AARHUS UNIVERSITY

Coversheet

This is the accepted manuscript (post-print version) of the article.

Contentwise, the post-print version is identical to the final published version, but there may be differences in typography and layout.

How to cite this publication

Please cite the final published version:

Christensen, J. (2018). Biased, Not Blind: An Experimental Test of Self-Serving Biases in Service Users' Evaluations of Performance Information. *Public Administration*, 96(3), 468-480. DOI: 10.1111/padm.12520

Publication metadata

Title:	Biased, Not Blind: An Experimental Test of Self-Serving Biases in Service Users' Evaluations of Performance Information
Author(s):	Julian Christensen
Journal:	<i>Public Administration</i> , 96(3), 468-480
DOI/Link:	https://doi.org/10.1111/padm.12520
Document version:	Accepted manuscript (post-print)

General Rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Biased, Not Blind: An Experimental Test of Self-Serving Biases in Service Users' Evaluations of Performance Information

Author: Julian Christensen, Aarhus University (julian@ps.au.dk).

ABSTRACT

Based on literature about motivated reasoning, this article proposes that choosing a public service provider from among competing options may bias service users in a positive direction when evaluating the performance of their chosen provider. Users are expected to defend their choice through processes of goal reprioritization, meaning that they will alter the weight they assign to given pieces of information depending on the (in)convenience of that information.

This article uses nine experimental studies to test this expectation on students who had recently chosen to study at one university instead of competing universities. As expected, findings show signs of biases in students' evaluations, but the biases are small and not consistently significant. Thus, prior research may have been too pessimistic regarding the general potential of performance information in the public sector.

INTRODUCTION

How do users of public services make sense of performance information? That question is essential for an assessment of the potential of performance information as a tool for performance improvements. Following recent years' explosion in performance measurement (Van Dooren, 2011), which has led scholars to label the beginning of the twenty-first century the 'era of governance by performance management' (Moynihan, 2008, p. 3) or even 'the age of quantified performance' (Van de Walle & Roberts, 2011, p. 215), this question is more relevant than ever before.

Performance information can be collected for many purposes (Behn, 2003), but most would agree that *the* major purpose of measuring performance is to make performance improvements (Ibid.). One prevalent strategy for pursuing this goal is to publish performance information while letting citizens choose freely from among competing providers of public services (Jilke, 2015). Thus, marketization efforts have given citizens all over the world the ability to choose between competing options when, for example, placing their child in a kindergarten, choosing a university for their education, or receiving treatment at a hospital, and people are free to leave in favor of competing providers if they are dissatisfied with the services they receive. In other words, citizens have been given an opportunity to hold organizations accountable by reacting through channels such as voice and exit in case of unsatisfactory performance (Dowding & John, 2007; Hirschman, 1970; James & Moseley, 2014; Paul, 1992). Consequently, decision-makers should be incentivized to react to poor performance (Jilke et al., 2016; Le Grand, 2010) by using performance information to make informed, performance-enhancing decisions.

However, the strategy of publishing performance information while giving service users free choice between competing providers can only work effectively as long as users respond systematically to performance information. Responsible decision-makers will

only face meaningful incentives to continuously maximize performance if the users' performance evaluations reflect the providers' actual performance (Stipak, 1979) as reported in the performance information. The psychology of information processing is therefore highly relevant to the study of performance information.

Recent psychologically informed literature has demonstrated how people's performance evaluations can often be distorted by cognitive (Andersen & Hjortskov, 2015; Olsen, 2013, 2015) and politically motivated biases (Baekgaard & Serritzlew, 2016; Baekgaard et al., 2017; Christensen et al., 2018; Hvidman & Andersen, 2016; James & Van Ryzin, 2017; Marvel, 2015, 2016). These findings of distorted performance evaluations all question the extent to which (dis)satisfaction among users of service providers can be seen as a meaningful reflection of the provider's actual performance.

This article aims at broadening and nuancing the literature about motivated biases with the potential to distort people's performance evaluations. Existing literature has done an excellent job pointing to the relevance of psychological insights when studying performance information (Grimmelikhuijsen et al., 2017; James et al., 2017), and there are studies that have contributed importantly by uncovering the distortive impact of *politically* motivated reasoning whenever performance information has political implications. This is, for example, relevant when people evaluate benchmark data about public and private providers of welfare services, in which case ideologically founded attitudes towards the size of the public sector become likely to distort people's evaluations (Baekgaard & Serritzlew, 2016; Baekgaard et al., 2017). It is also the case when highly contested policies are evaluated (James & Van Ryzin, 2017). However, while the distortive impact of political attitudes is well documented in existing literature, much less is known about the relevance of motivated biases in the absence of a link between political attitudes and performance information.

The last notion is important as a narrow focus on political biases overlooks the

fact that performance information is far from always politicized. For instance, instead of comparing public and private providers of a welfare service, it may be more relevant for a service user to compare neighboring public (or private) providers of the service while ignoring more distant private (or public) options. Findings of politically motivated reasoning have led researchers to fundamentally ‘question the potential of performance information’ as a performance-enhancing tool (Baekgaard & Serritzlew, 2016, p. 80), but such pessimism would not be warranted if the relevance of motivated biases is limited to politicized information.

The purpose of this article is thus to investigate the relevance of motivated biases in a case where political attitudes are not relevant to people’s evaluations. Focus will be on public service users’ evaluations of their own service provider’s performance in situations where they have chosen the provider from among competing options. As will be argued in the next section, there are theoretical reasons to expect that motivated (but choice-driven instead of politically driven) biases will be relevant in this case, but this has not hitherto been empirically investigated.

This article uses literature about motivated reasoning to propose that choosing a service provider will motivate service users to defend their choice. The users are thus expected to be biased in a positive direction when evaluating the performance of their own provider. Specifically, it is hypothesized that users will engage in goal reprioritization (Christensen et al., 2018), where they opportunistically alter the weight they assign to different pieces of performance information depending on the (in)convenience of that information. Users are expected to find performance information important if it shows their provider to perform well (compared to competing providers) while downplaying the importance of information showing their provider to perform less well.

This article uses nine experimental studies to test the hypothesis on samples of

university students who have recently chosen to pursue a given degree at a given university. Three conceptually similar survey experiments were run on three independent samples. The experiments were designed to overcome the challenge of identifying causal effects in a situation where self-selection is likely to cause endogeneity problems. Findings suggest that students are indeed biased when assessing the performance of their own universities, but the results also show that the students are far from blind. Specifically, all nine tests show signs of self-serving goal reprioritization, but the effect sizes are negligible and often statistically insignificant. Thus, there are clear limits to people's biases, and prior research may have been too pessimistic regarding the general potential of performance information.

HOW CHOICES MAY MOTIVATE BIASED PERFORMANCE EVALUATIONS

The idea of performance information and choice as performance-enhancing tools for accountability is appealing from an intuitive point of view. Citizens (the end users of public services) are given a great deal of power as they can choose providers who fit them well and leave if they become dissatisfied. In turn, managers and employees in public organizations should be incentivized to maximize the quality of the services they deliver in order to keep their service users and attract new ones.

However, while this seems plausible from an intuitive point of view, a prerequisite for the strategy to work effectively is that users of public services do actually update their performance evaluations in cases where performance data gives reason to do so. In other words, in order for users of public services to react to poor performance, they must acknowledge the poor performance and be ready to recognize that they have made a wrong choice of service provider. According to literature on motivated reasoning, the realization of having made a wrong choice is in conflict with some basic psychological drives.

Psychologists have long recognized that motivation has implications for the processes by which people approach information in order to form and update knowledge. Motivation has implications for what issues people devote attention to, for the amount of effort they spend on information processing and for the mindset with which they approach information (Kruglanski & Webster, 1996). Sometimes, people are motivated to engage in thorough processing of a great deal of information in order to make correct judgments about issues at hand. When this is the case, people tend to approach information with an open mind and tend to be reluctant to commit to premature and definite conclusions (Ibid.: 264). This open-minded way of reasoning is also known as ‘accuracy-motivated reasoning’ (Kunda, 1990). However, there are also situations in which people are motivated to preserve, or ‘freeze’ on, certain judgments. In these situations, people seek to avoid new information that might force them to update their judgments and tend to approach information with a closed-minded defense of the conclusions they want to preserve (Kruglanski & Webster, 1996: 264). Many scholars use the term ‘directionally motivated reasoning’ to describe such closed-minded defenses of preselected conclusions (Kunda, 1990).

The insights above are central to literature on politically motivated reasoning where the focus is on the distortive impact of political identities and attitudes on people’s processing of information with political implications (Baekgaard & Serritzlew, 2016; Baekgaard et al., 2017; Bisgaard, 2015; Groenendyk, 2013; James & Van Ryzin, 2017; Kahan et al., 2017; Taber & Lodge, 2006). However, factors other than political identities and attitudes can also motivate closed-minded defenses of preselected conclusions. More generally, Kruglanski and Webster (1996) argue that people’s motivations are ‘determined by perceived benefits or costs of closed or open states as influenced by situational, cultural, or personality factors’ (Ibid.: 267).

A reoccurring theme in the broader literature on motivated reasoning is how choices tend to impact the way people process choice-relevant information. Before choices are made, people tend to approach information with different motivations than they do after the choices have been made. In predecisional phases, people are focused on making good choices (Gollwitzer, 1990: 62) and the benefits of an open mind are made salient by concerns about the risk of a judgmental mistake (Kruglanski & Webster, 1996: 264). However, in postdecisional phases, people tend to be motivated to avoid questioning the desirability of the choices that have been made (Gollwitzer, 1990: 65). The realization of having made a wrong choice is unpleasant to such an extent that people tend to enter a closed state of mind, seeking to defend their choices (Cooper, 2007: 10). As a result, there will often be a crystallization effect of choices, where people engage in self-serving reasoning and freeze on the choices that have been made (Brehm, 1956; Cooper, 2007; Festinger, 1957; Shultz et al., 1999).

An effective reasoning strategy that people can (consciously or subconsciously) use to remain confident in their choices is goal reprioritization (Christensen et al., 2018), where people retrospectively reweight the perceived importance of information depending on the (in)convenience of the information (Christensen et al., 2018; Festinger, 1957; Groenendyk, 2013; Simon et al., 1995). This implies that, following a choice between competing options, people will seek to magnify the perceived importance of advantages associated with their choice (relative to alternative options) while downplaying the importance of relative disadvantages associated with that choice. Compared to more blatant forms of directionally motivated reasoning, such as ignoring or even misinterpreting information (Baekgaard & Serritzlew, 2016; Kahan et al., 2017), goal reprioritization is psychologically appealing as it allows people to defend preselected conclusions while maintaining an ‘appearance of objectivity’ (Groenendyk, 2013: 50). Thus, people leverage

the fact that information is often ambiguous (Moynihan, 2008), which makes it possible to come up with reasonable arguments in support of choices having been made.

The theory above has implications in the context of public services whenever citizens can choose between competing service providers. Based on the theory, performance information may work well in terms of assisting people's choice of provider (that is, in the predecisional phases). However, the theory expects that choosing a provider will motivate people to engage in self-serving goal reprioritization in defense of that provider. Thus, in the postdecisional phases, people are expected to alter their importance evaluations of performance information in ways that allow them to find information important if their chosen provider is shown to perform well. In contrast, people are expected to downplay the importance of information showing their provider performing less well. This article's analysis will test for such postdecisional biases, and hypothesizes that:

Hypothesis: *Service users find performance indicators more important when their chosen service provider performs well on the indicators (compared to competing providers) and less important when their chosen provider performs less well.*

DESIGN AND DATA COLLECTION

A main concern related to testing this article's hypothesis is the difficulty of isolating signs of self-serving goal reprioritization from systematic but endogenous differences in importance evaluations resulting from self-selection (Blom-Hansen et al., 2015). Following the theory above, people are expected to find performance indicators important if their chosen public service provider performs well on the indicators and less important if the provider performs less well. However, such systematic differences in

importance evaluations may emerge through different processes, with very different theoretical implications.

One possibility is that people do indeed re-weigh performance indicators in response to performance information's (dis)conformity with their desire to defend their choice of provider. But similar differences in importance evaluations may also emerge through processes where people self-select into providers that perform well on indicators they find important. In that case, there would be reason to be optimistic with regard to performance information as a decision-improving tool.

In order to handle this analytical challenge and test the article's hypothesis properly, we will not only need data on how people evaluate the importance of performance indicators on which their chosen providers perform well/poorly. We will also need to know how people would evaluate the same information in a counterfactual situation where the same providers perform poorly/well on the same indicators. This has important design implications as it would not be possible to collect useful counterfactual data with an observational design. In order to test the article's hypothesis, it was therefore decided to run survey experiments where users of a public service were asked to evaluate the importance of constructed performance information regarding their own service provider and a competing provider. Thus, it was randomized whether the respondents' provider was shown to perform better or worse than the competing provider.

Three data collections were made, each including three independently randomized experiments designed to test the article's hypothesis on first-semester undergraduate students in Political Science at Aarhus University. This pool of respondents is advantageous as it consists of public service users (students) who have recently made a choice between competing and highly comparable service providers (universities) and who

should therefore, following the theory above, be in the midst of a self-serving defense of their choice of university and degree program.

In each experiment, respondents were asked to evaluate the importance of a specific piece of information about the performance of the Political Science program at Aarhus University compared to the performance at a competing university. Degrees in Political Science are offered at three universities in Denmark (Aarhus University, University of Copenhagen, and University of Southern Denmark). Tuition is free at all universities and citizens with a qualifying secondary education can apply for enrollment in any degree program at whichever university they choose. Thus, for the sake of generalizability, some participants were asked to evaluate information about Aarhus University compared to University of Copenhagen whereas others were asked to evaluate information about Aarhus University compared to University of Southern Denmark.¹

The article's experiments were embedded in surveys that were distributed to all students attending lectures in a first-semester Methods course on 28 September 2015, 12 September 2016², and 11 September 2017. Participants had applied for enrollment in Aarhus University's Bachelor's program in Political Science between March and July, and when data was collected in September, they had been enrolled in the program for about a month. Respondents were given time to answer the surveys in class and the number of Political Science students³ who provided answers to the experimental items was 234 in 2015, 232 in 2016, and 238 in 2017 (out of 275 students who were enrolled in Aarhus University's

¹ In the 2015 data collection, University of Copenhagen (CPH) was used as the comparison university; in the 2016 data collection, University of Southern Denmark (SDU) was used as the comparison university; and in the 2017 data collection, simple randomization was used to assign respondents to information using CPH or SDU as the comparison university.

² The 2016 data collection was preregistered on 30 August 2016 using the *Open Science Framework*. The preregistered analysis plan is publicly available on <https://osf.io/uwf3s/register/565fb3678c5e4a66b5582f67>.

³ Lectures are open to the public and are sometimes attended by students from programs other than Political Science. The theoretical expectations of the article do not apply to these people and thus respondents who were not studying Political Science at the time of the data collection were excluded from the analyses.

Bachelor's program in Political Science each year). A total of 89 respondents⁴ indicated that they had not chosen Political Science at Aarhus University as their first priority when applying for enrollment in a degree program. They had preferred another program and/or another university but their grades had not qualified them for enrollment in their preferred program. The purpose of this article is to investigate biases in students' performance evaluations when they have chosen their own degree program and university and thus the article's hypothesis is not expected to apply to students who did not choose their program themselves. Consequently, in accordance with the preregistered analysis plan, these respondents are excluded from the article's main analyses below, and the analyses are thus based on answers from 615 respondents. As can be seen in Appendix A1, however, including the students who did not choose Political Science at Aarhus University as their first priority does not lead to substantively altered results, although the effects are somewhat weakened.

The experiments were included in the surveys after a separate page with the following introduction text:

'On the next pages you will be asked to respond to information scenarios, all concerning the performance of the Department of Political Science at Aarhus University compared to the Department of Political Science at [COMPARISON UNIVERSITY].

Please take a look at the information and evaluate how important you find *each separate piece of information* for an *overall evaluation* of the quality of education at the two departments.'

⁴ 31 respondents in 2015, 37 respondents in 2016, and 21 respondents in 2017 had not chosen Political Science at Aarhus University as their first priority.

After this introduction page, the experiments were introduced in a randomized order. In each of the three experiments, respondents were asked to evaluate the importance of a performance indicator. They were presented with a table with constructed information⁵ on the proportion of Political Science students at Aarhus University and a comparison university who were dissatisfied with the supply of social events at their university (Experiment 1), the proportion of students who were dissatisfied overall with their education (Experiment 2), and the proportion of students who thought that they did not receive enough teaching (Experiment 3). Question wordings for each of the experiments are reported in Figure 1 below.⁶

The performance indicators were chosen to resemble indicators on which Danish universities are regularly measured in publicly available performance reports. Respondents saw the same comparison university in all three experiments but the respondent's treatment status (which university was shown to perform best) was independently randomized in each experiment.⁷ Each experiment was introduced on a separate page in the survey and information was always visible to the respondents while evaluating its importance.

- Figure 1 about here -

⁵ On the introduction page, respondents were told that they would be asked to respond to 'information scenarios,' but this was the only indication that the information was not real. Respondents received an oral debriefing in class, just after answering the survey, where they were told that the information had been fictional. Furthermore, a written debriefing was distributed to the respondents in which the purpose of the study was explained and links to results of real satisfaction surveys were provided. This procedure was approved by the Study Director at Aarhus University's Department of Political Science.

⁶ Figure 1 shows question wordings with University of Copenhagen as the comparison university, but the wording was identical in the other version of the survey with the exception that University of Southern Denmark was shown instead of University of Copenhagen (for exact wording of the experiments using University of Southern Denmark as the comparison university, see online supporting information S9).

⁷ Balance tests in online supporting information S1-S3 do not give reason to question the randomization. Imbalances exist with regard to age in the 2015 Experiment 1, with regard to mother's education in the 2017 Experiments 1-2, and with regard to political attitudes in the 2017 Experiment 2, but with a total of 66 tests this is unsurprising. Furthermore, additional analyses (not reported but available upon request) show no moderating impact of these variables on the experimental effects.

ANALYSIS

It is described above how three rounds of data collection were made, each including three different but highly related experiments to test the article's hypothesis on independent samples of university students. Figure 2 summarizes importance evaluations in all of the article's experiments and reports the results of two-tailed t-tests for statistically significant differences between importance evaluations in the experimental conditions.

- Figure 2 about here -

According to the article's hypothesis, respondents were expected to engage in goal reprioritization, where they would opportunistically alter the weight they assign to performance indicators such that they would find indicators more important when their own university was shown to perform best. In contrast, they were expected to downplay the importance of performance indicators if another university was shown to perform best. The results in Figure 2 are generally but weakly supportive of this expectation. In all nine tests, participants did indicate higher average levels of perceived importance of information when Aarhus University was shown to perform best compared to the scenario where a competing university was shown to perform best. Effects are statistically significant in one test (Experiment 1 in 2015 where the effect of Aarhus University performing best is a +0.85 increase in perceived importance, $p = 0.015$) and marginally significant in two tests (Experiment 1 in 2016 where the effect is +0.60, $p = 0.095$ and Experiment 3 in 2015 where the effect is +0.54, $p = 0.063$). The six remaining tests show statistically insignificant effects. Turning to the substantive magnitude of the effects, the effect of Aarhus University performing best is less than one point on the 11-point importance rating scale in all of the tests in Figure 2, and Cohen's d varies between 0.24 and 0.34 for the statistically significant

and marginally significant effects. While this is not unusual in social science research (Murnane & Willett, 2011: 92), these are quite small effect sizes (Cohen, 1988).

In addition to the analyses in Figure 2, explorative analyses are reported in Figure 3 below, where answers from the three years are pooled⁸ in order to increase statistical power. Each of the three samples are quite small, which, in combination with the small effect sizes, may explain the many statistically insignificant results in Figure 2. Power analyses (see online supporting information S8) reveal that Experiment 1 (with an estimated power of 0.96-0.97) is the only experiment which is well-powered using the three individual samples. Pooling the data increases the estimated power of Experiment 2 from an average of 0.26 in the individual samples to 0.63, while the estimated power of Experiment 3 increases from an average of 0.57 in the individual samples to 0.96. However, as can be seen in Figure 3, maximizing power by pooling the data does not lead to substantively altered conclusions. There is now a small positive, statistically significant effect of Aarhus University performing best in Experiment 1 (+0.68 on the 11-point importance rating scale; Cohen's $d = 0.26$; $p = 0.001$) and a marginally significant effect in Experiment 3 (+0.37 on importance rating scale; Cohen's $d = 0.15$; $p = 0.06$), whereas the effect remains insignificant in Experiment 2 ($p = 0.18$).⁹

- Figure 3 about here -

⁸ Additional analyses show no statistically significant differences in effects across data collections or across comparison universities (see online supporting information S4-S7).

⁹ Note that with an estimated power of 0.63, even the pooled version of Experiment 2 is rather under-powered. With an average effect of 0.23 on the 11-point importance rating scale and an average standard deviation of 2.49, 922 respondents would be needed in order to reach an estimated power of 0.8.

DISCUSSION AND CONCLUSION

This article contributes to a growing literature about motivated biases with the potential to distort people's processing of performance information. The purpose of the article was to investigate the relevance of motivated biases in a case where political attitudes are not relevant to people's evaluations. Prior literature has had an exclusive focus on politicized information, and findings of politically motivated reasoning have led researchers to be pessimistic about the performance-enhancing potential of performance information. It was argued in the introduction that such pessimism would not be warranted if the relevance of motivated biases were limited to politicized performance information.

The results of the analysis can be viewed from a pessimistic perspective and from an optimistic one. Based on insights from the literature on motivated reasoning, the article hypothesized that public service users will tend to engage in self-serving biases when evaluating information about their own chosen service provider. Service users were expected to engage in goal reprioritization, meaning that they would alter the weight they assign to individual pieces of information such that they would find information important if their provider was shown to perform well while downplaying the importance of information showing their provider to perform less well.

The general patterns in the article's results are consistent with this hypothesis, meaning that the findings overall are supportive of the existence of self-serving biases when students evaluate information about the performance of their own university versus competing universities. An important contribution of the article is thus to show that the relevance of motivated biases is *not* limited to politicized performance information. Pessimistic readers may use this to question mechanisms by which performance information is meant to improve public performance. Performance information can only be effective in

improving public performance if people react meaningfully to the information. If people's evaluations are distorted, their reactions are likely to be so as well.

However, this pessimistic reading of the analysis ignores important nuances in the results. From a more optimistic point of view, it is striking how stable respondents are in their importance evaluations. Effect sizes are small in all tests and even maximizing the studies' power by pooling data does not lead to consistently significant results. Thus, while students do indeed seem to be motivated to defend their chosen university, this motivation does not blind the students when evaluating their university's performance. In that sense, there is reason to be optimistic with regard to the performance-enhancing perspectives of performance information.

This article has taken an important step in nuancing our understanding of motivated biases in relation to performance information, but some limitations should be acknowledged. First, the experiments in this article used fictional performance information in order to create exogenous variation in the information. This made it possible to isolate signs of self-serving goal reprioritization from endogenous differences in importance evaluations, but at the same time, it might have weakened the experimental treatment as some respondents may have had prior knowledge about the universities' real performance. Thus, future research might benefit from developing designs using real information. Second, the article's surveys did not ask respondents about the degree to which they had actively considered attending the two comparison universities when choosing to study at Aarhus University. Such information could have been relevant as people's motivation to engage in self-serving reasoning may vary with the extent to which their choice of Aarhus University was an active rejection of the other universities. Future research is therefore encouraged to better distinguish between people based on the nature of their choice of service provider. Third, the article's experiments were not designed to demonstrate the causal mechanisms behind the biases under investigation.

The experiments allowed for the investigation of biases in a case where students had recently made a choice of university, but we cannot know for sure that this choice was the causal reason behind the biases. Future research is encouraged to address this shortcoming, for example by studying people's evaluations before and after choices are made or by varying access to choice in laboratory experiments.

A final issue that should be considered is the generalizability of the article's results. Can the results teach us anything about public service users more generally or are they limited to the case of university students? This is ultimately an empirical question, but some theoretically informed observations can be made.

For instance, the article's hypothesis was based on literature on choices tending to motivate biased reasoning. However, while choice has been introduced in relation to many public services all over the world, there are also services where citizens cannot choose between competing providers. Sometimes, services are too specialized for multiple providers to exist, and sometimes citizens are automatically assigned to certain providers, e.g. based on their address. This article does not have the ambition to make claims about cases where citizens cannot choose between competing providers. Other sources of bias may exist in these situations (Baekgaard & Serritzlew, 2016; James & Van Ryzin, 2017), but this article's choice-driven hypothesis should not be expected to apply in the absence of choice.

This still leaves open the question of generalizability to other services where users do have a choice between competing providers. Kruglanski and Webster's theory is relevant in relation to this question as they argue that people's approach to information will be determined by the perceived costs and benefits of a closed or open mind (Kruglanski & Webster, 1996: 267). This article's theory expected that choosing a certain provider should increase the psychological costs of an open mind, but the extent to which this is the case may vary from service to service. In fact, three factors make it likely that students are more

motivated to defend their choice than most other service users. First, educational choices have large effects on people's lives (Brewer et al., 1999; Card, 1999), meaning that students should be especially motivated to convince themselves that they have made a good choice. Second, compared to other services, education tends to be quite identity-defining. Students are not 'just' users of a university. They are *part of* their university and are therefore likely to develop a sense of 'belongingness' to and organizational identification with their university (Mael & Ashforth, 1992), meaning that there is also an identity-protective aspect of students' motivation to defend their choice. And finally, when students enroll in a degree program, it is likely that they will settle in the city where their university is located, make friends, start in student jobs, and so on. Therefore, although students may in theory be free to move to competing universities in cases of dissatisfactory performance, the transaction costs of doing so would be high, adding to the inconvenience of realizing that one's educational choice was not good.

Based on these arguments, there is reason to be even more optimistic with regard to the decision-improving potential of performance information. The case of university students may be considered most likely in terms of the article's hypothesis, meaning that choice-driven biases can be expected to be even less pronounced in relation to less identity-defining services with smaller effects on people's lives and with lower transaction costs associated with moving to competing providers.

It is clear from the discussion above that a great deal of work remains to be done in order to fully understand the role of motivated biases in the context of Public Administration. This article's analysis shows that motivated biases are relevant, even in relation to non-politicized performance information. However, the analysis also shows that there are clear limits to people's biases and thus it is likely that findings of politically motivated reasoning have led researchers to be too pessimistic regarding the performance-

enhancing potential of performance information. While motivated biases may fundamentally call into question the performance-enhancing potential of performance information in some contexts, it seems that these biases are less challenging in other contexts. Future research should focus on investigating the conditions under which biases are most – and least – severe.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article. Furthermore, a replication dataset and a code to reproduce the article's results is available at the Harvard Dataverse repository: <https://doi.org/10.7910/DVN/EH0D15>.

ACKNOWLEDGMENTS

I would like to thank Morten Hjortskov Larsen, Martin Baekgaard, Lene Aarøe, Søren Serritzlew, Asmus Leth Olsen, Jeanette Frank Nielsen, and the two anonymous reviewers for a lot of helpful inputs to this paper. Furthermore, I would like to thank the 2015, 2016, and 2017 cohorts of Political Science students at Aarhus University for the many null results in the paper – these are good news for performance information in the public sector!

REFERENCES

- Andersen, S. C., & Hjortskov, M. (2015). Cognitive Biases in Performance Evaluations. *Journal of Public Administration Research and Theory*, 26(4), 647–662.
<https://doi.org/10.1093/jopart/muv036>
- Baekgaard, M., Christensen, J., Dahlmann, C. M., Mathiasen, A., & Petersen, N. B. G. (2017). The Role of Evidence in Politics: Motivated Reasoning and Persuasion among Political Decision-Makers. *British Journal of Political Science*.
<https://doi.org/https://doi.org/10.1017/S0007123417000084>
- Baekgaard, M., & Serritzlew, S. (2016). Interpreting Performance Information: Motivated Reasoning or Unbiased Comprehension. *Public Administration Review*, 76(1), 73–82.
<https://doi.org/10.1111/puar.12406>
- Behn, R. D. (2003). Why Measure Performance? Different Purposes Require Different

- Measures. *Public Administration Review*, 63(5), 586–606.
<https://doi.org/doi.org/10.1111/1540-6210.00322>
- Bisgaard, M. (2015). Bias Will Find a Way: Economic Perceptions, Attributions of Blame, and Partisan-Motivated Reasoning during Crisis. *The Journal of Politics*, 77(3), 849–860. <https://doi.org/10.1086/681591>
- Blom-Hansen, J., Morton, R., & Serritzlew, S. (2015). Experiments in Public Management Research. *International Public Management Journal*, 18(2), 151–170.
<https://doi.org/10.1080/10967494.2015.1024904>
- Brehm, J. W. (1956). Postdecision changes in the desirability of alternatives. *The Journal of Abnormal and Social Psychology*, 52(3), 384–389. <https://doi.org/10.1037/h0041006>
- Brewer, D. J., Eide, E. R., & Ehrenberg, R. G. (1999). Does it pay to attend an elite private college? Cross-cohort evidence on the effects of college type on earnings. *The Journal of Human Resources*, 34(1), 104–123. <https://doi.org/10.2307/146304>
- Card, D. (1999). The causal effect of education on earnings. *Handbook of Labor Economics*, 3, 1801–1863. [https://doi.org/10.1016/S1573-4463\(99\)03011-4](https://doi.org/10.1016/S1573-4463(99)03011-4)
- Christensen, J., Dahlmann, C. M., Mathiasen, A., Moynihan, D. P., & Petersen, N. B. G. (2018). How do Elected Officials Evaluate Performance? Goal Preferences, Governance Preferences and the Process of Goal Reprioritization. *Journal of Public Administration Research and Theory*, 28(2), 197–211. <https://doi.org/doi.org/10.1093/jopart/muy001>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd Edition)* (2nd ed.). Hillsdale, N.J.: L. Erlbaum.
- Cooper, J. (2007). *Cognitive Dissonance: Fifty Years of a Classic Theory*. SAGE Publications Ltd.
- Dowding, K., & John, P. (2007). The Three Exit , Three Voice and Loyalty Framework : A Test with Survey Data on. *Political Studies*, 56(2), 288–311.
<https://doi.org/10.1111/j.1467-9248.2007.00688.x>
- Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.
- Gollwitzer, P. M. (1990). Action Phases and Mind-Sets. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of Motivation and Cognition: Foundations of Social Behavior, Volume 2* (pp. 53–92). New York: The Guilford Press.
- Grimmelikhuijsen, S., Jilke, S., Olsen, A. L., & Tummers, L. (2017). Behavioral Public Administration: Combining Insights from Public Administration and Psychology. *Public Administration Review*, 77(1), 45–56. <https://doi.org/10.1111/puar.12609>
- Groenendyk, E. W. (2013). *Competing Motives in the Partisan Mind: How Loyalty and Responsiveness Shape Party Identification and Democracy*. New York: Oxford University Press.
- Hirschman, A. O. (1970). *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*. Cambridge, MA.: Harvard University Press.
- Hvidman, U., & Andersen, S. C. (2016). Perceptions of Public and Private Performance:

- Evidence from a Survey Experiment. *Public Administration Review*, 76(1), 111–120. <https://doi.org/10.1111/puar.12441>
- James, O., Jilke, S. R., & Van Ryzin, G. G. (2017). Behavioural and experimental public administration: Emerging contributions and new directions. *Public Administration*, 95(4), 865–873. <https://doi.org/10.1111/padm.12363>
- James, O., & Moseley, A. (2014). Does Performance Information About Public Services Affect Citizens' Perception, Satisfaction, and Voice Behaviour? Field Experiments With Absolute and Relative Performance Information. *Public Administration*, 92(2), 493–511. <https://doi.org/10.1111/padm.12066>
- James, O., & Van Ryzin, G. G. (2017). Motivated Reasoning about Public Performance: An Experimental Study of How Citizens Judge Obamacare. *Journal of Public Administration Research and Theory*, 27(1), 197–209. <https://doi.org/10.1093/jopart/muw049>
- Jilke, S. (2015). *Essays on the Microfoundations of Competition and Choice in Public Service Delivery*. Erasmus University Rotterdam. Retrieved from https://sebastianjilke.files.wordpress.com/2015/03/dissertation_sj1.pdf
- Jilke, S., Van Ryzin, G. G., & Van De Walle, S. (2016). Responses to Decline in Marketized Public Services : An Experimental Evaluation of Choice Overload. *Journal of Public Administration Research and Theory*, 26(3), 421–432. <https://doi.org/10.1093/jopart/muv021>
- Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated Numeracy and Enlightened Self-Government. *Behavioural Public Policy*, 1(1), 54–86. <https://doi.org/doi.org/10.1017/bpp.2016.2>
- Kruglanski, A. W., & Webster, D. M. (1996). Motivated Closing of the Mind : “Seizing” and “Freezing.” *Psychological Review*, 103(2), 263–283. <https://doi.org/10.1037//0033-295X.103.2.263>
- Kunda, Z. (1990). The Case for Motivated Reasoning. *Psychological Bulletin*, 108(3), 480–498. <https://doi.org/http://dx.doi.org/10.1037/0033-2909.108.3.480>
- Le Grand, J. (2010). Knights and knaves return: public service motivation and the delivery of public services. *International Public Management Journal*, 13(1), 56–71. <https://doi.org/10.1080/10967490903547290>
- Mael, F., & Ashforth, B. E. (1992). Alumni and their alma mater : A partial test of the reformulated model of organizational identification. *Journal of Organizational Behavior*, 13(2), 103–123. <https://doi.org/https://doi.org/10.1002/job.4030130202>
- Marvel, J. D. (2015). Public Opinion and Public Sector Performance: Are Individuals' Beliefs About Performance Evidence-Based or the Product of Anti-Public Sector Bias? . *International Public Management Journal*, 18(2), 209–227. <https://doi.org/10.1080/10967494.2014.996627>
- Marvel, J. D. (2016). Unconscious Bias in Citizens' Evaluations of Public Sector Performance. *Journal of Public Administration Research and Theory*, 26(1), 143–158. <https://doi.org/10.1093/jopart/muu053>

- Moynihan, D. P. (2008). *The Dynamics of Performance Management: Constructing Information and Reform*. Washington D.C.: Georgetown University Press.
- Murnane, R. J., & Willett, J. B. (2011). *Methods Matter: Improving Causal Inference in Educational and Social Science Research*. New York: Oxford University Press.
- Olsen, A. L. (2013). Leftmost-digit-bias in an enumerated public sector ? An experiment on citizens ' judgment of performance information. *Judgment and Decision Making*, 8(3), 365–371.
- Olsen, A. L. (2015). Citizen (Dis)satisfaction: An Experimental Equivalence Framing Study. *Public Administration Review*, 75(3), 469–478. <https://doi.org/10.1111/puar.12337>.
- Paul, S. (1992). Accountability in public services: Exit, voice and control. *World Development*, 20(7), 1047–1060. [https://doi.org/10.1016/0305-750X\(92\)90130-N](https://doi.org/10.1016/0305-750X(92)90130-N)
- Shultz, T. R., Léveillé, E., & Lepper, M. R. (1999). Free Choice and Cognitive Dissonance Revisited: Choosing “Lesser Evils” Versus “Greater Goods.” *Personality and Social Psychology Bulletin*, 25(1), 40–48.
<https://doi.org/https://doi.org/10.1177/0146167299025001004>
- Simon, L., Greenberg, J., & Brehm, J. (1995). Trivialization: The Forgotten Mode of Dissonance Reduction. *Journal of Personality and Social Psychology*, 68(2), 247–260. <https://doi.org/10.1037/0022-3514.68.2.247>
- Stipak, B. (1979). Citizen Satisfaction With Urban Services : Potential Misuse as a Performance Indicator. *Public Administration Review*, 39(1), 46–52.
<https://doi.org/10.2307/3110378>
- Taber, C. S., & Lodge, M. (2006). Motivated Skepticism in the Evaluation of Political Beliefs. *American Journal of Political Science*, 50(3), 755–769.
<https://doi.org/https://doi.org/10.1111/j.1540-5907.2006.00214.x>
- Van de Walle, S., & Roberts, A. (2011). Publishing Performance Information: An Illusion of Control? In W. Van Dooren & S. Van de Walle (Eds.), *Performance Information in the Public Sector: How it is Used* (pp. 215–230). New York: Palgrave Macmillan.
- Van Dooren, W. (2011). Nothing New Under the Sun? Change and Continuity in the Twentieth-Century Performance Movements. In W. Van Dooren & S. Van de Walle (Eds.), *Performance Information in the Public Sector: How it is Used* (pp. 15–27). New York: Palgrave Macmillan.

FIGURES

Figure 1: Question wording in experiments (comparison university = Copenhagen)

“Please take a look at table below:

Experiment 1

	Aarhus	Copenhagen
Proportion of Political Science students who are dissatisfied with the supply of social events	20 pct	25 pct

Or

	Aarhus	Copenhagen
Proportion of Political Science students who are dissatisfied with the supply of social events	25 pct	20 pct

Experiment 2

	Aarhus	Copenhagen
Proportion of Political Science students who are overall dissatisfied with their education	17 pct	21 pct

Or

	Aarhus	Copenhagen
Proportion of Political Science students who are overall dissatisfied with their education	21 pct	17 pct

Experiment 3

	Aarhus	Copenhagen
Proportion of Political Science students who do not think they receive enough teaching	32 pct	40 pct

Or

	Aarhus	Copenhagen
Proportion of Political Science students who do not think they receive enough teaching	40 pct	32 pct

On a scale from 0 to 10 where 0 means “of very low importance” and 10 means “of very high importance”, how important do you find this information for an overall evaluation of the Political Science educational quality at Aarhus University and University of Copenhagen?”

Figure 2: Effect of relative performance on importance evaluations

Note: † = $p < 0.1$; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$. Effects of relative performance on perceived importance with 95% confidence intervals. Significant differences identified by comparing groups in two-tailed t-tests. $N = 203$ for each experiment in 2015, 195 for each experiment in 2016, and 216-217 for each experiment in 2017. “Other” = University of Copenhagen (CPH) in 2015, University of Southern Denmark (SDU) in 2016, and CPH or SDU (randomized) in 2017.

Figure 3: Effect of relative performance on importance evaluations when pooling data from the three data collections

Note: † = $p < 0.1$; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$. Effects of relative performance on perceived importance with 95% confidence intervals. Significant differences identified by comparing groups in two-tailed t-tests. N = 615 in experiment 1; 614 in experiments 2-3. “Other” = University of Copenhagen (CPH) in 2015, University of Southern Denmark (SDU) in 2016, and CPH or SDU (randomized) in 2017.

APPENDIX

A1: Effect of relative performance on importance evaluations when pooling data from the three data collections and including students who did not choose Political Science at Aarhus University as first priority

Note: [†] = $p < 0.1$; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$. Effects of relative performance on perceived importance with 95% confidence intervals. Significant differences identified by comparing groups in two-tailed t-tests. N = 704 in experiment 1; 703 in experiments 2-3. "Other" = University of Copenhagen (CPH) in 2015, University of Southern Denmark (SDU) in 2016, and CPH or SDU (randomized) in 2017.

ONLINE SUPPORTING INFORMATION

This is supplementary material for the article entitled “Biased, Not Blind: An Experimental Test of Self-Serving Biases in Service Users' Evaluations of Performance Information”.

The supplementary material contains:

1. Balance tests (S1-S3)
2. Test for differences in effects depending on comparison university (S4)
3. Test for differences in effects across samples (S5-S7)
4. Power estimations (S8)
5. Wording of experiments using University of Southern Denmark as comparison (S9)

Furthermore, a replication dataset and a code to reproduce the article’s results is available at the Harvard Dataverse repository: <https://doi.org/10.7910/DVN/EH0D15>.

S1: Balance tests (2015 data)

	Experiment 1: Social events		Experiment 2: General dissatif.		Experiment 3: Enough teaching	
Best performing university	AU	CPH	AU	CPH	AU	CPH
Female (proportion)	0.45	0.53	0.50	0.47	0.47	0.50
Age (mean)	20.97	20.50*	20.77	20.73	20.73	20.79
Father has higher education (proportion)	0.59	0.63	0.58	0.63	0.60	0.60
Mother has higher education (proportion)	0.75	0.78	0.75	0.78	0.74	0.78
Left-right orientation (mean) ¹	4.71	4.30	4.40	4.67	4.43	4.61
Sympathy twd. government (mean) ²	3.70	3.55	3.36	3.94	3.79	3.44
Identifies with a party (proportion)	0.44	0.41	0.43	0.43	0.46	0.38
Left-wing voter (proportion) ³	0.70	0.78	0.75	0.72	0.75	0.72
N	109	94	110	93	111	92

Note: † = $p < 0.1$; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$. Significant differences identified by comparing groups in two-tailed t-tests and proportion tests. ¹: Self-identified left-right orientation on scale ranging from 0-10 with higher values corresponding to more right-wing. ²: Self-identification on scale ranging from 0 (indicating very low sympathy towards current government/administration) to 10 (indicating very high sympathy towards current government/administration). ³: Respondents have been coded as left-wing voters if they would vote for the Social Democrats, the Danish Social Liberal Party, the Socialist People's Party, The Alternative, or the Red-Green Alliance in case of a parliamentary election at the time of the data collection.

S2: Balance tests (2016 data)

	Experiment 1: Social events		Experiment 2: General dissatif.		Experiment 3: Enough teaching	
Best performing university	AU	SDU	AU	SDU	AU	SDU
Female (proportion)	0.53	0.59	0.55	0.57	0.54	0.58
Age (mean)	20.53	20.71	20.50	20.75	20.61	20.63
Father has higher education (proportion)	0.62	0.66	0.64	0.65	0.69	0.59
Mother has higher education (proportion)	0.79	0.79	0.78	0.80	0.79	0.79
Left-right orientation (mean) ¹	4.28	4.30	4.20	4.39	4.31	4.26
Sympathy twd. government (mean) ²	3.12	2.85	2.87	3.11	3.07	2.89
Left-wing voter (proportion) ³	0.83	0.84	0.87	0.81	0.79	0.88
N	97	98	102	93	98	97

Note: † = $p < 0.1$; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$. Significant differences identified by comparing groups in two-tailed t-tests and proportion tests. ¹: Self-identified left-right orientation on scale ranging from 0-10 with higher values corresponding to more right-wing. ²: Self-identification on scale ranging from 0 (indicating very low sympathy towards current government/administration) to 10 (indicating very high sympathy towards current government/administration). ³: Respondents have been coded as left-wing voters if they would vote for the Social Democrats, the Danish Social Liberal Party, the Socialist People's Party, The Alternative, or the Red-Green Alliance in case of a parliamentary election at the time of the data collection.

S3: Balance tests (2017 data)

	Experiment 1: Social events		Experiment 2: General dissatif.		Experiment 3: Enough teaching	
	AU	Other	AU	Other	AU	Other
Best performing university						
Female (proportion)	0.54	0.60	0.60	0.53	0.57	0.57
Age (mean)	20.65	20.70	20.70	20.67	20.53	20.85 [†]
Father has higher education (proportion)	0.61	0.72	0.68	0.65	0.68	0.64
Mother has higher education (proportion)	0.73	0.85*	0.85	0.73*	0.78	0.79
Left-right orientation (mean) ¹	4.82	4.50	4.35	4.97*	4.57	4.77
Sympathy twd. government (mean) ²	3.95	3.89	3.53	4.30*	3.93	3.91
Left-wing voter (proportion) ³	0.66	0.76	0.74	0.68	0.73	0.69
N	114	103	110	106	115	101

Note: [†] = $p < 0.1$; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$. Significant differences identified by comparing groups in two-tailed t-tests and proportion tests. ¹: Self-identified left-right orientation on scale ranging from 0-10 with higher values corresponding to more right-wing. ²: Self-identification on scale ranging from 0 (indicating very low sympathy towards current government/administration) to 10 (indicating very high sympathy towards current government/administration). ³: Respondents have been coded as left-wing voters if they would vote for the Social Democrats, the Danish Social Liberal Party, the Socialist People's Party, The Alternative, or the Red-Green Alliance in case of a parliamentary election at the time of the data collection. "Other" = University of Copenhagen or University of Southern Denmark (randomized).

S4: Test for differences in effects depending on comparison university

	Experiment 1	Experiment 2	Experiment 3
AU best	0.55 [†] (0.30)	0.20 (0.32)	0.26 (0.28)
Copenhagen as comparison	0.20 (0.31)	0.27 (0.32)	0.49 [†] (0.28)
AU best X Copenhagen as comparison	0.23 (0.42)	0.16 (0.44)	0.18 (0.39)
Constant	4.48*** (0.22)	5.21*** (0.22)	6.17*** (0.20)
N	615	614	614
Performance indicator	Social events	General dissatisfaction	Enough teaching

Note: [†] = $p < 0.1$; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$. Ordinary Least Square with standard errors in parentheses.

S5: Test for differences in effects across samples: 2015 sample vs. 2016 sample

	Experiment 1	Experiment 2	Experiment 3
AU best	0.86* (0.35)	0.34 (0.36)	0.54 [†] (0.31)
2016 sample	-0.39 (0.36)	-0.89 (0.37)	-0.91** (0.32)
AU best X 2016 sample	-0.27 (0.50)	-0.15 (0.51)	-0.19 (0.44)
Constant	4.09*** (0.26)	6.54*** (0.26)	7.22*** (0.23)
N	398	398	398
Performance indicator	Social events	General dissatisfaction	Enough teaching

Note: [†] = $p < 0.1$; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$. Ordinary Least Square with standard errors in parentheses.

S6: Test for differences in effects across samples: 2015 sample vs. 2017 sample

	Experiment 1	Experiment 2	Experiment 3
AU best	0.86* (0.35)	0.34 (0.34)	0.54 (0.33)
2017 sample	1.82*** (0.35)	-2.50*** (0.34)	-1.43*** (0.34)
AU best X 2017 sample	-0.35 (0.48)	-0.17 (0.47)	-0.37 (0.46)
Constant	4.09*** (0.25)	6.54*** (0.25)	7.22*** (0.25)
N	420	419	419
Performance indicator	Social events	General dissatisfaction	Enough teaching

Note: [†] = $p < 0.1$; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$. Ordinary Least Square with standard errors in parentheses.

S7: Test for differences in effects across samples: 2016 sample vs. 2017 sample

	Experiment 1	Experiment 2	Experiment 3
AU best	0.59 [†] (0.35)	0.20 (0.37)	0.35 (0.35)
2017 sample	2.21*** (0.34)	-1.61*** (0.36)	-0.52 (0.35)
AU best X 2017 sample	-0.09 (0.48)	-0.03 (0.50)	-0.19 (0.49)
Constant	3.69*** (0.25)	5.65*** (0.26)	6.31*** (0.25)
N	412	411	411
Performance indicator	Social events	General dissatisfaction	Enough teaching

Note: [†] = $p < 0.1$; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$. Ordinary Least Square with standard errors in parentheses.

S8: Power Analysis

Power has been estimated in Stata based on observed effects and standard deviations in the article's three experiments. Ideally, power analyses would have been based on external studies but this was not possible, as the article's experimental designs have not been used in prior literature. Instead, power analyses have been run for each of the three experiments using average effects and standard deviations across the three data collections.

Experiment 1 (social activities)

- Average importance, AU best = 5.21
- Average importance, other best = 4.56
- Average standard deviation on DV: 2.47
- Power, 2015 (n=203) = 0.962
- Power, 2016 (n=195) = 0.955
- Power, 2017 (n=216) = 0.971
- Power, pooled (n=615) = 1
- N required for power of 0.8 = 116

Experiment 2 (general dissatisfaction)

- Average importance, AU best = 5.64
- Average importance, other best = 5.41
- Average standard deviation on DV: 2.49
- Power, 2015 (n=203) = 0.258
- Power, 2016 (n=195) = 0.250
- Power, 2017 (n=215) = 0.271
- Power, pooled (n=614) = 0.628
- N required for power of 0.8 = 922

Experiment 3 (enough teaching)

- Average importance, AU best = 6.79
- Average importance, other best = 6.44
- Average standard deviation on DV: 2.32
- Power, 2015 (n=203) = 0.571
- Power, 2015 (n=195) = 0.554
- Power, 2015 (n=215) = 0.596
- Power, pooled (n=614) = 0.962
- N required for power of 0.8 = 347

S9: Question wording in experiments (comparison university = Southern Denmark)

“Please take a look at table below:

Experiment 1

	Aarhus	Southern Denmark
Proportion of Political Science students who are dissatisfied with the supply of social events	20 pct	25 pct

Or

	Aarhus	Southern Denmark
Proportion of Political Science students who are dissatisfied with the supply of social events	25 pct	20 pct

Experiment 2

	Aarhus	Southern Denmark
Proportion of Political Science students who are overall dissatisfied with their education	17 pct	21 pct

Or

	Aarhus	Southern Denmark
Proportion of Political Science students who are overall dissatisfied with their education	21 pct	17 pct

Experiment 3

	Aarhus	Southern Denmark
Proportion of Political Science students who do not think they receive enough teaching	32 pct	40 pct

Or

	Aarhus	Southern Denmark
Proportion of Political Science students who do not think they receive enough teaching	40 pct	32 pct

On a scale from 0 to 10 where 0 means “of very low importance” and 10 means “of very high importance”, how important do you find this information for an overall evaluation of the Political Science educational quality at Aarhus University and University of Southern Denmark?”