



Quality measures in higher education: Norwegian evidence

Torberg Falch^{a,b}, Jon Marius Vaag Iversen^c, Ole Henning Nyhus^{c,*}, Bjarne Strøm^a

^a Department of Economics, Norwegian University of Science and Technology, N-7491 Trondheim, Norway

^b Department of Teacher Education, Norwegian University of Science and Technology, N-7491 Trondheim, Norway

^c NTNU Social Research, N-7491 Trondheim, Norway

ARTICLE INFO

JEL-codes:

D02

I23

I26

J24

J30

Keywords:

Higher education

Quality indicators

Value-added

ABSTRACT

We exploit rich administrative matched data for students and institutions to obtain quality measures across higher education institutions in Norway. Our primary quality indicators are based on individual income after leaving higher education within a value-added approach. Estimated quality indicators reveal significant differences in student outcomes across institutions, although the differences are much lower than raw income differences. “Old” and traditional universities appear in the upper part of the estimated quality distribution, while most of the smaller regional university colleges appear in the lower part. Students’ migration is challenging to handle appropriately, but we show that the estimated quality distribution is fairly robust to different assignments of students to institutions. Simple correlational analyses demonstrate that publicly available indicators based on subjective student assessments do not give reliable information about quality in higher education. This confirms earlier findings in the literature.

1. Introduction

The construction of quality indicators for schools and higher education institutions is a growing research area. OECD (2008) argues in favor of developing “accurate school performance measures” by using “value-added” models that account for differences in student composition across schools. The Obama administration took the initiative to publish so-called “College Scorecards” at the federal level, with information about costs, graduation rates, and earnings in colleges in the US (U.S. Department of Education, 2019). Several states in the US are developing indicator systems within higher education. Clotfelter, Ladd, Muschkin and Vigdor (2013) and Kurlaender, Carrell and Jackson (2016) are recent studies of the experience with such systems. Cunha and Miller (2014) construct and discuss the use of quality indicators for higher education institutions in Texas based on the value-added approach. We use a similar approach in the current paper, which is also inspired by the use of the value-added approach in the construction and application of teacher quality and school quality measures in compulsory education. Deming and Figlio (2016) give an overview of

the different accountability systems in US education.¹ One of their recommendations concerning higher education is that scrutiny should be positively correlated with public subsidization, indicating that measures on institutional performance should be highly relevant in, e.g., the Nordic countries.

Developing quality indicators in higher education is challenging due to several reasons. In contrast to primary and secondary education, higher education institutions differ in scope (education and research) and composition regarding study programs. Primary and secondary schools are more homogenous, with the main purpose of providing similar educational services. Obtaining quality indicators is difficult, even if the concept is narrowed to cover only educational services and defined as the extent to which education services provided increased likelihood of desired educational outcomes. Quality measures should ideally have a causal interpretation in the sense that student outcomes should reflect solely the contribution of the institutions and not systematic sorting of students with different characteristics across institutions and study programs.

The main contribution in this paper is to perform a detailed analysis

*The paper is part of a project financed by the Norwegian Ministry of Education and Research and builds partly on Strøm, Falch, Iversen and Nyhus (2016). Comments and suggestions from Jan Morten Dyrstad and a number of people in the Norwegian Ministry of Education and Research are gratefully acknowledged. We have also benefitted from many useful comments from two anonymous referees. Declarations of interest: none.

* Corresponding author.

E-mail addresses: torberg.falch@ntnu.no (T. Falch), jon.iversen@samforsk.no (J.M.V. Iversen), ole.nyhus@samforsk.no (O.H. Nyhus), bjarne.strom@ntnu.no (B. Strøm).

¹ Belfield et al. (2019) provide estimates of earnings returns to higher education institutions in England.

of quality differences among higher education institutions taking advantage of rich administrative data from Norway. We estimate quality indicators using matched data on student educational careers and subsequent labor market performance. Income measured in 2013 when most students are on average 30 years old (born between 1982 and 1985) is our primary outcome variable. We also present quality indicators based on other labor market outcomes such as unemployment and mismatch in the labor market. Such indicators may be relevant in Scandinavian countries, where labor markets are associated with strong labor unions, centralized collective bargaining, and a large share of public sector employees.

A considerable fraction of students move between institutions during their career in higher education. Yet, this fact has been given little attention in the research on returns to higher education and construction of quality indicators, as Andrews, Li and Lovenheim (2014) pointed out. Our benchmark model uses a combined approach where the students' contribution to the quality indicator is adjusted by the number of credits obtained from each institution attended. We study to what extent the estimated quality indicators using this assignment method differ from indicators obtained when assigning students to the institution where they initially enrolled or to the institution attended when leaving or graduating from higher education, traditionally used in the literature.

Most of the public discussion of quality in higher education is still based on rather crude information on traditional input measures like teacher/student ratios, expenditure per student, or variables using student assessment of quality based on survey data. It is an open question to what extent input measures or subjective student assessments are informative about quality. The relationship between long term outcomes and input measures as class size in compulsory education seems to vary a lot across countries and school systems, as exemplified by the diverging findings for Sweden and Norway in Falch, Sandsør and Strøm (2017), Fredriksson, Öckert and Oosterbeek (2013), and Leuven and Løkken (2020). Although many studies find a negative association between exam grades and class size in higher education, to our knowledge, no evidence exists as to the effect of class size in higher education on long-run outcomes.² As to subjective student assessments, Carrell and West (2010) and Braga, Paccagnella and Pellizzari (2014), using data from the US and Italy, respectively, find that student evaluations of university teachers are positively related to current outcomes (exam results). However, it is not systematically associated with long-term value-added-based measures of teacher quality. Motivated by this research, we analyze to what extent quality indicators based on input measures and self-reported student assessments at the institution level are associated with the quality indicators based on labor market outcomes.

Taking account of student sorting is one of the main challenges in estimating quality indicators. Kirkeboen, Leuven and Mogstad (2016) use variation in outcomes for students being closely below or above admission cutoffs in oversubscribed study programs to identify causal effects of study program on student outcomes. While this is undoubtedly a credible identification strategy, it is inherently local, and it is not clear how this strategy is informative about the situation for the majority of students located far away from either the cutoffs or enrolled in institutions and study programs without oversubscription. To provide a broader picture of the contribution of institutions, we choose a value-added approach as in Cunha and Miller (2014) to adjust for the fact that institutions recruit students with different abilities.

Value-added in the education literature is typically formulated as the

difference between test scores or grades at different points in time for the same student. This is not a feasible strategy in our case because exam grading practices have been shown to differ substantially across higher education institutions.³ Instead, we will use the candidates' performance in the labor market in terms of income and employment as relevant outcomes. Simply described, "value-added" indicators can be obtained in a regression framework as the effect of institution indicators, conditional on average grades from high school, and additional individual student control variables. Our value-added based indicators are constructed using individual register data from Statistics Norway and The Common Student System (FS), using income in 2013 as outcome measures for cohorts born 1982–85. Moreover, in a second step, we study the association between estimated quality indicators and institutional characteristics, including students' subjective assessment of institutions based on information obtained from the Norwegian database for higher education (DBH).

Our estimated income-based quality indicators reveal significant differences between higher education institutions, although the differences are much less than those obtained using unadjusted raw income differences. "Old" and traditional universities appear in the upper part of the quality distribution, while most smaller regional university colleges appear in the lower part. We also find that the estimated quality distribution is robust to the method used to assign students to institutions. Simple correlational analyses demonstrate that the estimated income-based quality indicators are not systematically associated with indicators based on subjective student assessments. However, student effort measured by self-reported hours of studying per week appears to be positively associated with the income-based quality indicator.

The rest of the paper is organized as follows: Institutional background is presented in Section 2, while Section 3 presents the data and the empirical framework. Constructions of quality measures are presented in Section 4. The correlations between our constructed quality indicators and publicly available characteristics of higher education institutions, as well as survey-based measures of student satisfaction, are shown in Section 5. Section 6 offers some concluding remarks.

2. Institutional background

The Ministry of Education and Research has the overall responsibility for higher education in Norway. Higher education is offered by three types of higher education institutions: Universities, scientific colleges, and university colleges. Since 2003 Norway has been following the objectives of the Bologna Process in the European higher education and has implemented a 3 + 2 + 3-degree system with a Bachelor's, Master's, and Ph.D. structure following the European standards.

The Norwegian Agency for Quality Assurance in Education (NOKUT) is an autonomous governmental agency that provides external supervision and control of the quality of Norwegian higher education. NOKUT accredits new study programs, controls the existing ones, and is also responsible for "Studiebarometeret" – the National Student Survey.⁴ The National Student Survey aims to strengthen the quality of work in higher education and give useful information about educational quality.

All Norwegian higher education institutions use a system of credits for measuring study activities considered equivalent to the European Credit Transfer and Accumulation System (ECTS). 60 ECTS credits are allocated to the workload of a full year of academic study, equivalent to 1500–1800 hours of study. 30 ECTS credits are normally allocated to

² The literature is reviewed in Bandiera et al. (2010) who find a significant negative relationship between exam grades and class size in UK universities and that the class size effect is highly non-linear across the class sizes observed. De Paola et al. (2013) find that test performance is negatively associated with class size in math, but unrelated to class size in language using data from an Italian university.

³ Møen and Tjelta (2010) reach this conclusion using information from a limited number of institutions offering study programs in economics and business administration in Norway. Strøm et al. (2013) reach the same conclusion using a much larger sample of Norwegian students and study programs.

⁴ Information about the survey can be obtained at <https://www.nokut.no/en/studiebarometeret/studiebarometeret/>

Table 1

Descriptive statistics on student background, the field of study, and graduation.

	Mean	Std. dev.
Born in 1982	0.258	0.437
Born in 1983	0.261	0.439
Born in 1984	0.251	0.434
Born in 1985	0.230	0.421
Female	0.603	0.489
First generation immigrant	0.027	0.161
Second generation immigrant	0.012	0.109
Parental edu. = MSc/PhD	0.169	0.374
Parental edu. = BSc	0.369	0.482
Parental edu. = High school	0.409	0.492
Average high school grades	4.117	0.649
(Average high school grades) ²	17.368	5.359
(Average high school grades) ³	74.957	34.204
Field of study (share of credits):		
<i>Humanities and arts</i>	0.137	0.270
<i>Teacher training and pedagogy</i>	0.146	0.315
<i>Social sciences and law</i>	0.157	0.301
<i>Business and administration</i>	0.122	0.295
<i>Natural sciences, vocational and technological subjects</i>	0.188	0.351
<i>Health, welfare, and sport</i>	0.234	0.391
<i>Other fields</i>	0.016	0.092

Note: $N = 53,996$.

one semester's full-time study. The academic year usually lasts for ten months and runs from August to June.

Completion of secondary education in a program for general studies, equivalent to passing the exam at the end of Norwegian secondary school (high school), is the general basic requirement for entry to Norwegian higher education institutions. Some study programs have special admission requirements, usually related to specialist subjects or fields of study from high school. The Norwegian Universities and Colleges Admission Service ("Samordna Opptak") coordinates the admission to ordinary undergraduate study programs at all universities, scientific colleges, university colleges, and some private university colleges in Norway. Details of the centralized admission process and the allocation of students to institutions and study programs are described in Kirkebøen et al. (2016) and Dyrstad, Sohlman and Teigen (2021).

This paper includes data from all state-owned higher education institutions. They are established at different times but with a specific composition of study fields and study programs. The different profiles of the institutions are historically a political choice. However, the institutions can change the size of admission in different study fields and study programs. They typically respond to trends in demand for different studies, but they have the power to steer their own profile to some degree.

3. Data and empirical specification

We exploit a rich individual data set based on Norwegian administrative registers to develop value-added quality indicators. This section describes the data and the specification of the regression model to obtain quality indicators.

3.1. Data

Individual data used in the paper are taken from different administrative registers and sources and merged by Statistics Norway using a unique personal identifier.

The Common Student System (*Felles studentsystem*, FS) provides detailed individual data on exams and student careers through higher education. This is a study administration system developed for universities, scientific colleges, and university colleges. The data source contains all exam results, including the number of credits and field of study, and identifies at which institution and point in time (year and semester) the exam was taken.

Table 2

Descriptive statistics on the outcome variables in 2013.

	Mean	Std. dev.
Income (NOK)	474,308	197,992
<i>Humanities and arts</i>	389,029	164,552
<i>Teacher training and pedagogy</i>	397,714	120,948
<i>Social sciences and law</i>	473,034	179,460
<i>Business and administration</i>	533,509	233,919
<i>Natural sciences, vocational and technological subjects</i>	583,236	212,663
<i>Health, welfare, and sport</i>	444,657	179,703
<i>Other fields</i>	490,197	206,291
Log (income)	12.969	0.534
Full-time employment in a reference week	0.811	0.392
Working hours in a reference week	32.702	10.635
Registered as unemployed in 2013	0.238	0.426
Number of days unemployed in 2013	30.161	75.196

Note: The descriptive statistics on income categorized by field of study are calculated based on a classification of individuals to the field of study with the most credits. 1.38 percent of the observations are omitted due to an equal number of credits at two or more fields.

Registers administered by Statistics Norway provide data on labor market outcomes, such as income, employment, working hours, occupation, and industry affiliation. It also includes individual background data on parental education and immigration background and data on degree completion and progression through the education system.

Data on institutional level variables such as resources and the number of applicants per enrolled student is taken from *The database for statistics on higher education* (DBH). DBH also provided access to the National student survey data administered by NOKUT and includes information on student satisfaction, motivation, effort, and students' subjective evaluation of the learning environment and relevance of instruction at the institution level.

We base our analyses on students that have attended universities, scientific colleges, and university colleges. The institutions included and the number of credits by field of study are shown in Appendix Table 1. We exploit four cohorts of students that are born 1982–1985. One reason for this is that individual information on grades from high school is only available from 2001. Also, in recent years, there has been considerable consolidation in the Norwegian college and university sector, resulting in many mergers. Another reason to focus on the cohorts born 1982–1985 is that most of them graduated prior to these mergers. Descriptive statistics on the students are provided in Table 1, while Table 2 presents descriptive statistics on the outcome variables.

The share of students born in 1982–83 is slightly higher than the share of students born in 1984–85. This is natural since more students in the latter cohorts are still in higher education and therefore excluded from the sample. Females amount to about 60% of the observations. This is consistent with official data on students in higher education.⁵ The share of first- and second-generation immigrants consists of only 2.7 and 1.2%, respectively. At face value, this indicates an underrepresentation of immigrants in higher education relative to their share of the total population. However, sample reductions partly cause the low share due to omitted information on other covariates. The variables describing parental education are based on the classification for the parent with the highest recorded education. About 17% of the individuals have at least one parent with MSc/Ph.D. or equivalent education. The share with parental education equal to BSc and high school amounts to 37 and 41%, respectively. The average grade from high school is 4.1, with a standard deviation of 0.65.

The most populated fields of study are "natural sciences, vocational and technological subjects" and "health, welfare, and sport". The share of credits in "humanities and art" is somewhat high because courses in

⁵ See <https://www.ssb.no/en/utdanning/statistikker/utuvh> for details on Norwegian students in higher education.

philosophy of science are compulsory at most university programs. We have chosen to classify credits within the study fields primary industries, transportation and communication, general, and unspecified as *other fields* since credits in these fields account for only 1.6% of all credits. Table 2 shows that there are quite substantial differences in mean income across different fields, varying from about NOK 390,000 for humanities and arts to about 583,000 for students who have attained mainly natural sciences. These relative differences also apply to other labor market outcomes, where attending business and administration, natural sciences, and transport and communication signals quite desired labor market outcomes.

Our data include outcome variables measured up to 2013 when the individuals in the sample were 28–31 years of age. One advantage of using labor market outcomes at a relatively young age is that it is a limited time gap between when actual institutional quality is spelled out and when the outcome is measured, which in our case is about 5–10 years. One disadvantage is that income trajectories are steeper earlier in the career than later. Therefore, income at a young age is a less reliable measure of lifetime income than income at a later age in longitudinal data. Current earnings are biased measures of lifetime earnings (Bhuller, Mogstad & Salvanes, 2017; Haider & Solon, 2006). Because of data limitations, however, income at a young age is often used in studies related to education. For example, Chetty, Friedman and Rockoff (2014) use income at age 28 in their analyses of teacher value-added.

In the sample, 81.1% were full-time employed in the reference week in the fall of 2013, whereas the mean number of working hours this week was 32.7. Unemployment was registered at least once during 2013 for 23.8% of the sample, while the mean number of days unemployed was 30.2.

3.2. Empirical strategy to obtain quality indicators

3.2.1. Quality measured by individual income

The value-added approach is a well-known approach used to construct institutional quality or teacher quality indicators based on individual student changes in achievement on tests or exams taken at different points in time for students exposed to the same institution or teachers. Koedel, Mihaly and Rockoff (2015) contain a review of the literature and an extensive evaluation of applications of the approach in compulsory and secondary schooling and teacher effectiveness studies. In principle, this approach can also be applied to postsecondary education and higher education institutions. However, to do so, we need objective measures of students' performance before and after being enrolled in a higher education institution. There are several reasons why school performance measures are more difficult to obtain in higher education than compulsory education. Exam results in higher education are a possible measure of knowledge acquisition. However, the literature suggests that grading practices vary systematically across courses and institutions and are likely to be used strategically to recruit students. De Paola (2011) uses data from an Italian university and shows that the tendency to inflate grades is higher for those degree courses which obtain a number of applications that are lower than the number of places they offer. For Norway, Møen and Tjelta (2010) and Strøm, Falch, Gunnes and Haraldsvik (2013) document substantial differences in grading practices between Norwegian higher education institutions.

In addition, heterogeneity across institutions is a challenge. In primary and secondary education, schools offer more or less the same education independent of location. Institutions have substantial freedom to choose and allocate resources to different disciplines and study programs in higher education. Thus, both systematic variation in grading practices and heterogeneity in subjects and courses across institutions suggest that the value-added approach to quality indicator construction needs to be re-formulated when applied to the higher education sector.

We argue that a better measure for valuing skills and knowledge obtained in higher education is individual labor market outcomes like income and employment after finishing the education program. Suppose

we were to follow the value-added approach as used in the education literature in a literal sense. In that case, we should also include outcome levels, i.e., income before enrollment in higher education. However, this would be misleading since most individuals were students with no income and labor market attachment before entering higher education. Instead, we include average grades from high school to measure productivity and income potential before enrolment in a higher education program. Quality differences between two higher education institutions that offer the same program can then be measured by the students' wages in the labor market when we control for initial skills measured by grades from high school. This approach is similar to that used in Cunha and Miller (2014). While wages would be a more satisfactory output measure, most studies, including ours, only have access to income or earnings information from administrative registers and thus represent a combination of wages and labor supply decisions. With these caveats in mind, we formulate the following model as a benchmark for our analyses:

$$Y_i = \alpha + \sum_{s=1}^N \beta_s E_{is} + \gamma_1 f(GPA_i) + \delta_j + \gamma_2 X_i + \sum_{p=1}^M \gamma_3 D_{ip} + \theta_c + \tau_r + \varepsilon_i, \quad (1)$$

Y_i is the logarithm of income for student i in 2013.⁶ E_{is} is the number of credits of individual i at institution s normalized to years of study (60 credits). The β_s -vector is our value-added measures and are the parameters of main interest. The model is estimated without restrictions, but for presentation purposes, the figures below normalize the β_s -vector such that $\sum_{s=1}^N \beta_s = 0$ as is common in the value-added literature. The results presented in the figure can thus be interpreted as approximately the percentage income differences between students attending different institutions.

$f(GPA_i)$ is a cubic function of student i 's average grade from high school. The cubic function is chosen over both first, second, and fourth-order polynomial because of its fit in explaining income conditional on all covariates. The choice of functional form for GPA turns out to have minimal impact on the estimated value-added measures. The model includes fixed high school effects (δ_j) that account for differences in grading practices between high schools (subscript j) as well as time-constant differences in unobservables across students growing up in different parts of the country.⁷

The X -vector includes controls for gender, immigration status, and parental education. We have included two dummy variables capturing first- and second-generation immigrants, respectively, whereas we include dummy variables for *i*) upper secondary education, *ii*) short higher education (BSc), and *iii*) long higher education (MSc) regarding highest parental education. Hence, unknown and lower secondary parental education is the reference category.

D_{ip} is a vector controlling for the share of credits obtained from different fields of study. The share of social science credits is the reference category.⁸ Ideally, the control variables should be measured *before* exposure to higher education. However, with large wage and income differences across programs and educational length, controls for field of study improve on interpretation. θ_c is fixed birth year (cohort) effects, τ_r is labor market region fixed effects, and ε_i is the error term. Controlling for labor market region fixed effects may account for potential peer-effects or that some institutions prepare their students for particular regional labor markets. However, since sorting into different labor

⁶ We use the pension-qualifying income as reported in the tax registry. This income measure is not top coded and includes labor income, taxable sick benefits, unemployment benefits, parental leave payments, and pensions, see Black et al. (2013, p. 132).

⁷ The individuals are assigned to the high school where they attended most courses.

⁸ See Table 2 for information on the classification of study fields.

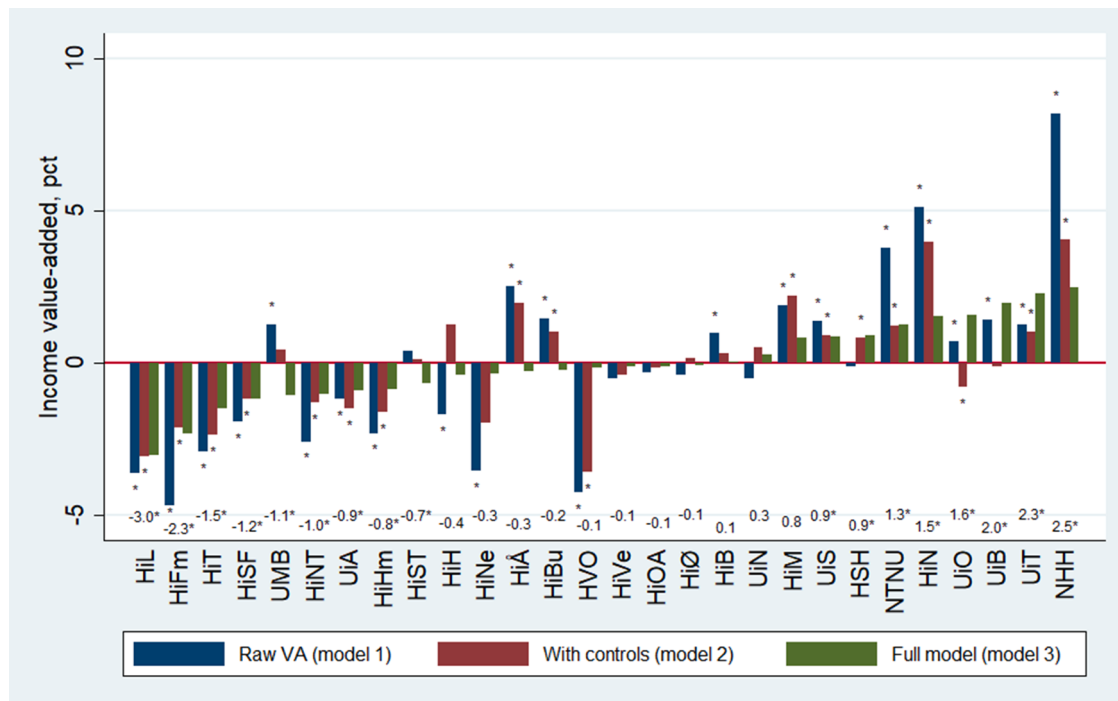


Fig. 1. Estimated quality indicators for Norwegian higher education institutions. Note: The bars height are defined as the point estimate in Appendix Table A2 (models 1, 2, and 3, respectively) subtracted the mean value. * indicates that the normalized value-added indicator is significantly different from 0 in the figure at the 10% level. The normalized income value-added estimates for model (3) (green bars) are provided in the lower part of the figure, just above the abbreviation for the institutions.

markets could be an endogenous outcome, we have chosen to include this term only in an extended version of our baseline model to address the sensitivity of our main results.

In our baseline specification, we allow credits from all institutions to explain student i 's income in 2013. In earlier studies, the E_{is} -vector is either dummy variables for “startup-institutions” or “exit-institutions”. We discuss the consequences regarding the different assignments of students to institutions in Section 4.3.

Whether a single value-added estimate or indicator is representative for all fields within an institution is a question. To allow for heterogeneity in quality within institutions, we expand the value-added approach in Eq. (1) by specifying a model that includes credits (normalized study year) that are specific to every study field at each institution (E_{isp}).⁹

$$Y_i = \alpha + \sum_{s=1}^P \sum_{p=1}^N \beta_{sp} E_{isp} + \gamma_1 f(GPA_i) + \gamma_2 X_i + \sum_{p=1}^M \gamma_3 D_{ip} + \delta_j + \theta_c + \varepsilon_i, \quad (2)$$

One of the purposes for constructing the value-added estimates is to correlate them with publicly available information about the institutions. Field-specific institutional information does not exist for all publicly available variables, and for some fields, the profile of many institutions implies that the number of students is insufficient to obtain reliable quality indicators at the field level. Therefore, we treat the empirical specification in Eq. (1) as our main approach. The estimated field-specific institution quality indicators obtained by Eq. (2) supplement the main results.

⁹ The study fields *Primary industries, Transport and communications, safety and security and other services, General or unspecified field of study* is treated as one study field because of the relative low number of students at these programs. Credits in these fields are included in the analyses, although we only present the estimates for the other six fields.

3.2.2. Alternative specifications

Our benchmark model use income as the dependent variable in Eq. (1). The argument for using this outcome is the traditional one that differences in individual wages for similar higher education programs reflect differences in individual productivity. Higher wages reflect high individual productivity, and productivity is increasing in skills and knowledge obtained through higher education (Cunha & Miller, 2014). Earnings or income-based measures of school quality have long traditions in research in education economics. Card and Krueger (1992) provide an early examination of the relationship between income and school quality in the US. Income is also used in several Norwegian studies. See, for example, Black, Devereux and Salvanes (2013) for a study on peer effects and Havnes and Mogstad (2015) for an analysis of the kindergarten expansion. Other studies using Norwegian data, such as Falch et al. (2017), use income as an outcome in a study of class size effects in compulsory education, while Kirkebøen et al. (2016) estimate the returns to higher education across study programs using income as the outcome variable.

One may argue that in Norway and other countries with high union coverage, centralized collective wage bargaining systems, and a large share of public sector employees, income measures are only weakly related to individual skills and productivity. We, therefore, extend the analysis in Eq. (1) by comparing institutional quality estimates based on income with quality estimates using other labor market outcomes as employment or unemployment incidence, hours worked, and labor market mismatch.

A critical challenge in all estimation of quality indicators is student selection. While the value-added approach extended with a battery of individual and regional controls may reduce the problem, we cannot trust that it is eliminated. Deming (2014) exploits randomization of students across schools to test the validity of value-added models to obtain causal school effects. He concludes that value-added-based school effects are unbiased predictors of actual achievement and thus well suited to measure school effectiveness.

To evaluate different estimation methods, Guarino, Reckase and

Wooldridge (2015) conduct a simulation exercise to evaluate different approaches to estimate teacher effectiveness. While their analysis reveals that no method accurately captures true teacher effects, they conclude that value-added indicators based on a student-level dynamic specification using a single lag of student achievement were the most robust estimator. Guarino, Stacy and Wooldridge (2019) use both simulated and administrative data from Georgia to compare estimates of school effectiveness based on the value-added approach with estimates from a “beating the odds” type approach. The latter compares actual school performance with predicted outcomes from a regression at the school level controlling for observable school characteristics, including student composition and enrollment size measures. They find that the value-added method provides the most credible measures of school effectiveness in terms of being less subject to bias due to student sorting across schools.

While this evidence seems to support the use of value-added-based quality indicators, a limitation is that all income or labor market-based quality indicators result from several factors that are hard or impossible to identify. Estimated quality differences may reflect differences across institutions in the general culture to encourage students to aim for high-paid jobs. Thus, one may argue that income or labor market-based quality indicators will be biased measures of the educational quality of the institutions in a narrow sense. This should be kept in mind when judging empirical results.

4. Empirical results

This section first reports estimated institutional quality indicators using the value-added approach in Eq. (1). Our benchmark model uses income as the outcome measure in Section 4.1. In 4.2, we present the relationship between indicators based on income with indicators based on alternative labor market outcomes like employment, unemployment, and labor market mismatch. In Section 4.3, we compare the income-based indicators using different strategies to assign students to institutions.

4.1. Income-based quality indicators

This section reports results from estimating value-added-based indicators of institutional quality as described in Section 3.2 and formally presented in Eq. (1). We use individual income measured when most students are about 30 years old as the outcome variable. The income-based indicators are presented in Fig. 1. To illustrate the implications of model specification, Fig. 1 includes results from three different specifications. Appendix Table A2 presents the full model results. The bars in Fig. 1 refer to $100 \times$ estimated coefficient to each institution and can approximately be interpreted as the percentage income difference between students attending an institution relative to the average institution.

The blue bars in the figure are a raw indicator without any controls, except for the year of birth (cohort fixed effects). Including controls in the model reduces the differences across institutions. The second model in Fig. 1 (red bars) conditions on individual characteristics as well as

fixed high school effects, while the third model (green bars) in addition condition on the field of study (see Appendix Table 2, column (2) and (3), respectively).

Not surprisingly, the estimated “quality” differences shrink substantially when control variables are included. As an example, the raw income premium for students from the Norwegian School of Economics (NHH) of 8.2% is reduced to 4.0% when controls are included and further to 2.5% when controls for fields of study are included. It also appears that the four oldest universities (University of Oslo, Norwegian University of Science and Technology (NTNU), University of Bergen, and University of Tromsø), together with NHH, all appear among the institutions with the highest estimated quality in all specifications.¹⁰ On the other end of the spectrum are some smaller regional university colleges.

GPA from high school is the individual control variable with the largest effect on the quality variable. There seems to be important sorting across institutions related to ability. Fig. 1 also shows that the estimated institutional differences depend on the inclusion of study field controls in the model. In particular, the quality indicators for regional colleges historically established with a profile towards professions in the public sector, such as teachers and nurses, improve when the model condition on study field. Nevertheless, the overall conclusion is that the “old” and traditional universities appear in the upper part of the quality distribution while small university colleges are in the lower part. This is largely independent of whether study field controls are included or not.

While Fig. 1 presents the main results based on estimation of (1), we have also estimated institutional quality indicators when quality is allowed to vary between study fields within institutions. Appendix Table 3 reports the results from estimated versions of Eq. (2), and there appear to be some differences between study fields. Due to different historical profiles, some institutions have an insufficient number of students to estimate reliable quality differences in this specification. This must be taken into account when judging the results.¹¹ The correlation with the overall quality indicator is highest for studies in social sciences, business, and health studies (the correlation coefficient is around 0.60 for all three cases) and lowest for studies in humanities & arts (the correlation coefficient is -0.09).¹² Overall, the results, the broad conclusion that the old and traditional universities appear in the upper part of the quality distribution still holds, except the study field Humanities & Arts. Humanities & Arts is the study field with clearly the lowest return in the labor market, see Appendix Table 1, but it is outside the scope of the present paper to analyze why this study field seems to differ from the others in terms of quality variation across institutions.

Another approach to investigate the extent to which the results depend on the institutions’ profile is to exclude public sector employees from the model. The public sector is heavily unionized, and there is little

¹⁰ One exception might be the Norwegian University of Life Sciences (UMB). UMB has a long tradition as an agricultural university college, and has similarities with NHH in the sense that they historically have a national profile with national responsibilities, in contrast to the more recent regional colleges. The main study programs at UMB are officially classified within the field “Natural sciences, vocational and technical subjects”, covering 60% of the students (see Appendix Table 1). This field has a high payoff in the labor market (see Appendix Table 2, column 3 and 4), which explain the drop in the income premium in the model conditioning on field of study. The estimated income premium is unaffected by excluding the few students registered in Humanities & Arts at UMB.

¹¹ A formal test reject at conventional levels that equation (1) is a valid simplification of equation (2), indicating that there are significant quality differences across study fields within institutions.

¹² In addition, there are two large outliers in the estimates for the study field Humanities & Arts, which both are related to few observations at the relevant institution (see Appendix Table 1). They are not, however, the reason for no correlation with the overall quality indicator in Figure 1.

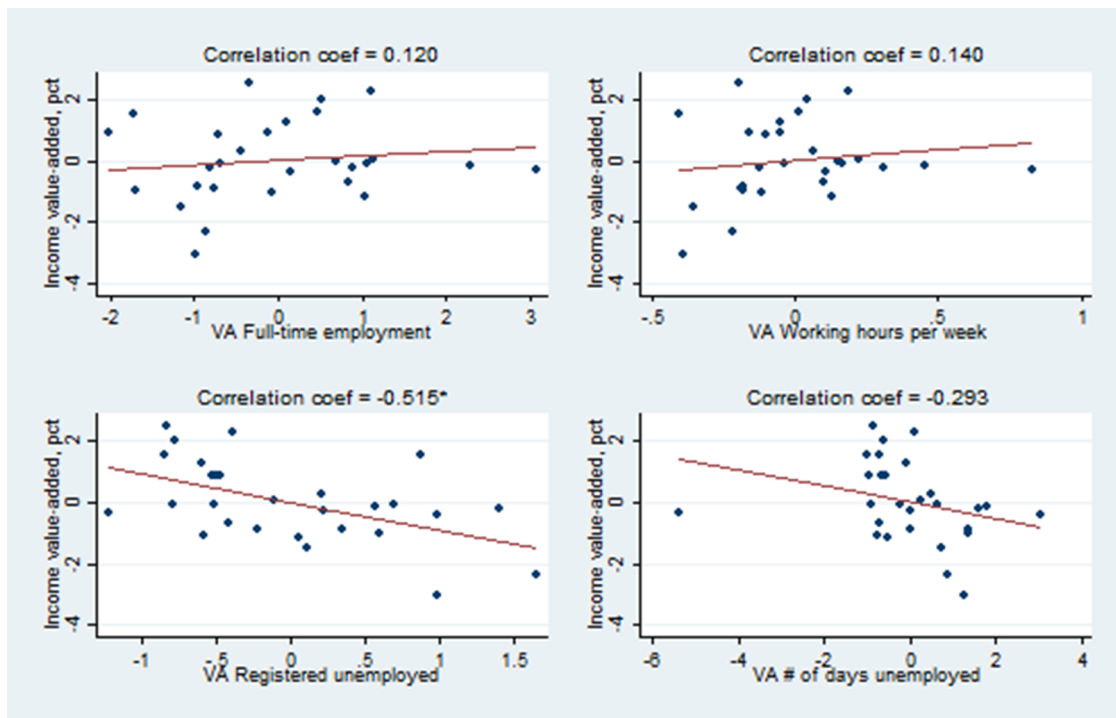


Fig. 2. Relationships between the quality indicator from the benchmark model and quality indicators using models based on employment and unemployment status
* indicates significant correlation at the 10% level.

income variation within occupations. This exercise reduces the sample by 21,935 observations but does not alter the quality indicators based on the full model much. There is a small increase in estimated institutional quality for university colleges in cities with relatively high income in the private sector (Stavanger and Bergen) and reduced estimated institutional quality for some small university colleges.¹³ As further evidence on this issue, column (4) in [Appendix Table 2](#) reports the estimated quality effects when we control for labor market region fixed effects to account for potential peer-effects or that some institutions prepare their students for particular regional labor markets. It appears that most of the estimated quality effects are unchanged. However, the effect for an institution located in Stavanger (UIS) is reduced compared to the main specification, presumably because this area is the center of the high-wage petroleum industry in Norway.

Since the sample only includes four cohorts, and the outcome variable is measured in the single year 2013, there are limited possibilities to investigate the impact of institutions on lifetime income and the age-income profile. We have estimated the model separately for the two youngest and oldest cohorts (28–29 and 30–31 years of age when the outcome is measured, respectively). The findings indicate that the quality indicator is larger for the oldest cohorts for NHH, indicating a steeper age-income profile within business education than for other educations. At the same time, the results are stable for the traditional universities. For the regional university colleges, the results are mixed. Some institutions obtain higher value-added for the oldest cohorts, while others obtain higher value-added for the youngest cohorts.

4.2. Indicators based on alternative labor market-based outcomes

A specific feature of the labor market in Norway and other Scandinavian countries is the important role of centralized collective wage bargaining institutions and the large share of workers employed in

public sector institutions with limited flexibility in wage setting. Thus, in this section, we investigate to what extent using other labor market outcomes in the value-added approach generates comparable results to the income measure. We consider three different alternative outcomes: Employment, unemployment, and a measure of labor market mismatch.

4.2.1. Employment and unemployment

We use two variables to measure employee performance in our sample of students. The first is a dummy variable taking the value 1 if the student is employed in a full-time position in the reference week in 2013 and 0 otherwise.¹⁴ Our second measure is the number of hours worked in the reference week.

We also use two variables to measure the incidence of unemployment. The first variable is a dummy that takes the value 1 if the student was registered as unemployed at some point in 2013 and 0 otherwise. The second variable measures the number of days being unemployed in 2013.

We estimated the value-added model in [Eq. \(1\)](#), replacing log income with these alternative outcomes. To save space and to concentrate on the comparison of estimated quality indicators from the benchmark model with indicators based on alternative outcomes, [Fig. 2](#) shows scatter plots of the estimated quality indicators based on alternative outcomes and the benchmark model. We expect a positive association between the quality indicators when employment is the outcome, while we expect a negative association when using unemployment as an outcome. [Fig. 2](#) shows that this is the case. Still, the relationship between the indicators is weak, with correlation coefficients of 0.12 and 0.14 for the employment-based indicators and -0.52 and -0.29 for the unemployment-based indicators. One reason for the moderate relationships is that the variation in the employment and unemployment measures is low: most of the students have full-time positions, the variation in hours worked is low, and the incidence of unemployment is low. A specific problem with the employment variables is that they refer

¹³ The results from this regression, and other regressions only reported in the text or in figures, are available from the authors on request.

¹⁴ The reference week is the third week in November.

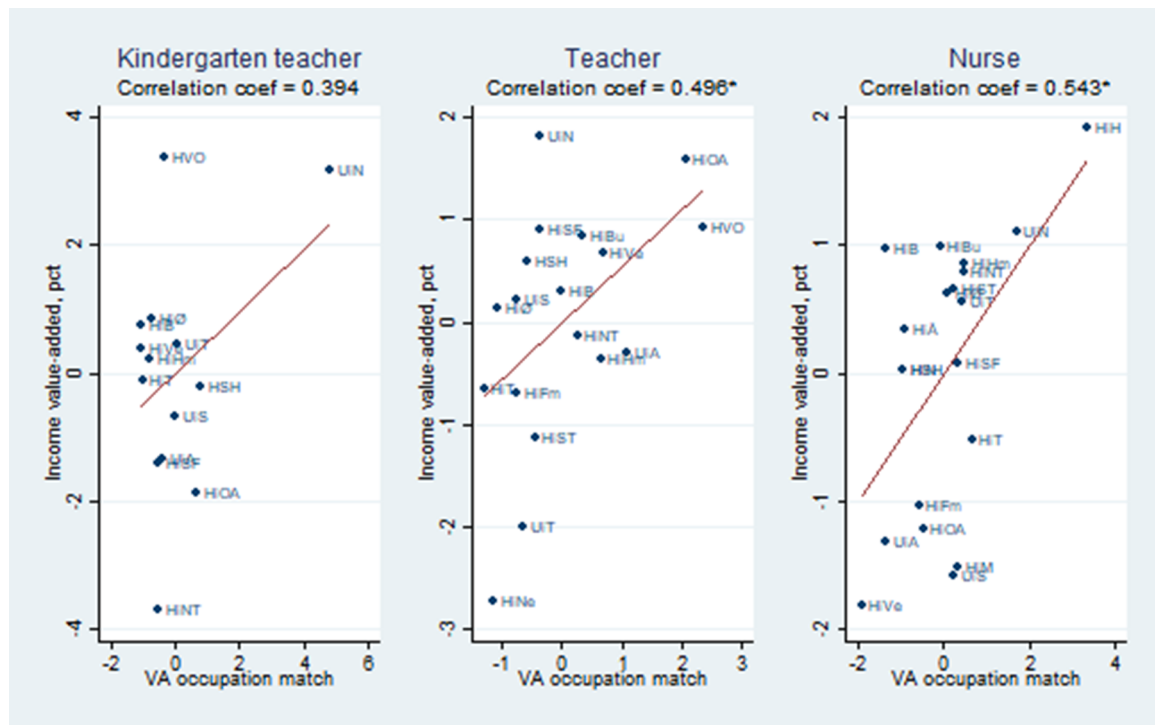


Fig. 3. Relationships between the quality indicator from the benchmark model and quality indicators using models based on labor market mismatch
* indicates significant correlation at the 10% level. N equals 1909, 2601, and 3592, respectively.

to one specific week in the year and thus may include substantial measurement errors.

4.2.2. Labor market mismatch

It is traditionally expected that highly skilled individuals are more likely to obtain a job that corresponds to his/her educational background. For example, highly skilled nursing students are more likely to obtain a nurse job than low-skilled students with the same education. In educations where students acquire skills that are relevant in narrowly defined occupations, this is likely to be the case. On the other hand, this is less relevant for educations where students acquire skills relevant in a broad range of occupations. This might be highly applicable for some professions in higher-paid occupations outside the professions, like administration and management. Thus, we do not expect a very high correlation between the mismatch and value-added analyses.

In order to measure labor market mismatch, we use three subsamples of students with narrowly defined education; students with education as kindergarten teachers, students with ordinary teacher education, and nurse education. A common feature of these groups is that the main employers are public sector or non-profit private institutions (kindergartens, schools, elderly care, and hospitals), with limited pay flexibility. We then use register-based information on the extent to which students with these specific educations are employed in occupations corresponding to their education. Thus, we measure if students with teacher education are registered as employed in a teacher occupation category in 2013 or if students with nurse education are registered in a nurse occupation category in 2013.

Thus, for each separate subsample of students, we estimated Eq. (1) with income replaced with a dummy equal to 1 if the individual is employed in an occupation relevant for the education and 0 if the occupation is irrelevant or if the individual is unemployed. In Eq. (1), all credits from higher education (E_{ig}) are assigned to the institution where each individual obtained the relevant education.

Fig. 3 shows the cross-plots of the estimated quality indicators based on the occupation match outcome and the benchmark indicators using income as the outcome variable. The analyses are based on similar

samples. Since these narrowly defined groups of students are enrolled in subsamples of institutions, the number of included institutions is lower than the benchmark model. Fig. 3 shows that institutional quality indicators based on occupation match and income are positively associated, although the correlation is not statistically significant for kindergarten teachers.

Taken literally, the relatively weak associations between indicators based on the alternative labor market outcomes and the benchmark model may suggest that income does not account for important aspects of institutional quality. However, in our view, this is a premature conclusion. The variation in outcomes like employment and unemployment incidence is clearly limited for this group of employees since most of them are employed in full-time positions. Even though wage variation is limited by centralized collective bargaining, in many cases, employees with similar educational backgrounds may end up in many different jobs with different wage levels. Further, to keep the analysis in line with most of the international literature, we use quality indicators based on income in the rest of the paper.

4.3. Assignment of students to institutions

As discussed in the introduction, a tricky question is how to deal with the fact that a substantial fraction of students move across institutions during their career in higher education.¹⁵ Since we do not know why students change institutions, it is difficult to predict how this will bias the results. In our benchmark model, the student's contribution to the institutions' quality indicator is the number of credits the particular student has obtained from the particular institution attended. However, our rich administrative register data makes it possible to assign students based on other approaches often used in the literature and compare the results from alternative assignment methods to those obtained from the benchmark model.

¹⁵ In our sample, the share of students finishing their higher education career at the same institution as they initially enrolled is around 70%.

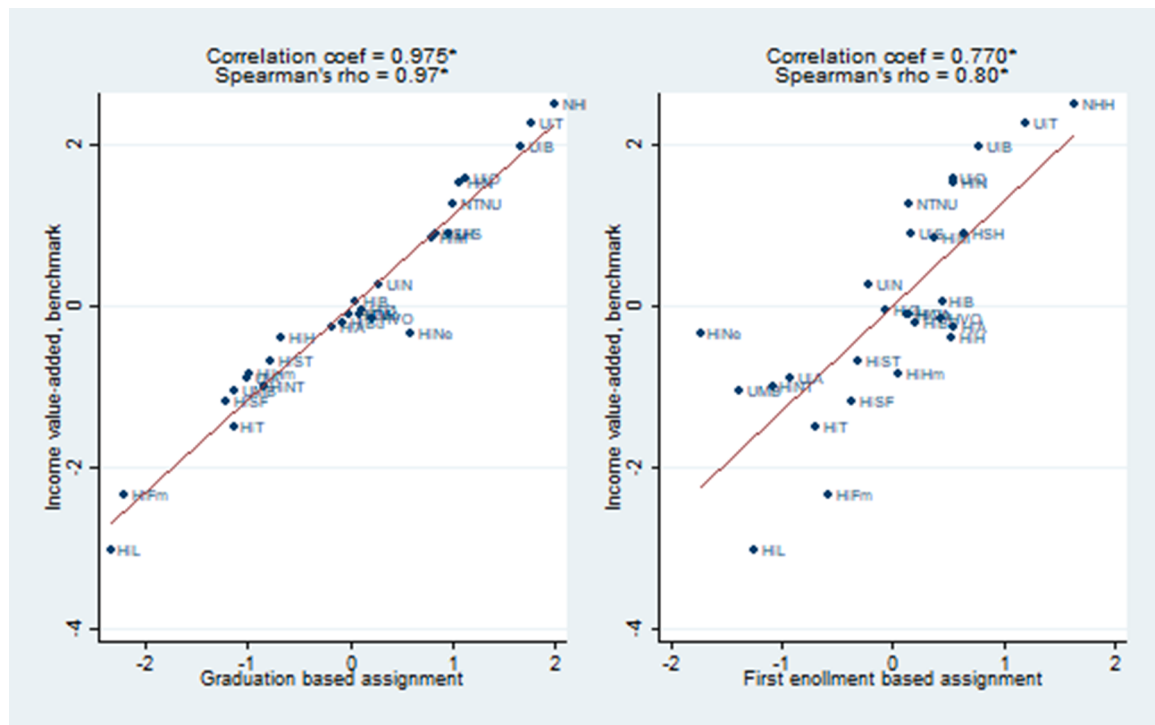


Fig. 4. Relationships between the quality indicator from the baseline model and quality indicators using different approaches to assign students to institutions
* indicates significant correlation at the 10% level.

The existing studies have used two main strategies to assign students to institutions. Cunha and Miller (2014) use data from Texas and assign students to the institution they first entered. Clotfelter et al. (2013) apply the same strategy using data on community colleges in North Carolina. On the other hand, Lindahl and Regner (2005), in a study of Swedish higher education institutions, assign students to the institution in which they were enrolled when exiting higher education. Both

strategies have possible weaknesses. Assigning students to an entering institution that only offers lower grade education, or where students for other reasons typically stay for a short time, will be attributed the income effects for students that most of the time attend and finish their education at other institutions. If the entering institutions are of lower quality than the “exiting institutions”, the entering institutions’ quality estimates will be biased upward.

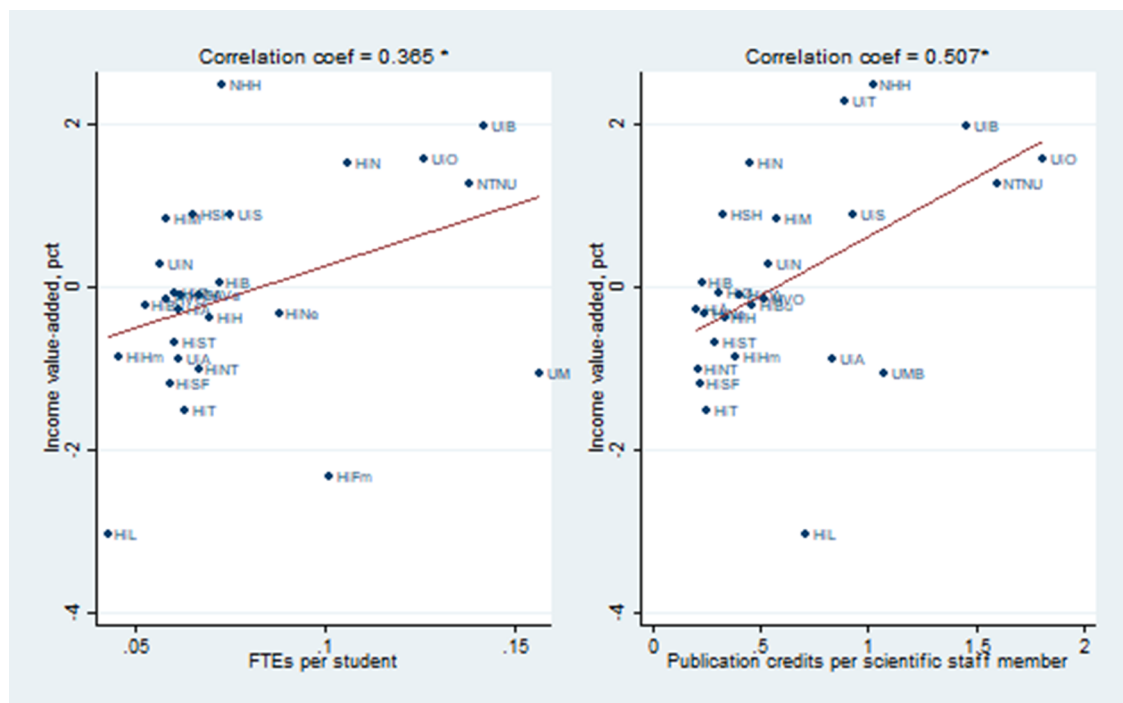


Fig. 5. Relationships between the quality indicator from the benchmark model and input characteristics
* indicates a significant correlation at the 10% level.

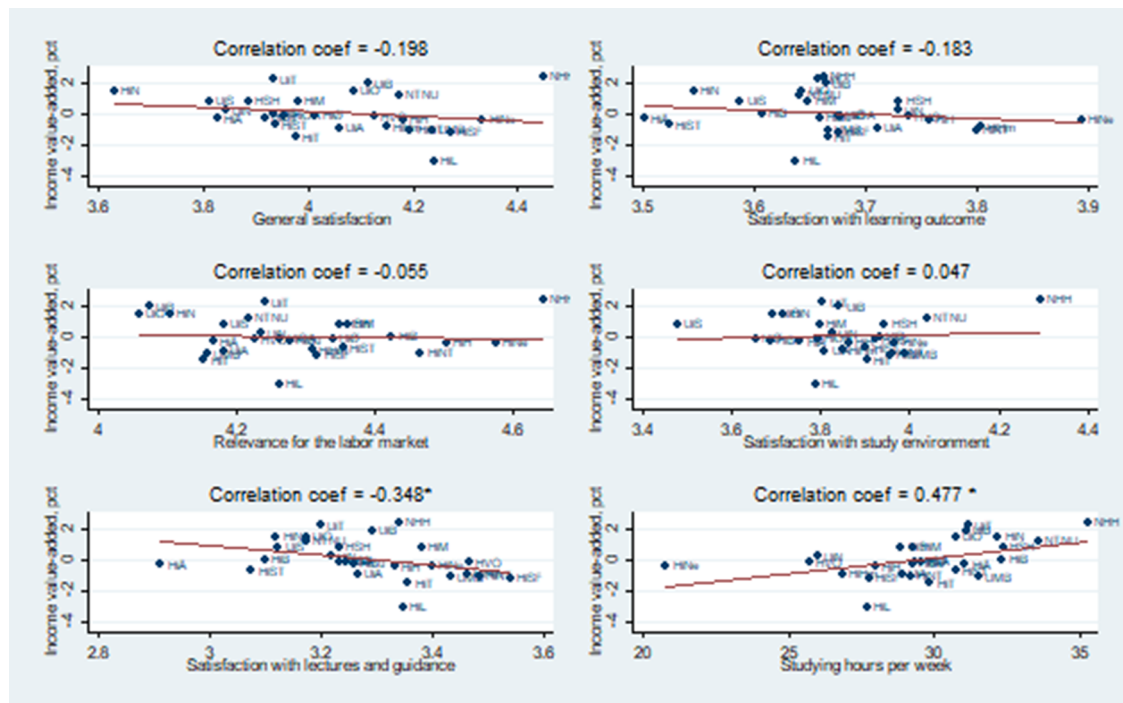


Fig. 6. Relationships between the quality indicator from the benchmark model and student assessments

* indicates a significant correlation at the 10% level.

In the opposite case, where students are assigned to the institution where they finish their higher education career, the contribution of institutions where students typically spend their first years is not reflected in the quality indicators. As an example, institutions that produce very good candidates at the lower level, but where the best candidates finish their education at another institution, the contribution of these first attended institutions will not be reflected in the quality indicators. This might lead to a downward bias in the estimated quality indicators for the “startup-institutions”.

An alternative view of the assignment problem is that student mobility creates measurement error in the institution assigned to the student, which may generally bias all institutions’ contribution toward zero. In this case, it is difficult to predict how the different assignment methods will affect the estimated quality distribution.

About 30% of the students change institution during their studies in the sample. Fig. 4 compares the estimated quality indicators obtained from the benchmark model (log income as outcome) with indicators estimated when assigning students to the institution where they entered higher education and with indicators assigning students to the institution they graduated.

Indicators using exit-based assignment and assignment based on entering institution are highly correlated with indicators using the benchmark model, with correlation (rank correlation) coefficients of 0.98 and 0.77 (0.97 and 0.80). Thus, we conclude that the income-based indicators are fairly robust with respect to the approach used when assigning students to institutions in the value-added framework. If researchers have to choose between using entry or exit assignment of institution, our results indicate that future analyses should use the latter assignment in analyses of institutional quality.

5. The association between quality indicators and institutional characteristics

This section discusses to what extent the labor market-based quality indicators are statistically associated with readily available institutional characteristics and subjective student satisfaction measures. Section 5.1

presents the relationship with some traditional input measures. Section 5.2 shows the relationship with student survey-based indicators, while Section 5.3 considers measures based on exam grade value-added and student attainment.

5.1. Teacher-student ratio and faculty quality indicators

Above, we saw that the old traditional universities were among the institutions with the highest estimated institutional quality. A potential reason for this pattern is that these institutions have more resources and high-quality faculty. Although we cannot identify the causal effects of these factors, it is instructive to see whether our estimated quality measures are systematically related to such variables measured at the institutional level.

Fig. 5 shows scatter plots between our income-based institutional quality indicator and two institutional characteristics: The number of full-time members of the scientific staff per student and a measure of scientific publications per full-time faculty at the institution. The first variable is a traditional resource measure, fairly similar to the concept of teacher-student ratios in compulsory schooling. The second variable can be viewed as an indicator of faculty quality in terms of research orientation and research quality.¹⁶ Both correlations are positive and statistically significant at the 10% level, see Fig. 5. This is to some extent driven by the old traditional universities having more resources and better research outcomes than the other institutions. We conclude that income-based value-added quality measures seem to be positively related to resources and faculty quality.

¹⁶ We use publication credits as defined and measured by the government. The publication credits measure the number of publications weighted by the quality of the publications. Scientific journals are sorted into three categories, level 0, level 1 and level 2, roughly reflecting scientific quality. Publications in level 0 receive no publication credits, while publications in level 1 and level 2 journals receive credits, with level 2 journals receiving the highest number of credits per publication.

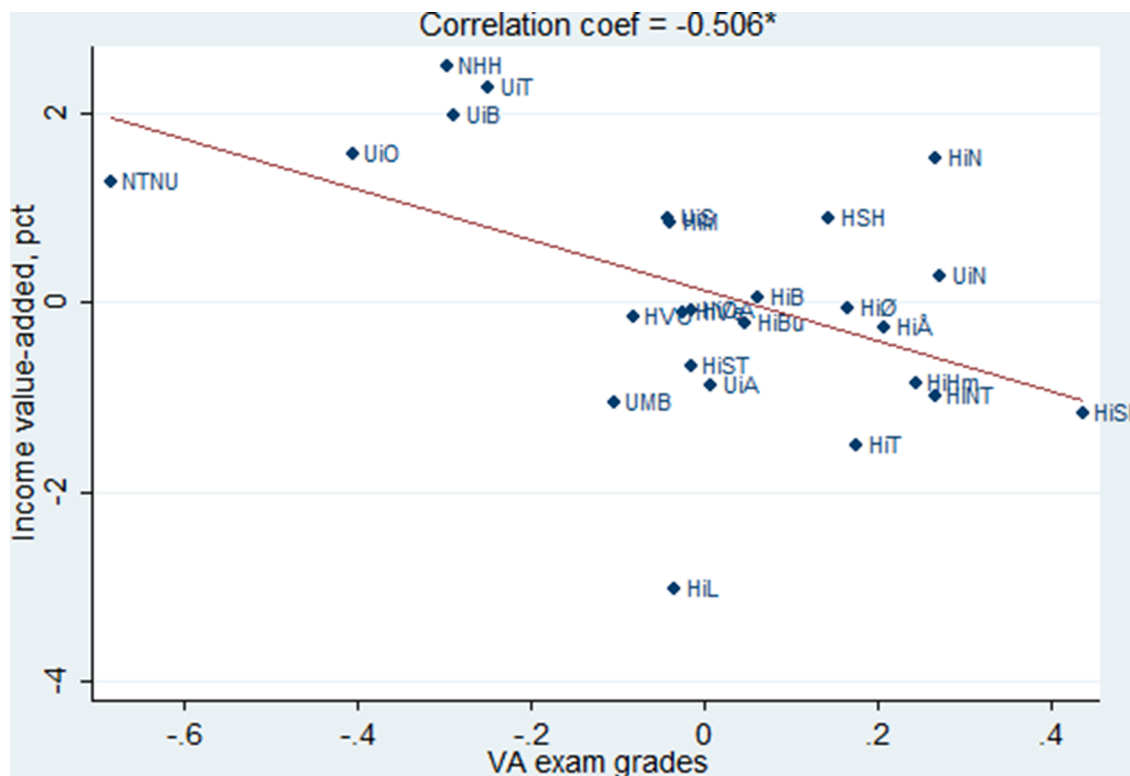


Fig. 7. Relationship between the quality indicator from the benchmark model and a quality indicator based on exam results

* indicates a significant correlation at the 10% level.

5.2. Student evaluations

Using data from the US, Carrell and West (2010) show that students' subjective evaluations of university teachers are not systematically related to long-term value-added-based measures of teacher quality. Braga et al. (2014), using data from an Italian university, even report negative associations. We have no similar measure of teacher quality. Still, it is interesting to compare our institutional quality indicators obtained using labor market outcomes with surveys measuring students' subjective assessment of the institutions. Ideally, we would like to have data from surveys undertaken when the students in our register data were actually enrolled in higher education. However, information is not available from that period. Instead, we use the survey undertaken in 2013 by NOKUT ("Studiebarometeret"). That year, approximately 55,000 students enrolled in all public higher education institutions were invited to participate in the survey. About 18,000 students participated, implying a response rate of only 32%. With these caveats in mind, we investigate the extent to which our quality indicators are associated with survey results from this year.

Fig. 6 shows scatter plots of our estimated quality indicators against the average score at the institutional level of student assessments of several learning-related characteristics. Interestingly, the only student-based survey indicator that correlates positively with the quality indicator is the number of study hours per week. In general, we can conclude that self-reported student satisfaction does not correlate positively with our income-based value-added quality measures. Actually, four of the five indicators for student satisfaction are negatively associated with our measure of institutional quality, and the association with satisfaction

with lectures and guidance is even significantly negative at the 10% level.

Our results resemble the finding in Carrell and West (2010) and Braga et al. (2014) that subjective student evaluations of teacher quality are not positively associated with long-term objectively measured teacher quality indicators. On the contrary, we find negative correlations between subjective student evaluation indexes and the quality indicator in our data. The most interesting finding is that the subjective measure of student effort appears to be positively associated with our income-based quality measure.

5.3. Exam grade value-added and measures of student attainment

In the introduction, we argued that institutional quality estimates based on students' exam results in higher education are unlikely to give reliable information since grading practices vary systematically across institutions. Strøm et al. (2013) estimated such exam-based quality indicators for the higher education institutions in Norway in a value-added framework by regressing exam grades against GPA from high school and similar individual characteristics as above, using students from approximately the same cohorts. Fig. 7 presents a scatter plot of the benchmark labor market-based quality indicator and the indicator based on exam grades estimated in Strøm et al. (2013). It appears that the two indicators are *negatively* and significantly associated. To the extent that income after exit from higher education measures important elements of true quality, this indicates that exam results are not very informative when assessing the quality of higher education institutions. The reason seems to be systematic variation in grade inflation across institutions as

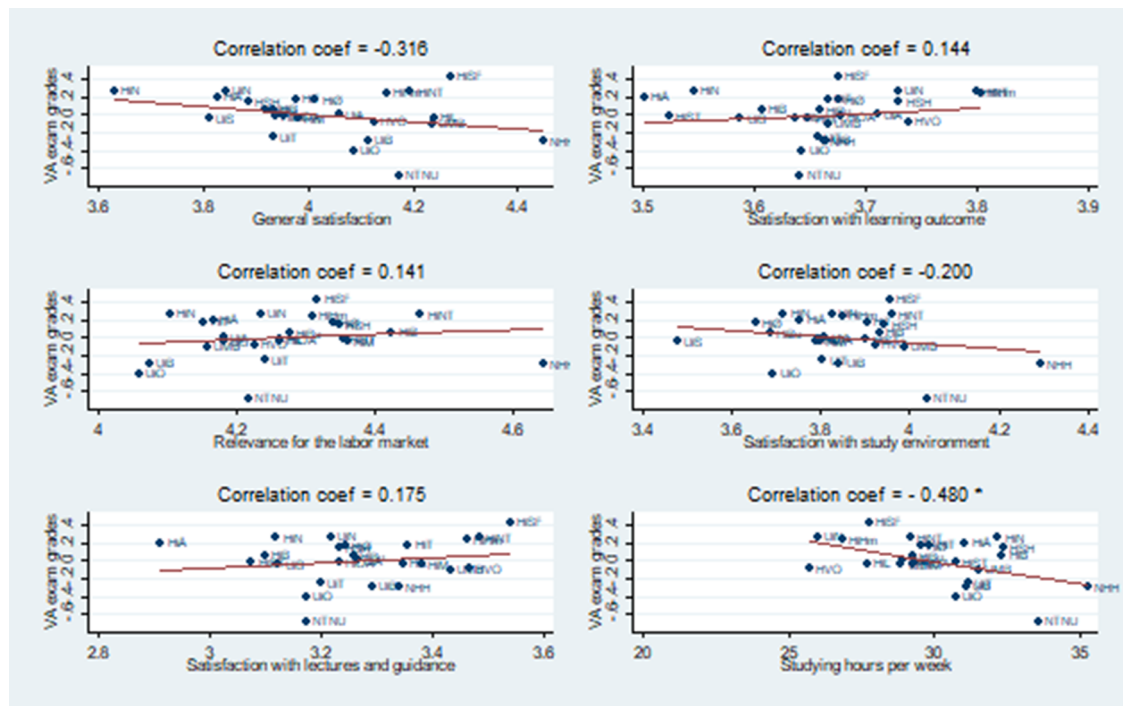


Fig. 8. Relationships between a quality indicator based on the exam results and student assessments

* indicates a significant correlation at the 10% level.

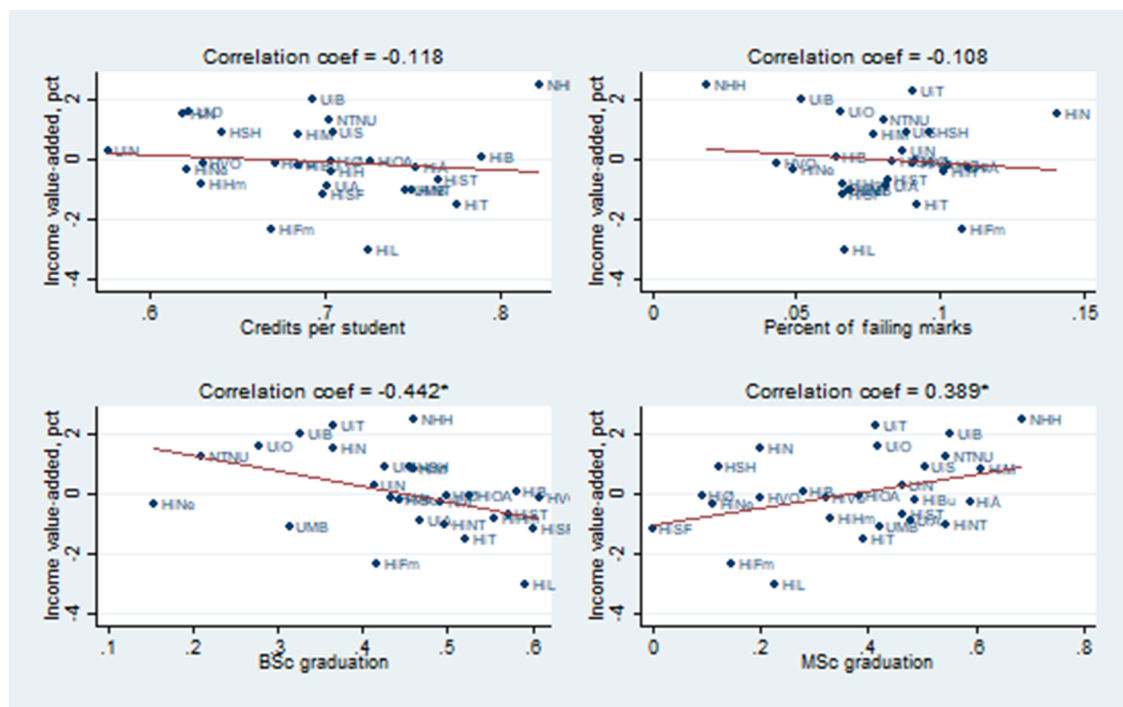


Fig. 9. Relationships between the quality indicator from the benchmark model and quality indicators based on student attainment

* indicates a significant correlation at the 10% level.

found in Strøm et al. (2013), and also in other countries, i.e., see the evidence from Italy in De Paola (2011).

Carrel and West (2010) and Braga et al. (2014), using data from the US and Italy, respectively, find that student evaluations of university teachers, fairly similar to those analyzed in Fig. 5 above, are positively related to short-run exam results, while negatively associated with long-run outcomes. We investigate to what extent these findings apply to our preferred quality measure in Norwegian higher education institutions. Fig. 8 presents the correlation between exam-based value-added and the same subjective student satisfaction and effort measures as used in Fig. 6. The overall impression is that the association between value-added exam results and student satisfaction differs somewhat from Fig. 6. Most importantly, while Fig. 6 indicates that student effort (studying hours per week) appears to be *positively* and significantly associated with our income-based quality indicator, effort turns out to be *negatively* associated with exam results. Qualitatively, this is in line with the finding in Carrel and West (2010) and Braga et al. (2014). This finding also confirms our worries that exam results are unlikely to give reliable information since grading differences between institutions are not taken into account, see also De Paola (2011), Møen and Tjelta (2010), and Strøm et al. (2013).

Our final outcomes measured at the institutional level are four measures of student attainment provided by *The database for Statistics on Higher Education* (DBH). These are the number of credits per student, the percent of failing exams, and the share of students completing a Bachelor's or Master's degree. Fig. 9 presents the associations between the institutional quality measure and student attainment. Our income-based quality measure is unrelated to the number of credits per student and the extent of failing the exams. However, it is significantly correlated with the completion of degrees. As expected, the quality indicator is positively related to graduation of master's programs. The negative relationship with the share of bachelor students completing the studies is surprising. The figure shows that the graduation rate is generally low at the old and traditional universities. It might be that these institutions have higher standards that require more effort, which results in high dropout rates at the bachelor level and high completion rates at the master level.

6. Concluding remarks

We exploit rich administrative data on individual labor market outcomes for students in an attempt to measure quality differences across higher education institutions in Norway. We estimate institution quality indicators using a value-added approach with individual income measured at age 28–31. Our primary control variables are high school grades, a battery of individual characteristics, and high school fixed effects. This approach's estimated quality indicators reveal significant differences across institutions, although the differences are much lower than unadjusted raw income differences. "Old" and traditional universities appear in the upper part of the estimated quality distribution, while most smaller regional university colleges appear in the lower part.

One may argue that in countries like Norway with inflexible wage-setting due to centralized collective bargaining and large public sectors, wages are less likely to reflect individual skills and productivity than in countries with less regulated and unionized labor markets. Therefore, we also investigate the relationship between income-based quality indicators with indicators based on other labor market outcomes. The income-based indicators appear to be systematically associated with employment and unemployment incidence indicators. While wage differences may appear small in countries with highly regulated

and unionized labor markets, in many cases, employees with similar educational backgrounds may possibly end up in many different jobs with different wage levels. Our results suggest that students from institutions highly ranked on the income-based quality indicator also are more able to find jobs that match their education than other students.

The assignment of students to institutions is a tricky question when a substantial fraction of students move between institutions during their higher education careers. We show that the estimated quality distribution is robust to whether students are assigned to their entering institution or their graduation institution, although the latter seems preferable.

We find by simple correlational analyses that our estimated income-based quality indicator is not associated with indicators based on subjective student assessments. This confirms earlier findings in the literature that subjective student assessments are unlikely to give reliable information about quality in higher education. An important finding is that student effort measured by self-reported hours of studying per week appears to be positively associated with the quality indicator. In addition, the indicator is positively correlated with the research performance of the faculty at the institution.

The institutions use the European grading system, where each individual's performance should be objectively assessed independent of institution, and that the system should provide a distribution of grades that has a fairly normal distribution. The results in the paper question whether this is the case. We find that our income-based quality indicator is significantly negatively associated with a value-added indicator using grades at exams (conditional on high school grades). In addition, the estimated quality measure using value-added exam results is negatively related to student effort. These results are consistent with weak institutions using easy grading of exams, which is at least partly revealed in the labor market.

7. Author statement

This is a joint statement on author contributions regarding the article «Quality measures in higher education: Norwegian evidence», ref. submission ECOEDU_2020_440.

CRedit authorship contribution statement

Torberg Falch: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Jon Marius Vaag Iversen:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Ole Henning Nyhus:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Bjarne Strøm:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Appendix A

Appendix Table 1-3.

Table A1

Institutions and the total number of normalized study years (60 credits) in the sample.

	Name	Total years	Human-ities and arts	Teacher training and pedag.	Social sc. & law	Busi-ness & admin	Natural sc., voc. & tech.	Health, welfare, and sport	Other fields
UiB	University of Bergen	19,146	4067	164	9183	–	2575	3108	49
UiO	University of Oslo	26,922	6657	1826	11,828	–	2956	3626	30
UiT	University of Tromsø	8041	1038	1091	1802	561	779	2635	135
NHH	Norwegian School of Economics	5310	–	–	–	5274	–	–	36
UMB	Norwegian University of Life Sciences	4010	119	59	389	495	2386	–	561
NTNU	Norwegian Uni. of Science and Techn.	28,262	4152	1214	4887	1055	14,170	2363	422
UiA	University of Agder	10,139	1272	2576	709	1623	1683	2202	76
HiB	Bergen University College	9558	–	3480	–	818	2446	2778	36
UiN	University of Nordland	3249	205	451	372	957	134	1082	48
HiBu	Buskerud University College	3387	280	400	252	859	996	600	–
HiFm	Finnmark University College	921	20	186	20	168	35	492	–
HiH	Harstad University College	574	–	8	–	195	–	359	12
HiHm	Hedmark University College	4403	116	1865	316	509	134	1104	359
HiL	Lillehammer University College	3884	259	228	967	1110	–	1320	–
HiM	Molde University College	1647	–	–	153	435	182	592	286
HiN	Narvik University College	646	–	–	–	37	384	217	9
HiNe	Nesna University College	379	50	183	–	–	66	76	4
NiNT	Nord-Trøndelag University College	3876	215	1158	74	291	466	1483	189
HiOA	Oslo og Akershus University College	16,115	883	3575	735	1198	2404	7319	–
HISF	Sogn og Fjordane University College	3384	142	551	146	561	205	1778	–
UiS	University of Stavanger	9144	1008	1783	708	1709	1917	1955	64
HSH	Stord/Haugesund University College	2767	–	935	–	397	223	979	233
HiST	Sør-Trøndelag University College	10,056	197	1614	–	1698	2388	4138	22
HiT	Telemark University College	5542	691	2195	–	659	599	1389	9
HiVe	Vestfold University College	3482	158	1442	149	342	541	588	263
HVO	Volda University College	2380	774	233	572	–	63	735	3
HiØ	Østfold University College	4650	378	1239	106	532	927	1467	–
HiÅ	Ålesund University College	1960	46	–	–	487	606	688	134
	Total	193,834	22,727	28,456	33,368	21,970	39,265	45,073	2980

Table A2

Value-added measures with log income in 2013 as the dependent variable.

Model acronym in Fig. 1	(1) Raw VA	(2) With controls	(3) Full model	(4) With region FE
UiB	0.0540*** (0.00147)	0.0411*** (0.00198)	0.0468*** (0.00229)	0.0455*** (0.00196)
UiO	0.0467*** (0.00206)	0.0344*** (0.00198)	0.0427*** (0.00224)	0.0377*** (0.00209)
UiT	0.0523*** (0.00354)	0.0527*** (0.00358)	0.0497*** (0.00361)	0.0502*** (0.00354)
NHH	0.122*** (0.00410)	0.0828*** (0.00393)	0.0519*** (0.00451)	0.0477*** (0.00345)
UMB	0.0526*** (0.00407)	0.0469*** (0.00381)	0.0164*** (0.00433)	0.0178*** (0.00378)
NTNU	0.0778*** (0.00382)	0.0547*** (0.00275)	0.0397*** (0.00192)	0.0368*** (0.00175)
UiA	0.0281*** (0.00329)	0.0273*** (0.00364)	0.0182*** (0.00335)	0.0206*** (0.00398)
HiB	0.0496*** (0.00326)	0.0458*** (0.00288)	0.0276*** (0.00311)	0.0260*** (0.00355)
UiN	0.0346*** (0.00410)	0.0475*** (0.00359)	0.0298*** (0.00406)	0.0271*** (0.00480)
HiBu	0.0545*** (0.00458)	0.0526*** (0.00406)	0.0249*** (0.00529)	0.0252*** (0.00518)
HiFm	−0.00692 (0.00921)	0.0213** (0.0105)	0.00377 (0.0107)	0.00461 (0.00623)
HiH	0.0230*** (0.00865)	0.0548*** (0.0123)	0.0232* (0.0121)	0.0179 (0.0128)
HiHm	0.0168*** (0.00350)	0.0262*** (0.00402)	0.0185*** (0.00431)	0.0222*** (0.00426)
HiL	0.00367 (0.00619)	−0.0117* (0.00626)	−0.00322 (0.00656)	0.000316 (0.00588)
HiM	0.0588*** (0.00626)	0.0645*** (0.00593)	0.0355*** (0.00679)	0.0328*** (0.00704)
HiN	0.0910*** (0.00794)	0.0821*** (0.00673)	0.0422*** (0.00612)	0.0378*** (0.00525)
HiNe	0.00437 (0.0149)	0.0228 (0.0137)	0.0237* (0.0120)	0.0361*** (0.00474)
HiNT	0.0140*** (0.00304)	0.0295*** (0.00409)	0.0170*** (0.00423)	0.0222*** (0.00437)
HiOA	0.0367*** (0.00314)	0.0408*** (0.00326)	0.0261*** (0.00247)	0.0173*** (0.00242)
HiSF	0.0204*** (0.00378)	0.0305*** (0.00446)	0.0153*** (0.00484)	0.0186*** (0.00482)
UiS	0.0535*** (0.00253)	0.0516*** (0.00327)	0.0359*** (0.00299)	0.0200*** (0.00369)
HSB	0.0388*** (0.00509)	0.0509*** (0.00444)	0.0360*** (0.00423)	0.0357*** (0.00385)
HiST	0.0436*** (0.00306)	0.0438*** (0.00285)	0.0202*** (0.00289)	0.0198*** (0.00284)
HiT	0.0107** (0.00425)	0.0187*** (0.00430)	0.0120*** (0.00399)	0.0160*** (0.00536)
HiVe	0.0347*** (0.00517)	0.0384*** (0.00548)	0.0260*** (0.00602)	0.0326*** (0.00543)
HVO	−0.00261 (0.00582)	0.00653 (0.00742)	0.0255*** (0.00691)	0.0238*** (0.00792)
HiØ	0.0357*** (0.00366)	0.0440*** (0.00392)	0.0264*** (0.00391)	0.0315*** (0.00452)
HiÅ	0.0651*** (0.00502)	0.0622*** (0.00597)	0.0244*** (0.00704)	0.0229*** (0.00605)
Born in 1983	−0.0249*** (0.00521)	−0.0237*** (0.00513)	−0.0248*** (0.00494)	−0.0215*** (0.00437)
Born in 1984	−0.0593*** (0.00603)	−0.0573*** (0.00599)	−0.0600*** (0.00609)	−0.0653*** (0.00582)
Born in 1985	−0.0709*** (0.00555)	−0.0708*** (0.00576)	−0.0765*** (0.00573)	−0.0771*** (0.00520)
Female	—	−0.216*** (0.00876)	−0.166*** (0.00733)	−0.160*** (0.00688)
First-generation immigrant	—	0.0173 (0.0158)	0.00186 (0.0143)	−0.00917 (0.0129)
Second-generation immigrant	—	0.0346* (0.0185)	0.0125 (0.0163)	0.00267 (0.0163)
Parental edu. = MSc	—	0.0190* (0.0112)	0.0253** (0.0114)	0.00728 (0.0115)
Parental edu. = BSc	—	0.0213* (0.0112)	0.0298** (0.0116)	0.0188 (0.0119)

(continued on next page)

Table A2 (continued)

Model acronym in Fig. 1	(1) Raw VA	(2) With controls	(3) Full model	(4) With region FE
Parental edu. = High school	–	0.0262** (0.0108)	0.0274** (0.0111)	0.0180 (0.0114)
Average high school grades	–	1.060*** (0.241)	0.949*** (0.227)	1.123*** (0.203)
(Average high school grades) ²	–	–0.261*** (0.0597)	–0.228*** (0.0563)	–0.273*** (0.0505)
(Average high school grades) ³	–	0.0228*** (0.00485)	0.0194*** (0.00459)	0.0234*** (0.00413)
Humanities and arts (share)	–	–	–0.212*** (0.0138)	–0.195*** (0.0112)
Teacher training and pedagogy (share)	–	–	–0.00141 (0.0112)	0.0213* (0.0119)
Business and administration (share)	–	–	0.191*** (0.0110)	0.190*** (0.00867)
Natural sciences, voc. & tech. subjects (share)	–	–	0.226*** (0.00969)	0.221*** (0.0111)
Health, welfare, and sport (share)	–	–	0.0703*** (0.0104)	0.0901*** (0.0106)
Other fields (share)	–	–	0.0730* (0.0414)	0.0820* (0.0425)
High school FE	No	Yes	Yes	Yes
Labor market region FE	No	No	No	Yes
Observations	53,996	53,996	53,996	53,459

Note. A constant term is included in all models. Standard errors are clustered at the region of residence at age 16. ***, **, and * indicates statistical significance at the 1, 5, and 10% level, respectively.

Table A3

Value-added measures by subject field with log income in 2013 as the dependent variable.

	Subject field Humanities & arts	Teacher training & pedagogy	Social sciences & law	Business & administration	Natural sciences, voc. & tech. subjects	Health, welfare & sport
UiB	–0.0151* (0.00881)	0.1402*** (0.03317)	0.0456*** (0.00383)	–	0.0361*** (0.00529)	0.0978*** (0.00395)
UiO	–0.0019 (0.00544)	0.0599*** (0.00683)	0.0447*** (0.00338)	–	0.0177*** (0.0062)	0.0954*** (0.00414)
UiT	0.0007 (0.01365)	0.0507*** (0.00768)	0.0438*** (0.00592)	0.0406*** (0.00751)	0.0068 (0.01565)	0.0837*** (0.00418)
NHH	–	–	–	0.0523*** (0.00514)	–	–
UMB	–0.3729** (0.14178)	–0.0168 (0.04295)	0.0798*** (0.0163)	0.0262 (0.0172)	0.0350*** (0.00634)	–
NTNU	–0.0098 (0.00688)	0.0469*** (0.00907)	0.0122** (0.00485)	0.1249*** (0.01332)	0.0560*** (0.00323)	0.0831*** (0.00631)
UiA	–0.0112 (0.01747)	0.0306*** (0.00723)	–0.0019 (0.01293)	0.0360*** (0.00843)	0.0444*** (0.00657)	0.0139 (0.00842)
HiB	–	0.0447*** (0.007)	–	0.0217*** (0.008)	0.0697*** (0.00516)	0.0249*** (0.0058)
UiN	–0.0682 (0.04682)	0.0769*** (0.01137)	0.0420** (0.01664)	0.0165* (0.00896)	0.0012 (0.01702)	0.0445*** (0.00907)
HiBu	0.0021 (0.02496)	0.0626*** (0.0172)	–0.0199 (0.0232)	–0.0025 (0.01062)	0.0508*** (0.00773)	0.0469*** (0.00649)
HiFm	0.1876 (0.13117)	0.0535*** (0.01998)	–0.0718 (0.05797)	–0.0805 (0.0575)	0.0102 (0.04181)	0.0236** (0.0105)
HiH	–	–0.0598 (0.15279)	–	–0.0215 (0.02391)	–	0.0491*** (0.01446)
HiHm	–0.0161 (0.04686)	0.0451*** (0.00721)	–0.0961*** (0.02238)	0.0392** (0.01972)	–0.014 (0.02375)	0.0522*** (0.00843)
HiL	0.0095 (0.04321)	0.0506** (0.02255)	–0.0363** (0.01587)	–0.0337*** (0.01164)	–	0.0351*** (0.00767)
HiM	–	–	0.0329 (0.02196)	0.0187 (0.01697)	–0.0169 (0.01567)	0.0267** (0.01187)
HiN	–	–	–	0.0204** (0.00842)	0.0569*** (0.00934)	0.0561*** (0.01014)
HiNe	0.0204 (0.03066)	0.0543*** (0.0136)	–	–	–0.0723 (0.05686)	0.0681*** (0.01232)
HiNT	–0.0303 (0.02194)	0.0790*** (0.0108)	–0.0033 (0.04531)	–0.0243** (0.01107)	–0.0705*** (0.01413)	0.0361*** (0.00703)
HiOA	0.0069 (0.01204)	0.0502*** (0.00724)	–0.0133 (0.01439)	0.0222*** (0.00683)	0.0592*** (0.00489)	0.0350*** (0.00592)
HiSF	0.0662*	0.0415**	–0.0162	–0.0282***	0.0561**	0.0274***

(continued on next page)

Table A3 (continued)

	Subject field Humanities & arts	Teacher training & pedagogy	Social sciences & law	Business & administration	Natural sciences, voc. & tech. subjects	Health, welfare & sport
	(0.03521)	(0.01681)	(0.03117)	(0.00929)	(0.0214)	(0.00889)
UiS	0.0016	0.0390***	0.0529***	0.0105	0.1037***	0.0008
	(0.01094)	(0.00703)	(0.00865)	(0.0074)	(0.00849)	(0.00675)
HSB	–	0.0524***	–	–0.0005	0.1421***	0.0242**
		(0.0064)		(0.01039)	(0.01211)	(0.01012)
HiST	0.0453*	0.0425***	–	0.0296***	0.0484***	0.0145***
	(0.02386)	(0.00667)	–	(0.00553)	(0.00568)	(0.00545)
HiT	–0.0032	0.0324***		–0.0026	0.0417***	0.0067
	(0.01435)	(0.00753)		(0.01207)	(0.00828)	(0.00699)
HiVe	–0.0505	0.0493***	0.0029	–0.0278	0.0642***	0.0263*
	(0.03486)	(0.00761)	(0.02584)	(0.0208)	(0.00978)	(0.01497)
HVO	–0.0236	0.1277***	0.0295	–	0.0249	0.0220*
	(0.02823)	(0.02758)	(0.01995)		(0.04132)	(0.01155)
HiØ	0	0.0499***	0.0612***	–0.0061	0.0366***	0.0417***
	(0.02985)	(0.00984)	(0.01905)	(0.01612)	(0.00815)	(0.00718)
HiÅ	0.1017	–	–	–0.0355***	0.0575***	0.0313***
	(0.11926)			(0.0117)	(0.01166)	(0.01146)
Birth year FE	Yes					
SES variables	Yes					
HS grades (3rd pol.)	Yes					
High school FE	Yes					
Labor market region	No					
FE						
Study field shares	Yes					
Observations	53,996					

Note. All coefficients are estimated in a single model with interaction terms between institutions and subject field. Interaction terms between institutions and *other fields* are also included but not reported. Standard errors are clustered at the region of residence at age 16. ***, **, and * indicates statistical significance at the 1, 5, and 10% level, respectively.

References

- Andrews, R., Li, J., & Lovenheim, M. F. (2014). Heterogeneous paths through college: Detailed patterns and relationships with graduation and earnings. *Economics of Education Review*, 42, 93–108.
- Bandiera, O., Larcinese, V., & Rasul, I. (2010). Heterogeneous Class Size Effects: New Evidence from a Panel of University Students. *Economic Journal*, 120(549), 1365–1398.
- Belfield, C., Britton, J., Buscha, F., Dearden, L., Dickson, M., van der Erve, L., Sibiet, L., Vignoles, A., Walker, I., & Zhu, Y. (2019). The returns to higher education: New evidence from administrative data in the UK. *mimeo*.
- Bhuller, M., Mogstad, M., & Salvanes, K. G. (2017). Life-cycle earnings, education premiums, and internal rates of return. *Journal of Labor Economics*, 35(4), 993–1030.
- Black, S. E., Devereux, P. J., & Salvanes, K. G. (2013). Under Pressure? The Effect of Peers on Outcomes of Young Adults. *Journal of Labor Economics*, 31(1), 119–153.
- Braga, M., Paccagnella, M., & Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review*, 41, 71–88.
- Card, D., & Krueger, A. B. (1992). Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States. *Journal of Political Economy*, 100(1), 1–40.
- Carrell, S. E., & West, J. E. (2010). Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors. *Journal of Political Economy*, 118(3), 409–432.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, 104(9), 2633–2679.
- Clotfelter, C. T., Ladd, H. F., Muschkin, C. G., & Vigdor, J. L. (2013). Success in Community College: Do Institutions Differ? *Research in Higher Education*, 54(7), 805–824.
- Cunha, J. M., & Miller, T. (2014). Measuring value-added in higher education: Possibilities and limitations in the use of administrative data. *Economics of Education Review*, 42, 64–77.
- De Paola, M. (2011). Easy grading practices and supply–demand factors: Evidence from Italy. *Empirical Economics*, 41(2), 227–246.
- De Paola, M., Ponzio, M., & Scoppa, V. (2013). Class size effects on student achievement: Heterogeneity across abilities and fields. *Education Economics*, 21(2), 135–153.
- Deming, D. J. (2014). Using School Choice Lotteries to Test Measures of School Effectiveness. *American Economic Review*, 104(5), 406–411.
- Deming, D. J., & Figlio, D. (2016). Accountability in US Education: Applying lessons from K-12 experience to higher education. *Journal of Economic Perspectives*, 30(3), 33–56.
- Dyrstad, J. M., Sohlman, M. M., & Teigen, T. H. (2021). Government funding incentives and study program capacities in public universities. *Mimeo*.
- Falch, T., Sandsør, A. J., & Strøm, B. (2017). Do smaller classes always improve Students' Long-run outcomes? *Oxford Bulletin of Economics and Statistics*, 79(5), 654–688.
- Fredriksson, P., Öckert, B., & Oosterbeek, H. (2013). Long-term effects of class size. *The Quarterly Journal of Economics*, 128(1), 249–285.
- Guarino, C. M., Reckase, M. D., & Wooldridge, J. M. (2015). Can value-added measures of teacher performance be trusted? *Education Finance and Policy*, 10(1), 117–156.
- Guarino, C. M., Stacy, B. W., & Wooldridge, J. M. (2019). Comparing and assessing the consequences of two different approaches to measuring school effectiveness. *Educational Assessment, Evaluation and Accountability*, 31, 437–463.
- Haider, S., & Solon, G. (2006). Life-Cycle variation in the association between current and lifetime earnings. *American Economic Review*, 96(4), 1308–1320.
- Havnes, T., & Mogstad, M. (2015). Is universal child care leveling the playing field? *Journal of Public Economics*, 127, 100–114.
- Kirkeboen, L. J., Leuven, E., & Mogstad, M. (2016). Field of study, earnings, and self-selection. *Quarterly Journal of Economics*, 131(3), 1057–1111.
- Koedel, C., Mihaly, K., & Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review*, 47, 180–195.
- Kurlaender, M., Carrell, S., & Jackson, J. (2016). The promises and pitfalls of measuring community college quality. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 2(1), 174–190.
- Leuven, E., & Løkken, S. A. (2020). Long-term impacts of class size in compulsory school. *Journal of Human Resources*, 155(1), 309–348.
- Lindahl, L., & Regnér, H. (2005). College choice and subsequent earnings: Results using Swedish sibling data. *Scandinavian Journal of Economics*, 107(3), 437–457.
- Møen, J., & Tjelt, M. (2010). Grading standards, student ability and errors in college admission. *Scandinavian Journal of Educational Research*, 54(3), 221–237.
- OECD. (2008). *Measuring Improvements in Learning Outcomes: Best Practices to Assess the Value-Added of Schools*. OECD. Paris.
- Strøm, B., Falch, T., Gunnes, T., & Haraldsvik, M. (2013). *Karakterbruk og kvalitet i høyere utdanning*. Report number 03/13. Center for economic research at NTNU.
- Strøm, B., Falch, T., Iversen, J. V., & Nyhus, O. (2016). *Kvalitetsindikatorer i universitets- og høyskolesektoren*. Report number 05/16. Center for economic research at NTNU.
- U.S. Department of Education. (2019). *College Scorecard*. Retrieved from <https://collagescorecard.edu.gov>.