

Disciplinary Differences in Academics' Perceptions of Performance Measurement at Nordic Universities

Johan Söderlind & Lars Geschwind

Department of Learning in Engineering Sciences, KTH Royal Institute of Technology, Stockholm, Sweden

Article Type: Research Article

Corresponding Author: Johan Söderlind, johanso2@kth.se

Funding: This work was supported by the Research Council of Norway under grant number 237782 and Riksbankens Jubileumsfond grant number FSK15-1059:1.

Cite as: Söderlind, J., & Geschwind, L. (2020). Disciplinary differences in academics' perceptions of performance measurement at Nordic universities. *Higher Education Governance & Policy*, 1(1), 18-31.

Access: <https://dergipark.org.tr/en/pub/hegp/issue/55277/758741>

Disciplinary Differences in Academics' Perceptions of Performance Measurement at Nordic Universities

Johan Söderlind^{1*} & Lars Geschwind²

^{1,2}Department of Learning in Engineering Sciences, KTH Royal Institute of Technology, Stockholm, Sweden

Abstract

As performance measurement has become increasingly common at Nordic universities, concerns have been raised that disciplinary differences create difficulties in comparing academic performances. To better understand the potential of utilising performance measures for the management of academic work, this study explores how academics perceive governance and steering based on performance measurement. Building on an established typology of the disciplines that distinguishes the hard sciences from the soft and the applied sciences from the pure, we ask how academics perceive performance measurement depending on their disciplinary affiliation. The empirical material consists of a survey sent to academics in four Nordic countries. Our results show there are clear differences in the attitudes toward performance measurement between academics from different disciplines. Academics from the hard applied sciences are more positive about performance measurement than any other group, and academics from the soft pure sciences are more negative. These findings are consistent with notions about the poor adaptation of metrics to publication practices within the soft sciences and greater sensitivity to performance measurement among the applied sciences. The main contribution of the article is to provide empirical data that support the notion that performance measures are accepted to different degrees in different disciplines.

Keywords: Disciplines, performance measurement, Nordic universities

Introduction

“Science and technology departments commonly become entrepreneurial first. Social sciences departments, aside from economics and business, find the shift more difficult and lag behind... Uneven adoption of new ways should be expected.” (Clark 1998, p. 88).

In recent decades, measuring academic performance has become a common practice at Nordic universities, as elsewhere. Student success is meticulously assessed using retention and throughput measures (Campbell & Oblinger, 2007), and course evaluations are used to judge teaching effectiveness (Spooren, Brockx, & Mortelmans, 2013). Research output is measured based on a variety of publication and citation indicators, which are used in allocating funds (Aagaard, 2015; Hammarfelt, Nelhans, Eklund, & Åström 2016; Hicks, 2012), guiding research activities (Mingers & Wilmott, 2013) and hiring and promotion decisions (Hammarfelt & Rushforth, 2017). Worldwide university rankings utilise some of these measures, as well as indicators of institutional reputation, which makes it increasingly important for universities to publicly demonstrate their merit (Harvey, 2008; van Vught & Westerheijden, 2010). Alternative metrics are also being developed to gauge the impact of research in wider contexts, such as public policy documents, mass media and social networks (Bornmann, 2014; Piwowar, 2013). The importance of universities in their surrounding society is another focus for policymakers, as indicators are being developed to assess the societal impact of academic work (Gulbrandsen & Slipersaeter, 2007).

* **Corresponding Author:** Johan Söderlind, johanso2@kth.se

¹**ORCID:** [0000-0002-4280-9866](https://orcid.org/0000-0002-4280-9866); ²**ORCID:** [0000-0003-2983-5573](https://orcid.org/0000-0003-2983-5573)

(Research Article) Cite as: Söderlind, J., & Geschwind, L. (2020). Disciplinary differences in academics' perceptions of performance measurement at Nordic universities. *Higher Education Governance & Policy*, 1(1), 18-31.

Received: February 28, 2020; **Revised:** June 4, 2020; **Accepted:** June 6, 2020; **Published:** June 30, 2020

The purpose of the present article is to contribute to the discussion of how these developments have impacted the higher education (HE) sector. In particular, we highlight the importance of disciplinary cultures, which is an issue that has often been disregarded. Studies of organisational behaviour are generally little concerned with internal organisational dynamics, as the organisation is often treated as a singular entity. In addition, disciplinary differences are frequently overlooked in studies exploring reactions to new public management (NPM) reforms and performance measurement (e.g. Hansen et al., 2019; Kallio, Kallio, Tienari, & Hyvönen, 2016; Paradeise, Reale, Bleiklie, & Ferlie, 2009; Pinheiro, Geschwind, Hansen, & Pulkkinen, 2019). However, it is well known that universities are comprised of a multitude of academic tribes and territories (Becher & Trowler, 2001) and that despite much reform that emphasises the institution as an entity, academics rely on their disciplines for identity formation (Henkel, 2000). Performance measurement is a good example of such reform, as it is often applied uniformly to universities and university systems, with little regard for the disciplinary differences that affect the conditions under which university departments and individual academics operate.

We therefore concur with Becher (1994) that the distinctive characteristics of disciplinary cultures must be recognised in HE research and believe it is imperative to include this perspective in analysing performance measurement. Whereas many scholars and practitioners alike have identified differences in the way academics respond to performance measurement, as illustrated by the quote from Burton Clark above, few studies have in a more systematic way addressed this issue. For this reason, we explore how attitudes toward measuring the performance of academic work vary between scientific disciplines. We ask whether academics from some disciplines are more inclined to accept governance and steering based on performance measures while others are more sceptical about metrics. Specifically, we address the following research question: *How do academics from different disciplines perceive the performance measurement of academic work?* We approach this question by disentangling some notions about how and why scientific disciplines differ and by developing some tentative ideas about how this is related to performance measurement. We continue with an analysis of a survey sent to academics at universities in four Nordic countries, thus gauging their attitudes about several aspects of performance measurement. In conclusion, we sum up the key findings and discuss their implications for further research, policy and practice.

Academic Disciplines

For academics, the disciplines are the primary units of membership and identification. They have a significant impact on academic work, as they shape attitudes, beliefs and practices (Becher & Trowler, 2001; Biglan, 1973a; 1973b). However, the concept of academic disciplines has been defined in a number of different ways, and it is not always clear what separates the disciplines from each other. As noted by Sugimoto and Weingart (2015), various conceptualisations rely on aspects that are cognitive, social, communicative, nominal, historical or institutional. Discussions about the scientific disciplines usually refer to several of these aspects. For example, Stichweh (2009) combines most of them when he illustrates the historical twists and turns involved in the formation of the scientific disciplines. He emphasises how scientists took on role differentiation based on the field of study and institutionalised scholarly communities with specific communications systems around these fields. Although the relative importance of these aspects may be discussed, the relationship between them should be understood as complex and intertwined. For our present purposes, an academic discipline will be understood as an academic community with institutionalised goals and structures for the creation, communication and dissemination of scholarly knowledge.

Analytical framework

Academic cultures differ not only between disciplines but also within them. Yet, a discipline is often taken to be a useful level of analysis because disciplines correspond roughly to the departments within universities, which often is the organisational manifestation of the structure of knowledge. However, the major differences between academic cultures discussed in the literature are often found between aggregations of disciplines. Although Trowler (2014) makes a strong argument that there are major differences within disciplines that are obscured by macro-level categorisations, we argue that the

endless complexity of (sub)disciplinary variations will yield less insight compared to what is possible by simply aggregating disciplines under a number of major categories. Therefore, we will apply a fourfold typology of the disciplines adopted from previous studies (Becher & Trowler 2001; Biglan, 1973b; Stoecker, 1993). On one hand, the categorisation distinguishes the hard sciences from the soft, and on the other hand it distinguishes the pure sciences from the applied (see Table 1). While it may be desirable to further explore the differences in attitudes toward performance measurement, our study is a first attempt to distinguish differences at this aggregate level.

Table 1. Typology of scientific disciplines

	Applied	Pure
Hard		
Soft		

In attempting to map the academic disciplines, some heuristic dimensions have repeatedly been applied. The most common one is the distinction between the hard and the soft sciences. This division is based on several observations and ideas but is often related to what Kuhn (1962) describes as normal or paradigmatic science, where knowledge is progressively accumulated upon previous findings. In contrast, pre- or non-paradigmatic science is characterised by conflicting foundational premises, which prevents a linear progression of knowledge accumulation. It is the hard sciences that are expected to demonstrate a stronger consensus about theory and methods than what may be observed in the soft sciences. However, existing studies have been unable to establish any difference in the cumulateness of the sciences (Cole, 1983; Hedges, 1987). In contrast, a related idea concerns the degree of theoretical integration whereby empirical facts are connected to theoretical formulations, which has been found to be higher in the hard sciences (Smith, Best, Stubbs, Johnston, & Archibald, 2000). Similarly, the idea that the more complex the object of study, the more difficult it is to study has been proposed. Due to technical, ethical and practical considerations, the methods applied will therefore differ between disciplines. While experiments constitute a reliable and powerful method of inference, complex phenomena are often hard to study in this way. Instead, observations and other less rigid methods are used. The complexity of the object of study therefore affects the ability of a discipline to achieve consensus by reaching conclusive evidence and settling intellectual debates (Fanelli & Glänzel, 2013).

Another commonly made distinction is that between the pure and the applied sciences. Here, the main difference is supposed to lie in the purpose of the scientific endeavour. While scientists belonging to the former group are said to seek knowledge for its own sake, those belonging to the latter are instead interested in the practical application of research. The dichotomy between pure and applied science has been constructively critiqued (Gibbons et al., 1994; Stokes, 1997) but is deeply entrenched. In particular, it underpins the linear model of innovation, which holds that innovation occurs in a linear process starting with basic research, followed by applied research and ending in product development (Godin, 2006).

It has been suggested that a difference between the pure and applied sciences is their responsiveness and openness to their external environment. For instance, it has been claimed that the pure sciences are essentially self-regulating, while the applied sciences are open to external influence (Becher & Trowler, 2001, p. 176ff). Reasons are that the relevance of teaching and research in the applied sciences makes external actors prone to invest resources to promote particular goals, but also to more generally interact with academics in order to influence the outcome of academic work. The origins of the disciplines are also said to differ, as disciplines within the pure sciences are established through processes internal to the scientific system, while those within the applied sciences are often established because of external demands (Becher & Trowler, 2001, p. 171). Furthermore, it has been proposed that an important mechanism reinforcing the difference between pure and applied science is the structure of the academic training within these two categories (El-Khawas, 1996). The difference is that

scholars in the pure sciences are trained for purposes of scientific discovery, whereas scholars in the applied sciences are subject to substantial professional preparation. This also means that the prospects for academics in the applied sciences to move between the university and professional practice are better than for academics in the pure sciences.

Categorising the disciplines

For examples of how the disciplines may be classified in accordance with the fourfold typology, we turn to previous research on this topic (Becher & Trowler, 2001; Biglan, 1973b; Stoecker, 1993). According to these studies, the hard applied sciences include agriculture, engineering and clinical medicine; the hard pure sciences include biology, mathematics, geology and physics; the soft applied sciences include education, law, social administration and economics; and the soft pure sciences include sociology, theology, languages, literature, history and philosophy. The categorisation of any specific discipline may be debatable and will always be a matter of judgement, particularly as it can be done on a number of grounds.

In the present study, the classification of academics is based on survey responses to the following survey item: ‘My research/field of science is classified as’. Possible answers were derived from the Organisation for Economic Co-operation and Development (OECD, 2007) classification of science and technology fields, including natural sciences (n=1,160); engineering and technology (n=749); medical and health sciences (n=838); agricultural sciences (n=130); social sciences (n=1,366); humanities (n=870); and other (n=176). Using this classification adheres to standard conceptualisations of disciplines and scientific fields. However, a limitation of the survey is the low granularity of the answers. This entails some problems with the categorisation of the respondents, further discussed below, that could have been better solved would the answers have been given at a more detailed level. It also prevents analyses at a more detailed level, but for present purposes this is not considered a problem, because the study attempts to provide a wide overview rather than a fine-grained comparison between all the disciplines.

Based on the argumentation above, we grouped engineering and technology, medical and health sciences and agricultural sciences to form what we call the applied sciences. Following the examples of previous studies, we then categorised the applied sciences as hard applied, the natural sciences as hard pure, the social sciences as soft applied and the humanities as soft pure (see Table 2). The group ‘other’ was excluded. While there should be little controversy regarding most of these categorisations, it is clear that the social sciences contain both applied disciplines (e.g. education, law and economics) and pure disciplines (e.g. sociology and anthropology). Social sciences may therefore be conceptualised as something in between the pure and the applied soft sciences. Compared to the humanities, however, they should be considered more applied than pure, and we will therefore categorise them as applied soft (but with this caveat, as indicated by the parenthesis in Table 2).

Table 2. Categorisation of scientific disciplines

	Applied	Pure
Hard	Applied sciences	Natural sciences
Soft	(Social sciences)	Humanities

Although the method used implies some difficulty in identifying each respondent’s specific discipline, the survey responses are assumed to correspond well with our fourfold typology. However, it should be noted that the categorisation of academics is primarily based on their research interests. It therefore neglects teaching and other duties, which may affect the disciplinary culture in which the respondents are situated. However, the choice to focus on the respondents’ research was made because most ideas about what defines the disciplines primarily revolve around aspects of research.

Performance measurement of academic work

As noted in the introduction, the HE sector has seen an increase in performance measurement. This may be seen in the light of several concurrent developments emphasising how the role of universities has shifted in recent decades. Elzinga (1997) notes that an epistemic drift has occurred within the research sector, implying a shift in emphasis from internal quality control to the external assessment of relevance. Slaughter and Leslie (1997) describe the increasing marketization of the HE and research sectors, giving rise to academic capitalism where universities are becoming competitors in a global market for students, faculty and funding. It has been suggested that these changes have given rise to 'the entrepreneurial university' (Clark, 1998), as the role of universities has shifted from that of cultural institutions to that of corporate enterprises in the knowledge industry (Bleiklie, 1998). In many countries, there has also been increasing managerialism. This often implies new practices and management tools, including different forms of performance measurement (Amaral, Meek, & Larsen, 2003; Deem, Hillyard, & Reed, 2007).

These developments may be understood as part of the wider NPM reforms of the public sector that have occurred in most Western countries in recent decades, where inspiration is drawn from management practices in the private sector. This often includes the professionalization of management, the devolution of responsibilities, the formalisation of relationships, increased competition, explicit performance standards and a stronger emphasis on output control and performance measurement (Hood, 1991; Pollitt & Bouckaert, 2004). Many of these changes have been observed in the HE sector. In particular, there is increasing competition for resources, stronger accountability systems have been developed, branding has become an important activity and there has been increasing autonomy of universities, which most often is coupled with stronger output control and more performance measurement (Christensen, 2011; Elzinga, 2012; Ferlie, Musselin, & Andresani, 2008).

These changes have all affected the Nordic HE sector, which in recent decades have seen a strong wave of reforms that has followed international trends. Since the early 1990s, there has also been a dramatic rise in student numbers and growing research budgets that have prompted stronger demands for efficiency and accountability. While having a long tradition of state control where also important stakeholders have had a significant influence, the Nordic higher education sector has experienced a shift towards more autonomy for the HEIs, but also more marked-based governance (Gornitzka et al. 2004). In all of the Nordic countries, the institutional autonomy has been a central feature of higher education reform, as has been the introduction of governance models based on management by results. All countries have also seen the establishment of organisations designated for the evaluation of higher education (Fägerlind & Strömqvist, 2004). Also performance-based funding has been an increasingly important governance tool as national governments have delegated authority to the HEIs. Within the HEIs, managerial modes of governance have been introduced in all countries at the expense of collegial decision-making structures (Ahola et al., 2014). Mergers has also been a prominent aspect of the Nordic higher education landscape in recent years (Pinheiro et al., 2016). Although the higher education landscapes of the Nordic countries also exhibit significant differences, it is clear that many reforms in recent decades have had similar aims and rationales, and have been induced by similar pressures. Among these changes, performance measurement constitutes a common denominator that promotes accountability and efficiency in resource management, and that often is believed to enhance quality.

In the present study, we explore how the increasing performance measurement of academic work is perceived by academics at Nordic universities. Although the study of macro-level trends is important to understand how HE systems are affected by NPM reforms, it is also essential to explore perceptions at the individual level. This enables an increased understanding of the impact of NPM reforms on local conditions and practices, which should significantly affect not only the performances achieved within the system but also such things as the social and ideational aspects of academic work. How academics interact in a system where measures abound could drastically affect the basic preconditions for academic work, as could changes in the perception of what matters and what does not. Exploring these issues necessitates an interest in individual perceptions of performance measurement.

Disciplinary differences

Disciplinary differences among scholars are expressed in a number of ways, including the attitude toward performance measurement. Buela-Casal and Zych (2012) state there are significant variations between disciplines in the attitude toward research metrics. One common explanation for this is that prevailing research indicators are designed for the hard sciences but are also unreflectively used for the soft sciences. The consequence is that the performance of the former group is overemphasised compared with that of the latter (Donovan, 2007; Hicks, 2004). This is also the experience of many scholars in the humanities who argue that ‘bibliometrics do not fit with the purpose and rationale of research in the humanities’, as noted by Hammarfelt and de Rijcke (2015, p. 74). These authors see a misfit between disciplinary norms and external demands manifested in quantitative evaluation systems. However, they also see notable changes in the publication practices within the humanities, as academics are conforming to these measures.

An important reason why we should expect varying attitudes toward performance measurement between different disciplines is that prevalent research metrics have been developed specifically for some disciplines. As a result, the metrics are well adapted to the publication practices within these disciplines but may suit other disciplines quite poorly. Another reason for variations relates to the different needs of academics in different disciplines. As noted by Whitley (2007), researchers in the applied sciences often have an easier time finding funding from external sources than researchers in the pure sciences, which makes the latter group more tolerant of and responsive to evaluation systems of importance to these funders. This has also been illustrated by Reale and Seeber (2011), who show that different university departments are responsive to different stimuli, depending on their perceived need for financial or reputational resources. It should also be noted that teaching practices vary greatly between disciplines, which in turn affects teaching metrics, such as student ratings and retention rates (Neumann, 2001). However, whether discipline is a factor that affects the attitudes of academics toward performance measurement is a question that has rarely been quantitatively studied to any substantial degree.

Empirical Material and Data Analysis

This study explores data from a survey sent to academics and academic managers at universities in Norway, Sweden, Finland and Denmark. The survey is part of a larger comparative research project that studied the relation between reform, organisational change and performance in Nordic universities in the last couple of decades (see Pinheiro et al. 2019). The survey contains a large range of questions on themes including, but not limited to, organisational structures, performance management, incentives, funding arrangements, autonomy and control and local atmosphere. The data were collected in the fall of 2015 and the spring of 2016. A total of 5,489 respondents completed the survey (Norway n=1,340, Sweden n=701, Finland n=1,044, Denmark n=2,404), and the overall response rate was 15 percent. Because of varying HE systems and availability of sampling frames, the sampling was somewhat different in the four countries. The representativeness of the responses varies to some extent between the countries but is generally good with regard to institutional affiliation, discipline, seniority and gender. A closer look at disciplinary representativeness shows that social scientists are somewhat overrepresented, while agricultural scientists and medical and health scientists are slightly underrepresented. Other disciplinary groups are well represented.

To measure the attitudes of the respondents, they were asked to give their opinion about a number of statements. A 5-point Likert scale was used, with responses ranging from *strongly disagree* (1) to *strongly agree* (5). To visualise our results, we present the percentage of respondents who answered *agree* (4) or *strongly agree* (5) in relation to the statements in the survey. However, with this type of data it is hard to draw firm conclusions about the strength of the respondents’ attitudes. This is because we cannot assume a normal distribution of the answers regarding attitudes, which means that expectations are difficult to estimate. Therefore, our analysis mainly focuses on differences *between* the four disciplinary groups. The analysis employs non-parametric mean rank comparisons, where global differences between groups are tested using the Kruskal–Wallis H test (α : .05). If global differences are found, the groups are analysed pairwise using the Bonferroni post hoc test (α : .05). Some comparisons will also be made between the survey items. These comparisons will be made

using the Wilcoxon signed-rank test ($\alpha: .05$) and only for the whole sample of respondents, regardless of disciplinary affinity.

Measuring attitudes toward performance measurement

The attitudes toward performance measurement explored in this study take a number of perspectives into account, which is reflected in the survey design. It includes ideas about the accuracy of performance measures, reasons for measuring performance and the consequences thereof. It also includes notions about how performance measurement affects the working environment and the behaviour of the respondents, including their performance in teaching and research. The nine survey items and the dimensions they target are given in Table 3.

Table 3. Survey items

Survey items	Dimension
1. Internal procedures for measuring academic performance are in accordance with my understanding of academic performance	Validity
2. In my opinion, performance measurements increase transparency and fairness	Transparency
3. In my opinion, performance measurements are signs of mistrust	Mistrust
4. Control and evaluation of my work is a legitimate task	Legitimacy
5. Teaching performance measurements have a positive impact on the atmosphere surrounding academic work	Atmosphere
6. Research performance measurements have a positive impact on the atmosphere surrounding academic work	Atmosphere
7. Internal procedures for measuring academic performance have an impact on my decisions regarding academic work	Behaviour
8. Measurements increase my performance in teaching	Performance
9. Measurements increase my performance in research	Performance

Survey item 1 gauges the accuracy of performance measures. The purpose of the item is to assess how the respondents perceive the validity of performance measures, particularly whether the measures cover the vital aspects of teaching and research performance. This is an important and ongoing debate with regard to various performance measures (Gläser & Laudel, 2007; van Raan, 2005). Unlike many other studies of performance metrics, the question does not specify a particular type of performance measure but instead probes the perceived validity of performance measurement more generally. Survey items 2 and 3 explore attitudes toward the reasons for using performance measurement. An expected effect of performance measurement is to increase transparency and potentially fairness. In an academic setting, transparency and fairness may be seen as instrumental in distributing rewards to the rightful recipients (Aksnes & Rip, 2009). Survey item 2 gauges whether academics understand performance measures as promoting these goals. A related question is whether performance measurement is understood as a consequence of mistrust (Porter, 1995). From this perspective, the implementation of performance measures can be understood as a strategy to establish objective expectations and requirements for academic work. However, this may imply an overreliance on the metrics at the expense of professional judgement. Survey item 3 explores whether academics consider the use of performance measures to be motivated by mistrust.

Survey item 4 sums up the previous questions by exploring the perceived legitimacy of evaluating academic work. Legitimacy is important for any organisation or process, as it affects its potential impact; the more acceptance it has, the greater the possibility for effects (Deephouse & Suchman, 2008). Whether evaluations of academic work are understood as legitimate by the academics is therefore important. Here, we do not specifically ask about performance measurement but rather about control and evaluation. This may therefore be understood to capture a wider variety of tools to monitor and evaluate academic work, including qualitative assessments such as peer reviews. While this requires caution in interpreting the results, it may also function as a control to note whether potential differences between disciplines are specific to quantitative tools or if there are general patterns of acceptance and scepticism of evaluative tools.

An important question is whether performance measurement has any effect on academic practices. Performance measurement can stimulate positive competitiveness and provide incentives for

academics to increase their performance. However, it can also distort incentives to induce suboptimal behaviour and can have a negative effect on work motivation (Espeland & Sauder, 2007; Kallio & Kallio, 2014; Osterloh, 2010). Survey items 5 and 6 explore the perceived impact on the working environment within teaching and research, respectively. Similarly, survey item 7 gauges the perceived impact of performance measurement on behaviour. If performance measures do not affect behaviour, they are ineffective in steering academic work. This is important, as metrics are often understood mainly as management instruments rather than as indicators of academic quality (Söderlind & Geschwind, 2019). Because the effects on behaviour are self-reported, they are likely to be underestimated. However, the effects on behaviour do not necessitate performance effects. Survey item 8 and 9 explore the perceived performance effects for teaching and research, respectively. Again, the effects are self-reported, meaning it is difficult to observe disparities between perceived behaviour and actual behaviour. Therefore, a limitation of the present study is that no conclusions can be made about what academics do because it merely reports on the perceptions of academics. The performance effects of measuring academic work are thus beyond the scope of this article.

Results

As can be seen in Table 4, the main finding of our study is that there are significant differences between the disciplines with regard to how they perceive academic performance measurement. The result of the mean rank comparisons for each item is demonstrated by the Kruskal–Wallis H, which shows that all items indicate significant differences ($p < .05$) between the disciplinary groups. Academics in the applied sciences particularly stand out as being the most positive about performance measurement, and those in the humanities are the most negative, as shown by the Bonferroni post hoc tests. Academics in the natural sciences and the social sciences are generally positioned in between academics in the applied sciences and academics in the humanities. Most often, there is no significant difference between academics in the natural sciences and those in the social sciences. There is only a significant difference in their beliefs about whether performance measurement impacts the decisions they make in their academic work. Academics in the natural sciences think they are less affected by performance measures than academics in any other group. Using the Wilcoxon signed-rank test, we can also see whether performance measures have a greater impact on teaching compared to research. The results show there is no statistically significant difference between the effects of performance measurement on the atmosphere in teaching versus the atmosphere in research ($Z = .123$, $p = .902$). However, performance measurement is perceived to increase performance within research more than within teaching ($Z = 7.095$, $p = .000$).

Taking a closer look at the various dimensions included in the analysis, we can see that the disciplinary groups are quite homogenous in their perceptions of the validity of performance measures (item 1), but academics in the applied sciences differ by displaying a more positive attitude. Regarding whether performance measurement is understood to increase transparency and fairness (item 2), the results show that academics in the applied sciences agree to a greater extent than those in the other groups and that academics in the humanities disagree more. The reverse pattern is observed when the academics are asked whether measurements are signs of mistrust (item 3). Here, we see that academics in the humanities have a greater tendency to agree with the statement than academics in the other groups, and academics in the applied sciences have a lesser tendency. Whether control and evaluation are considered legitimate (item 4) differs between the disciplinary groups in basically the same way but with a minor difference. Academics in the applied sciences are more positive than those in the natural sciences and the humanities but do not differ statistically from those in the social sciences. Academics in the social sciences are more positive than those in the humanities, which means that academics in the humanities are more negative than those in the applied sciences and the social sciences.

In terms of the effects of performance measurement, there are statistically significant differences between the disciplines regarding the effects on the working environment, the behaviour of the academics and their performance. Concerning the impact on the work atmosphere, academics in the applied sciences report a stronger effect than academics in all the other disciplines, and academics in the humanities report a weaker effect. This is true for both teaching metrics (item 5) and research

metrics (item 6). However, whether performance measurement has an impact on academics' decision making (item 7) differs from the previous results. Here, only academics in the natural sciences stand out, as they agree to a lesser extent than academics in the other groups. How measures are perceived to impact the performance of academics also differs between the disciplines. Regarding teaching performance (item 8), academics in the applied sciences rate the statement higher than those in the social sciences and the humanities, but there is no difference in relation to academics in the natural sciences. Academics in the humanities score lower than those in all other groups. In terms of research performance (item 9), academics in the applied sciences agree more than those in the other groups, and those in the humanities disagree more.

Table 4. Academics' attitudes toward performance measurement by discipline

		N	% that agree (Likert 4-5)	Kruskal- Wallis H	p	Bonferroni post hoc tests, α: .05
1. Internal procedures for measuring academic performance are in accordance with my understanding of academic performance	Natural sciences	824	23.9	23.9	.000	Applied sciences are significantly higher than all other groups.
	Applied sciences	1,203	28.3			
	Social sciences	1,034	25			
	Humanities	602	20.4			
	Total	3,663	25.1			
2. In my opinion, performance measurements increase transparency and fairness	Natural sciences	913	31.9	60.1	.000	Applied sciences are significantly higher than all other groups, and humanities are significantly lower than all other groups.
	Applied sciences	1,324	40.7			
	Social sciences	1,109	30.6			
	Humanities	644	21.7			
	Total	3,990	32.8			
3. In my opinion, performance measurements are signs of mistrust	Natural sciences	912	30.6	58.6	.000	Humanities are significantly higher than all other groups, and applied sciences are significantly lower than all other groups.
	Applied sciences	1,321	26.9			
	Social sciences	1,112	34.8			
	Humanities	638	39.7			
	Total	3,983	32			
4. Control and evaluation of my work is a legitimate task	Natural sciences	884	49.1	26.8	.000	Applied sciences are significantly higher than natural sciences and the humanities, and social sciences are significantly higher than the humanities.
	Applied sciences	1,298	54.5			
	Social sciences	1,088	51			
	Humanities	647	43.3			
	Total	3,917	50.5			
5. Teaching performance measurements have a positive impact on the atmosphere surrounding academic work	Natural sciences	686	17.2	108	.000	Applied sciences are significantly higher than all other groups, and the humanities are significantly lower than all other groups.
	Applied sciences	987	18.4			
	Social sciences	890	12.4			
	Humanities	514	6.6			
	Total	3,077	14.4			
6. Research performance measurements have a positive impact on the atmosphere surrounding academic work	Natural sciences	733	15.4	113	.000	Applied sciences are significantly higher than all other groups, and the humanities are significantly lower than all other groups.
	Applied sciences	1,040	23.1			
	Social sciences	906	14.7			
	Humanities	519	7.7			
	Total	3,198	16.4			
7. Internal procedures for measuring academic performance have an impact on my decisions regarding academic work	Natural sciences	828	28.5	37.3	.000	Natural sciences are significantly lower than all other groups.
	Applied sciences	1,217	37.6			
	Social sciences	1,035	35.8			
	Humanities	603	34.7			
	Total	3,683	34.6			
8. Measurements increase my performance in teaching	Natural sciences	667	20.5	99.7	.000	The humanities are significantly lower than all other groups, and applied sciences are significantly higher than social sciences.
	Applied sciences	956	22			
	Social sciences	871	16.4			
	Humanities	509	8.4			
	Total	3,003	17.7			
9. Measurements increase my performance in research	Natural sciences	732	24.2	86	.000	Applied sciences are significantly higher than all other groups, and the humanities are significantly lower than all other groups.
	Applied sciences	1,023	29.4			
	Social sciences	888	24.9			
	Humanities	514	12.3			
	Total	3,157	24.1			

If we look at the degree to which academics agree with the statements, we can see that one fourth of respondents consider performance measures to be accurate (item 1), and one third think they increase transparency (item 2). Just 15 percent of the respondents think performance measures affect the work atmosphere positively (item 5 and 6). More than one third think performance measures impact their decisions (item 7); 18 percent of those in teaching experience positive effects on performance (item 8), as do 24 percent of those in research (item 9). Using the Wilcoxon signed-rank test, we can compare these differences and see that the perceived effect on decision making is significantly greater than the effect on performance in teaching ($Z = 18.912$, $p = .000$) and research ($Z = 14.515$, $p = .000$). This indicates that the academics alter their behaviour in ways that are perceived to have no effect on their performance. Although it is difficult to draw firm conclusions from these percentages without proper comparison points, they are interpreted as indicating a general scepticism about the statements. Despite this, the results show that evaluation and control are considered legitimate (item 4) by half of the respondents, and just a third think performance measurement is a sign of mistrust (item 3). Therefore, the attitudes toward performance measurement are not entirely negative.

Concluding Discussion

In this study, we investigated how disciplinary cultures affect academics' attitudes toward performance measurement at Nordic universities. We proceeded from an established typology of the academic disciplines, distinguishing the hard sciences from the soft and the applied sciences from the pure. Our results show that there are clear differences in the attitudes toward performance measurement between the disciplines. Thus, our study confirms the importance of previous calls to include the disciplinary perspective in analyses of HE (Becher, 1994). Disciplines affect cultures perhaps more than anything else in academia. Taking account of these differences is therefore paramount to understand how academic work is influenced by organisational changes, such as performance measurement.

Our main finding is that academics from the applied sciences generally have a more positive attitude regarding performance measurement than academics in any other group, and academics in the humanities generally have a more negative attitude. This pattern is consistent across most of our survey items, indicating that the finding is robust and that it describes a general attitude toward performance measurement. The survey items span dimensions such as the perceived validity, transparency and legitimacy of performance measures and whether they are seen to indicate mistrust. They gauge how academics experience the effect of performance measurement on the work atmosphere, their decision making in their job and their academic performance. Additionally, we note that the perceived legitimacy of evaluation and control demonstrates similar differences between the disciplines as the other survey items. This indicates that the observed differences in attitudes are not specific to performance measurement but are part of a larger pattern whereby academics in the applied sciences show the greatest acceptance of evaluative tools, and academics in the humanities are the most sceptical.

Previous studies have suggested that disciplinary cultures are important in explaining varying attitudes and behaviour among academics (Becher & Trowler, 2001; Biglan, 1973a; 1973b). Based on our results, we argue that this includes their attitudes toward performance measurement. Possible explanations for our findings include on the one hand the varying traditions of using academic performance measures and on the other hand varying accountability relationships. That performance measures have been developed specifically for disciplines within the hard sciences makes them better adapted to the particular practices that prevail there, and it makes academics more familiar with them. The soft sciences have eventually been included in analyses with these tools, despite the inadequate ability to accurately describe the research outputs (Donovan, 2007; Hammarfelt & de Rijcke, 2015; Hicks, 2004). With weaker traditions of performance measurement and inadequate validity, it is hardly surprising that the attitudes among the soft sciences are more sceptical. With regard to accountability relationships, it has been suggested that academics from the applied sciences are more sensitive to demands from external actors than academics from the pure sciences (Becher & Trowler, 2001; Whitley, 2007). As performance measures often make up the interface between external actors with limited capability to assess scientific research, metrics become important tools to convey success and

legitimise academic endeavours. Academics in the applied sciences are therefore expected to adhere more to performance measurement than academics in the pure sciences, as the former group depends more on support from external actors and may therefore be more accustomed to communicating about their work in similar terms.

While this study has provided empirical data to show how academics differ between the various disciplines, further research should heed Trowler's (2014) argument that macro-level categorisations of the disciplines often obscure the many differences within them. Future studies could advantageously explore these issues within the different disciplines. This would perhaps shed more light on the specific mechanisms giving rise to the disciplinary differences in the attitudes toward and experiences with academic performance measurement. Nevertheless, our results contribute to current debates on the impact of performance measurement in HE by highlighting the importance of disciplinary cultures and by providing empirical data on the impact of the disciplines. We hope that our study provides university actors a general understanding of the different perspectives and opinions regarding performance measurement, as it reflects the varying conditions under which academic work is conducted. This conclusion also serves as an important reminder to policymakers and managers interested in comparing academic performances across disciplines. Measurement systems that neglect these aspects are bound to meet resistance. Although there are statistical techniques to account for some of the measurement problems, such as the varying publication practices, it is also important to appreciate and consider the various attitudes toward performance measurement among academics. This is imperative, as any evaluation system will suffer without a reasonable level of acceptance among its subjects.

Ethical Note

In the national contexts of this study, no ethical clearance has been deemed necessary. Ethical considerations have continuously been made through the project, which has been conducted in agreement with The European Code of Conduct for Research Integrity, as defined by All European Academies (found here: <https://allea.org/code-of-conduct/>).

Funding

This work was supported by the Research Council of Norway under grant number 237782 and Riksbankens Jubileumsfond grant number FSK15-1059:1.

References

- Aagaard, K. (2015). How incentives trickle down: Local use of a national bibliometric indicator system. *Science and Public Policy*, 42(5), 725-737.
- Ahola, S., Hedmo, T., Thomsen, J.-P., & Vabø, A. (2014). *Organisational features of higher education: Denmark, Finland, Norway & Sweden*. Oslo: NIFU. Accessed (June 1, 2020): <https://www.nifu.no/en/publications/1162368>
- Aksnes, D. W., & Rip, A. (2009). Researchers' perceptions of citations. *Research Policy*, 38(6), 895-905.
- Amaral, A., Meek, L., & Larsen, I. M. (Eds.). (2003). *The higher education managerial revolution?*. Dordrecht: Springer.
- Becher, T. (1994). The significance of disciplinary differences. *Studies in Higher Education*, 19(2), 151-161.
- Becher, T., & Trowler, P. R. (2001). *Academic tribes and territories*. Buckingham: Open University Press.
- Biglan, A. (1973a). The characteristics of subject matter in different academic areas. *Journal of Applied Psychology*, 57(3), 195-203.

- Biglan, A. (1973b). Relationships between subject matter characteristics and the structure and output of university departments. *Journal of Applied Psychology*, 57(3), 204-213.
- Bleiklie, I. (1998). Justifying the evaluative state: New public management ideals in higher education. *Journal of Public Affairs Education*, 4(2), 87-100.
- Bornmann, L. (2014). Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. *Journal of Informetrics*, 8(4), 895-903.
- Buela-Casal, G., & Zych, I. (2012). What do the scientists think about the impact factor?. *Scientometrics*, 92(2), 281-292.
- Campbell, J. P., & Oblinger, D. G. (2007). Academic Analytics. *EDUCAUSE Quarterly*. Accessed (September 01, 2019): <https://library.educause.edu/-/media/files/library/2007/10/pub6101-pdf.pdf>
- Christensen, T. (2011). University governance reforms: potential problems of more autonomy?. *Higher Education*, 62(4), 503-517.
- Clark, B. R. (1998). *Creating entrepreneurial universities: Organizational pathways of transformation*. Bingley: Emerald.
- Cole, S. (1983). The hierarchy of the sciences?. *American Journal of Sociology*, 89(1), 111-139.
- Deem, R., Hillyard, S., Reed, M., & Reed, M. (2007). *Knowledge, higher education, and the new managerialism: The changing management of UK universities*. Oxford: Oxford University Press.
- Deephouse, D. L., & Suchman, M. (2008). Legitimacy in organizational institutionalism. In R. Greenwood, C. Oliver, R. Suddaby, & K. Sahlin-Andersson (Eds.), *The Sage Handbook of Organizational Institutionalism* (pp. 49-77). London: Sage.
- Donovan, C. (2007). The qualitative future of research evaluation. *Science and Public Policy*, 34(8), 585-597.
- El-Khawas, E. (1996). One professoriate, or many?: Assessing aspects of differentiation among academics. *Tertiary Education & Management*, 2(2), 146-152.
- Elzinga, A. (1997). The science-society contract in historical transformation: With special reference to "epistemic drift". *Social Science Information*, 36(3), 411-445.
- Elzinga, A. (2012). Features of the current science policy regime: Viewed in historical perspective. *Science and Public Policy*, 39(4), 416-428.
- Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, 113(1), 1-40.
- Ferlie, E., Musselin, C., & Andresani, G. (2008). The steering of higher education systems: A public management perspective. *Higher education*, 56(3), 325-348.
- Fanelli, D., & Glänzel, W. (2013). Bibliometric evidence for a hierarchy of the sciences. *PLoS One*, 8(6), e66938.
- Fägerlind, L., & Strömquist, G. (Eds.). (2004). *Reforming higher education in the Nordic countries: Studies of change in Denmark, Finland, Iceland, Norway and Sweden*. Paris: International Institute for Educational Planning. Accessed (June 1, 2020): <http://www.iiep.unesco.org/en/reforming-higher-education-nordic-countries-studies-change-denmark-finland-iceland-norway-and-sweden>
- Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (Eds.). (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. London: Sage.

- Gläser, J., & Laudel, G. (2007). The social construction of bibliometric evaluations. In R. Whitley & J. Gläser (Eds.), *The changing governance of the sciences: The advent of research evaluation systems*, (pp. 101-123). Dordrecht: Springer.
- Godin, B. (2006). The linear model of innovation: The historical construction of an analytical framework. *Science, Technology, & Human Values*, 31(6), 639-667.
- Gornitzka, Å., Stensaker, B., Smeby, J-C., & De Boer, H. (2004). Contract arrangements in the Nordic countries—solving the efficiency/effectiveness dilemma?. *Higher Education in Europe*, 29(1), 87-101.
- Gulbrandsen, M., & Slipersaeter, S. (2007). The third mission and the entrepreneurial university model. In A. Bonaccorsi & C. Daraio (Eds.), *Universities and strategic knowledge creation: Specialization and Performance in Europe* (pp. 112-143). Cheltenham: Edward Elgar Publishing.
- Hammarfelt, B., & de Rijcke, S. (2015). Accountability in context: Effects of research evaluation systems on publication practices, disciplinary norms, and individual working routines in the Faculty of Arts at Uppsala University. *Research Evaluation*, 24(1), 63-77.
- Hammarfelt, B., & Rushforth, A. D. (2017). Indicators as judgment devices: An empirical study of citizen bibliometrics in research evaluation. *Research Evaluation*, 26(3), 169-180.
- Hammarfelt, B., Nelhans, G., Eklund, P., & Åström, F. (2016). The heterogeneous landscape of bibliometric indicators: Evaluating models for allocating resources at Swedish universities. *Research Evaluation*, 25(3), 292-305.
- Hansen, H. F., Geschwind, L., Kivistö, J., Pekkola, E., Pinheiro, R., & Pulkkinen, K. (2019). Balancing accountability and trust: university reforms in the Nordic countries. *Higher Education*, 78(4), 557-573.
- Harvey, L. (2008). Rankings of higher education institutions: A critical review. *Quality in Higher Education*, 14(3), 187-207.
- Hedges, L. V. (1987). How hard is hard science, how soft is soft science? The empirical cumulativeness of research. *American Psychologist*, 42(5), 443-455.
- Henkel, M. (2000). *Academic identities and policy change in higher education*. London: Jessica Kingsley Publishers.
- Hicks, D. (2004). The four literatures of social science. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of Quantitative Science and Technology Research* (pp. 473-496). Dordrecht: Springer.
- Hood, C. (1991). A public management for all seasons?. *Public Administration*, 69(1), 3-19.
- Kallio, K. M., & Kallio, T. J. (2014). Management-by-results and performance measurement in universities—implications for work motivation. *Studies in Higher Education*, 39(4), 574-589.
- Kallio, K. M., Kallio, T. J., Tienari, J., & Hyvönen, T. (2016). Ethos at stake: Performance management and academic work in universities. *Human Relations*, 69(3), 685-709.
- Kuhn, T. (1962) [Reprinted 2012]. *The structure of scientific revolutions*. Chicago: The University of Chicago Press.
- Mingers, J., & Willmott, H. (2013). Taylorizing business school research: On the ‘one best way’ performative effects of journal ranking lists. *Human Relations*, 66(8), 1051-1073.
- Neumann, R. (2001). Disciplinary differences and university teaching. *Studies in Higher Education*, 26(2), 135-146.
- OECD. (2007). Revised field of science and technology (FOS) classification in the Frascati manual. Accessed (June 1, 2020): <https://www.oecd.org/science/inno/38235147.pdf>

- Osterloh, M. (2010). Governance by numbers. Does it really work in research?. *Analyse & Kritik*, 32(2), 267-283.
- Pinheiro, R., Geschwind, L., & Aarrevaara, T. (Eds.). (2016). *Mergers in higher education: The experience from Northern Europe*. Cham: Springer.
- Pinheiro, R., Geschwind, L., Hansen, H. F., & Pulkkinen, K. (Eds.). (2019). *Reforms, organizational change and performance in higher education: A comparative account from the Nordic Countries*. Cham: Palgrave Macmillan.
- Paradeise, C., Reale, E., Bleiklie, I., & Ferlie, E. (Eds.). (2009). *University governance: Western European comparative perspectives*. Dordrecht: Springer.
- Piwowar, H. (2013). Altmetrics: Value all research products. *Nature*, 493(7431), 159.
- Pollitt, C., & Bouckaert, G. (2004). *Public management reform: A comparative analysis*. Oxford: Oxford University Press.
- Porter, T. M. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton: Princeton University Press.
- Reale, E., & Seeber, M. (2011). Organisation response to institutional pressures in Higher Education: The important role of the disciplines. *Higher Education*, 61(1), 1-22.
- Slaughter, S., & Leslie, L. L. (1997). *Academic capitalism: Politics, policies, and the entrepreneurial university*. Baltimore: The Johns Hopkins University Press.
- Smith, L. D., Best, L. A., Stubbs, D. A., Johnston, J., & Archibald, A. B. (2000). Scientific graphs and the hierarchy of the sciences: A Latourian Survey of Inscription Practices. *Social Studies of Science*, 30(1), 73-94.
- Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, 83(4), 598-642.
- Stichweh, R. (2009). Differentiation of scientific disciplines: Causes and consequences. In G. H. Hadorn (Ed.), *Unity of knowledge in transdisciplinary research for sustainable development*, Volume 1 (pp. 82-90). Oxford: EOLSS Publishers/UNESCO.
- Stoecker, J. L. (1993). The Biglan classification revisited. *Research in Higher Education*, 34(4), 451-464.
- Stokes, D. E. (1997). *Pasteur's Quadrant: Basic science and technological innovation*. Washington, DC: Brookings Institution Press.
- Sugimoto, C. R., & Weingart, S. (2015). The kaleidoscope of disciplinarity. *Journal of Documentation*, 71(4), 775-794.
- Söderlind, J., & Geschwind, L. (2019). Making sense of academic work: The influence of performance measurement in Swedish universities. *Policy Reviews in Higher Education*, 3(1), 75-93.
- Trowler, P. (2014). Depicting and researching disciplines: Strong and moderate essentialist approaches. *Studies in Higher Education*, 39(10), 1720-1731.
- van Raan, A. F. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, 62(1), 133-143.
- van Vught, F., & Westerheijden, D. F. (2010). Multidimensional ranking. *Higher Education Management and Policy*, 22(3), 1-26.
- Whitley, R. (2007). Changing governance of the public sciences: The consequences of establishing research evaluation systems for knowledge production in different Countries and scientific fields. In R. Whitley & J. Gläser (Eds.), *The changing governance of the sciences: The advent of research evaluation systems* (pp. 3-27). Dordrecht: Springer.