2005) correction, leading to standard errors robust to heteroskedasticity and cross correlation (in parentheses). Full set of year dummies included. Significance denoted as: *** p<0.01, ** p<0.05, * p<0.1.* + *in prices of 2006.*

Table A4b: Further sensitivity analyses (2)

Dependent variable: Gross value added (log)[+]			
	Higher instrument count through further lags (still collapsed)		Corrected education variable
Selected regressors	(4)	(5)	(6)
Share of overeducated workers in workforce	-0.0052*	-0.0001	-0.0015
	(0.0031)	(0.0025)	(0.0028)
Share of undereducated workers in workforce	-0.0051**	-0.0034*	-0.0049**
	(0.0024)	(0.0020)	(0.0020)
Number of observations	23,127	23,127	23,127
Number of firms	6,094	6,094	6,094
Arellano-Bond test for AR(2)	0.684	0.338	0.460
Hansen test of overid. restrictions	0.174	0.593	0.152
Number of instruments	132	92	104

*Source: IAB Establishment Panel 2004-2010. Two-step estimator applied using the Windmeijer (2005) correction, leading to standard errors robust to heteroskedasticity and cross correlation (in parentheses). Full set of year dummies included. Significance denoted as: *** p<0.01, ** p<0.05, * p<0.1. [+] in prices of 2006.*