

What Goes in Must Come Out? Frontline Workers Factor in Production Attributes when Assessing Organizational Performance

Paw Hansen^{a*}

*^aVIVE—The Danish Center for Social Science Research, Aarhus and Copenhagen,
Denmark*

**Corresponding author: Paw Hansen: pahh@vive.dk*

Provide short biographical notes on all contributors here if the journal requires them.

What Goes in Must Come Out? Frontline Workers Factor in Production Attributes when Assessing Organizational Performance

Lawmakers define formal policy goals and the methods by which goal attainment is measured. However, frontline workers may hold their own perspectives on what constitutes the formal goals of their organization. Although these implicit considerations can significantly influence the extent to which agreed-upon policies are put into practice, little is known about the attributes frontline workers factor in when assessing performance. We contribute to the existing literature by developing hypotheses suggesting that when assessing the attainment of organizational goals, frontline workers take into account not only the formal goals but also two attributes of the service production: (1) the peer efforts exerted to achieve these goals and (2) the clientele needs. To test our hypotheses, we conducted a conjoint survey experiment involving 1,200 public high-school teachers from 131 organizations. Our research provides evidence in support of our hypotheses, shedding light on the goal attainment considerations made at the street-level of public service delivery.

Keywords: performance; frontline work; deservingness; conjoint designs;

Introduction

Although lawmakers have the discretion to define the formal goals of policies, programs, and public service, the frontline workers putting them into practice often have their own ideas about those goals, including what they should be and when they are met (Lipsky, 2010). However, the effective execution of public policies depends on at least some degree of alignment between higher-level decision-makers and those at the street-level (Destler, 2017; Hood, 2012; May & Winter, 2009).

While we have numerous studies on how citizens and politicians make sense of performance (Baekgaard & Serritzlew, 2016; Christensen et al., 2018; James & Moseley, 2014; Moynihan et al., 2017; Olsen, 2017), much less is known about frontline workers (N. B. G. Petersen, 2020; N. B. G. Petersen et al., 2019). This remains problematic because frontline workers play a special role with regard to public service performance: they *produce* it.

Building on this premise, this study seeks to explore how frontline workers evaluate and consider the attainment of organizational goals. More specifically, it posits that frontline workers do not solely base their assessments on formal goals set by lawmakers; they also factor in two key attributes associated with service production: the peer efforts exerted in achieving these goals and the perceived external pressures facing these formal objectives. We ask the following research question: *How do public service production attributes affect frontline workers' performance evaluations?*

To test our hypotheses, we conducted a conjoint survey experiment involving 1,200 public high-school teachers from 131 high schools. In brief, we asked teachers to assess the formal goal attainment on a number of fictive high schools. For each high school we randomly assigned information on both outcomes explicitly defined as formal goals by lawmakers and production attributes concerning peer effort and external

pressures (specifically, clientele needs). Our findings offer empirical evidence in support of our hypotheses.

Our study reveals three significant implications for research and practice. Firstly, it highlights a critical aspect of democratic accountability: while lawmakers articulate goals as promises to the electorate (Christensen et al., 2018), discrepancies in understanding between frontline workers and policymakers can obscure the implementation of these intentions (Meyers & Nielsen, 2012). Secondly, the potential lack of alignment between formal goals and perceived criteria by frontline workers may affect the effectiveness of service delivery, posing challenges to organizational learning (Behn, 2003). Lastly, our findings emphasize focal points for public managers, pointing to peer efforts and clientele needs as pivotal areas for communication regarding goal attainment. Recognizing and incorporating production attributes into daily interactions with workers, often overlooked in formal goal-setting, becomes imperative for effective public management.

The remainder of this article comprises five sections. In the following section, we will present a novel theoretical account on why and how frontline workers incorporate production attributes when assessing organizational performance. Afterward, we will present our test case, before moving on to discussing our experimental research design. In the subsequent section, we will present our findings before. The article ends with a discussing of limitations and implications of our conclusions.

Theory

How do frontline workers factor in production traits when assessing performance? We begin our theoretical account by briefly sketching a stylized production of public

service (depicted in Figure 1). At its most basic form, the production of public service can be seen as consisting of three sequential steps. This line of thinking is sometimes termed the IOO or Input-Output-Outcome model (Andersen et al., 2016). As a running example, consider the IOO chain of a ‘people-changing’ organization such as a public high school. The *outcome(s)* are the changes made in students, such as learning or wellbeing. Preceding outcome(s) is the production process: the inputs and what is done to these inputs to produce the desired outcome(s). For example, outputs may be the various efforts made by teachers such as teaching in certain ways or implementing other initiatives to better teaching quality. Finally, the inputs are the available resources. Resources may be concrete, such as particular books or access to IT facilities, or they may concern the students meant to ‘change’ as a result of the production process. For instance, in the case of achieving a fixed test score on reading abilities (as an outcome), a school with a high share of students with dyslexia or other disadvantages can be said to have fewer resources, all other things being equal.

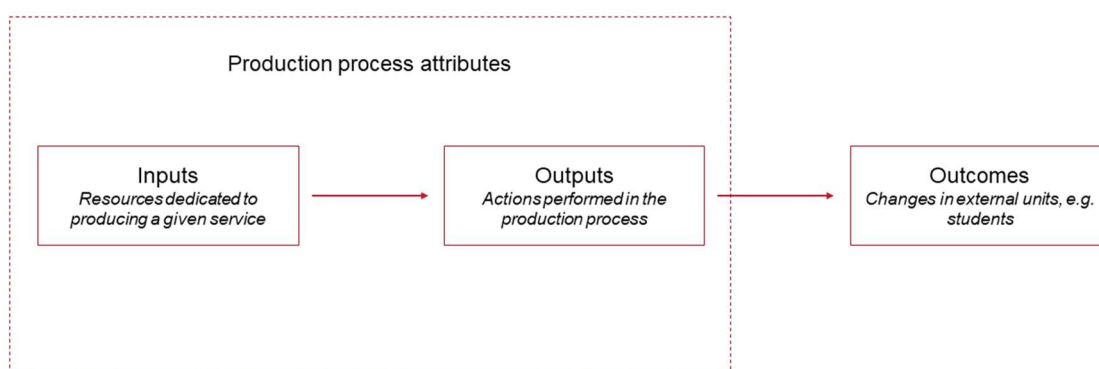


Figure 1. Sketch of a public service production

Although frontline workers produce the public service in question, *lawmakers* hold the discretion to define the formal goals of that service. Most often, goals will

center on outcomes (Andersen et al., 2016; Moynihan et al., 2011). One reason for this is that policy goals serve to explicate policy intentions, laying out the direction for the policy and specifying its aims (May & Winter, 2009). By extension, policy goals communicate to voters what lawmakers aim to achieve through their policies, hereby establishing a measure against which promises and objectives can be evaluated. And goals can then serve to holding elected officials accountable to the electorate (Christensen et al., 2018).

Against that backdrop, performance assessments have received wide attention in public administration research. Broadly, research has tended to focus on citizens and decision-makers with four pervasive themes having dominated the literature (James & Olsen, 2017): (1) the reception of performance indicators and how this is affected by motivated reasoning (e.g. Baekgaard & Serritzlew, 2016; Hvidman & Andersen, 2016), (2) how numerical factors influence citizens performance perceptions (e.g. Olsen, 2013), (3) the role of reference points (e.g. James & Moseley, 2014; Olsen, 2017) and (4) the valence of performance information, such as being relatively positive or negative (e.g. James, 2011).

To frontline workers, formal goals signal in which direction their organization should move (Meyers & Nielsen, 2012). These signals are important because without expressing goals in salient ways, “street-level bureaucracies will perform with internally generated objectives.” (Lipsky, 2010, 46). Alongside citizens, managers, and politicians, frontline employees actively engage in an ongoing struggle to determine which goals should take precedence within their organization (N. B. G. Petersen et al., 2019). Without at least some degree of goal alignment, there can be detrimental consequences for both performance and democratic efficiency since the successful

execution of policies heavily relies on the support and commitment of those responsible for delivering public services at the street-level (Destler, 2017; Hood, 2012).

Frontline workers differ from citizens and decision-makers because they themselves are *producers* of the public service receiving evaluation. One implication of this is that frontline workers may want to protect or enhance their self-image depending on whether performance measures show that they are doing well or not. For example, Petersen et al., (2019) showed that teachers are more likely to consider performance measures as valid, legitimate, and useful when performance measures showcase relatively high performance than relatively low performance. In addition, teachers in high-performing schools were more likely to take causal responsibility for their schools' performance.

By extension, we propose that when assessing organizational outcome performance, frontline workers will be prone to factoring in considerations related to the production process. In the subsequent section, we explain how this can be understood as a type of 'halo', and we present our two main hypotheses.

Which production attributes do frontline workers factor in? Peer efforts, clientele needs, and the role of the halo effect

A consistent discovery within the extensive literature on performance assessments is the tendency to evaluate everything about someone positively or negatively (Kahneman, 2011). This means that our assessment of one attribute often spills over into our judgment of all other attributes, magnifying the consistency of judgments to maintain straightforward and coherent explanatory narratives. This phenomenon is commonly known as the 'halo effect' (Belle et al., 2017). It is widely established that these halo effects systematically bias performance ratings, aligning ratings across different dimensions, sometimes disregarding or conflicting with available information. For

instance, raters might allow a rating on one criterion (e.g., punctuality) to influence ratings on subsequent criteria (e.g., task completion, customer communication) (Belle et al., 2017). Furthermore, these mechanisms not only influence our perceptions of individuals but also shape our perspectives on organizations

At least two major types of mechanisms drive the halo effect. From a cognition perspective, the halo effect allows individuals to simplify complex information and make quick judgements, serving as a cognitive shortcut. In addition, for many types of judgements, we have only incomplete information; employing a halo helps us extend existing information to fill information gaps. From an emotion perspective, the halo works because when we feel positively about something due to a specific trait, that emotion can spill over and influence our judgment about other qualities (Sinclair, 1988; see also Slovic et al., 2007).

In brief, we expect two production attributes to spill over, inflating or deflating frontline workers' performance assessments:

- *Peer efforts*
- External pressures, specifically, *clientele needs*

First, we expect that frontline workers will factor in *peer efforts*¹. Generally, taking action or observing someone else doing so is often seen as inherently better than inaction. (Patt & Zeckhauser, 2000). For example, experimental research has shown that people prefer antibiotics even when told it will have no effect and could possibly do harm (Thorpe et al., 2020). This 'action bias' is also sometimes used to explain why the goalkeeper in soccer too often dives to either the left or the right, when standing still would sometimes be preferable (Bar-Eli et al., 2007).

The same taste for ‘doing something’ is echoed in the deservingness literature where taking action may be a sign that someone is willing to put in the effort (M. B. Petersen, 2012). Noticing that someone is hardworking or effortful prompts us to think of them as deserving of our help (Jilke & Tummers, 2018)

Our basic argument here is that the same mechanism will cause frontline workers to think of a high school (consisting of peers) in positive ways, and that this will then spill over into their performance assessment. Thus, the ‘peer effort halo’ may be summed up as: “You work hard, you must be doing well.” In sum, we hypothesize the following:

H1: Frontline workers will assess organizational performance higher when peer effort is higher (as opposed to lower)

External pressures offer an alternative source of inflating frontline workers’ performance assessments. One particularly important source of external pressures are the needs of the clientele (May & Winter, 2009; Meyers & Nielsen, 2012). Organizations dealing with ‘challenging clients’—where achieving outcome goals is notably difficult—are considered to face greater external pressures compared to those for whom achieving these goals is easier.

Again, we may look the deservingness literature which has shown that even subtle cues of need will trigger a sense of compassion (M. B. Petersen et al., 2012) (M. B. Petersen, 2015; M. B. Petersen et al., 2012). By extension, we expect this sense of compassion to spill over, affecting the overall performance assessment.

In the education context, schools with a significant proportion of socio-economically disadvantaged children face increased external pressures, as meeting policy-defined student outcomes becomes relatively more challenging. Specifically, minority students often face low academic performance expectations, particularly from White teachers (Gershenson et al., 2016). Additionally, racial minorities are often

structurally disadvantaged in the allocation of opportunities and resources (Jilke & Tummers, 2018).

To summarize, concerning clientele needs, we hypothesize that:

H2: Frontline workers will assess organizational performance higher when clientele needs are greater (as opposed to lower)

Test case: Organizational Performance in Danish high-schools

To gauge the extent to which frontline workers factor in production attributes in their overall performance assessment, an ideal test case should meet three criteria. First, it should comprise a policy, program, or public service with clearly defined formal goals set by lawmakers. These goals should be widely communicated to stakeholders (specifically, frontline workers). Second, the test case should involve a considerable number of frontline workers who possess the autonomy to both adhere to and diverge from these formal goals. This discretionary power is crucial in understanding how the formal goals are interpreted and applied at the operational level, offering insights into the decision-making processes of frontline workers. Lastly, frontline workers' performance assessments should have some real-world consequences for the citizens in question.

In essence, a suitable test case should encapsulate a public performance management initiative, involve a substantial number of frontline workers with decision-making autonomy, and have discernible consequences for the citizens involved. In that way, the test case will provide valuable insights into the translation of policy goals into actionable outcomes at the operational level.

Based on these criteria, we chose to examine the Danish high-school reform, initiated in 2017/2018 ("Gymnasireformen"). Central to this reform were three specific

goals outlined in table 1. This case serves as an ideal study subject as it allows us to assess how much emphasis teachers place on each of these goals. Furthermore, we can also evaluate the extent to which teachers permit production criteria to influence their overall assessments of goal attainment (Ministry of Children and Education, 2019).

As it is key to be clear about the conceptual space of performance (Andersen et al., 2016), we note a few things here. Performance as understood in the context of our test case is first of all defined top-down by lawmakers. Although teachers, students, parents, and others may be considered stakeholders in the sense that they can affect performance, they have no pre-defined say in how performance is defined in the reform. Second, performance is defined by three formal goals (or ‘guiding objectives’ in the terms of the reform) and all three are to be measured using quantitative metrics. Concerning the focus of performance, all goals relate to an outcome rather than inputs (such as due process or student participation).

Table 1. Guiding objectives for the upper secondary education programs

	Guiding objective	Performance indicator
Objective 1: Socioeconomic mobility	The high schools must challenge all students regardless of their social background, so they become as skilled as they can.	Annual assessment of the development in students' grades concerning the following parameters: <ul style="list-style-type: none"> • Primary school grades • Parents' highest completed education • Origin • Gender
Objective 2: Transition to higher education	A larger proportion of students should start a higher education program.	Annual registration of transition rates to education on a national level.

Objective 3: Student wellbeing	The well-being in upper secondary education programs must be strengthened.	Annual well-being assessments.
--------------------------------	--	--------------------------------

Methods and Data

Our experimental protocol comprised five steps. First, we began the study by sending out email invitations to all Danish high-school teachers with publicly available contact information. Among them, 1,250 teachers completed at least some sections of the survey, yielding an approximate XX percent response rate. Table 1 presents descriptive information regarding the sampled participants.

Second, as part of the survey, we incorporated a novel conjoint experiment. We presented teachers with a series of tables displaying profiles of fictional Danish high schools. For each set of profiles, we posed questions related to organizational performance. We aimed to gauge the teachers' perceptions of performance through two specific questions. Firstly, we asked respondents to make a choice between the two fictional high schools by asking, "Which of the two high schools do you believe best fulfils the **formal goals** of the gymnasium?" (emphasis in the original¹). Additionally, to enhance the accuracy of these measurements and guard against potential errors (such as choosing Profile A when one actually preferred Profile B, see Clayton et al., 2023), we also asked respondents to rate the performance of each presented high school profile. Specifically, we asked, "How would you assess the fulfilment of the **formal goals** for

¹ We emphasized 'formal goals' in the query using bold font to make explicit that we were interested in the formal goals and not, say, the individual teachers' idea of what *should* be the goals of the gymnasium.

each of the gymnasiums?" (emphasis in the original) For the rating-based outcome, we used a ten-point scale ranging from "1: Does not meet the formal goals of the gymnasium at all" to "10: Completely fulfils the formal goals of the gymnasium." An example table from the experiment is shown in Figure 2.

	High school A	High school B
Socioeconomic mobility	Exam grades are far above what can be expected from student backgrounds	Exam grades are as can be expected from student backgrounds
Transition to higher education	Far fewer enrolled in higher education compared to the national average.	Slightly more enrolled in higher education compared to the national average.
Student wellbeing	Same as the national average	Far better than the national average
Teaching development	Have implemented many initiatives to improve the quality in the past year.	Have not implemented initiatives to improve the quality in the past year.
Working conditions	Has a high proportion of students with a non-Western background.	Has a low proportion of students with a non-Western background.

Figure 2 An Example Conjoint Table from the Experiment. The full set of possible attribute values are provided in Table A1.

Third, to assess the individual impact of specific pieces of information regarding the performance of high school profiles, we randomly assigned five different pieces of information. Among these, three directly addressed the formal goals of the reform—socioeconomic mobility, transition to higher education, and student wellbeing—while the remaining two focused on peer efforts and clientele needs. For the three aspects linked to the reform's formal goals, we selected values aligning with the reform's

objectives. Hence, for socioeconomic mobility, each level was put in relation to what could be expected from student backgrounds. For example, an attribute level could read “Exam grades are far above what can be expected from student backgrounds”. For the attributes concerning transition to higher education and student wellbeing, attribute values were instead put in relation to ‘the national average’.

Fourth, to gauge the particular impact of production attributes on performance evaluations, we assigned information regarding peer efforts and clientele needs. Regarding peer effort, we employed two distinct sets of wording—one focusing on past efforts (e.g., 'Have implemented many initiatives to improve quality in the past year.')

and the other on future efforts (e.g., ‘Significant focus on developing the teaching in the future.’). Concerning clientele need, we similarly utilized two sets of wording—one tied to student origin (e.g., 'Has a high proportion of students with a non-Western background.')

and the other linked to parents’ educational background (e.g., 'Has a high proportion of parents without higher education.'). For both types of production attributes, we included two subgroups of wordings to reduce the likelihood of one comparison unduly influencing the overall performance evaluation. In particular, our concern was that repeatedly asking respondents to assess multiple pairs related to student origin might lead them to believe the study primarily focused on that aspect, potentially introducing bias into our estimates. Importantly, despite variations in wording, all align with the core theoretical constructs under scrutiny. Table A1 provide a full list of attributes and their corresponding values.

Finally, regarding research ethics, we assured all respondents that their responses would remain anonymous, without any personal identification, and would not be shared, for instance with high school managers. Furthermore, we included contact

information at the survey's conclusion, inviting respondents to reach out for any concerns, questions, or clarifications they might have.

Table 2. Sample Descriptives

		Mean	Std. Dev.
Age		46.0	10.6
Work Experience		15.5	9.5
Tenure		12.1	8.8
		N	Pct.
Origin	Danish	1132	90.6
	Non-Danish	118	9.4

n = 1,250 respondents with a least one entry on the outcome variable.

Findings

Figures 1 and 2 illustrate our main findings. In both figures, we use dots to show the marginal means of performance ratings for each attribute level with horizontal lines displaying the corresponding 95 percent confidence intervals. These confidence intervals are computed in reference to the null hypothesis positing that the performance rating equals the overall mean rating, which, in our context, is 6.45. That is, we estimate whether the presentation of specific performance information leads respondents to deviate from the overall mean. As an example, consider the initial line in Figure 1, which illustrates that when exam grades are ‘far beyond what can be expected from student backgrounds’, teachers give an average performance rating of approximately 6.85 (with a 95 percent confidence interval ranging from 6.75 to 6.95). This rating significantly differs from the average performance rating (6.45), as evidenced by the confidence band not intersecting with the vertical dashed line.

Teachers Align with Lawmakers on Performance, Agreeing with All Three Formal Goals of the Reform

Figure 1 presents the results for the first three dimensions of our conjoint analysis, all of which concerned the formal goals of the reform. We find, that all three dimensions appear to influence teachers' performance ratings. In terms of *socioeconomic mobility*, schools with exam grades far above what can be expected from student backgrounds receive higher performance ratings, while those that underperform receive lower ratings.

Further, teachers also take into account *student wellbeing* when evaluating the organizational performance of schools. For instance, when student wellbeing is far better than the national average, schools receive ratings of 6.95 or about half a point higher than the sample average. Conversely, when wellbeing falls well below the national average, schools receive performance ratings of 5.73 or 0.72 points below the sample average.

Finally, Figure 1 affirms that teachers also factor in students' *transition to higher education* in their assessment of organizational performance. When far more students are enrolled in higher education compared to the national average, schools receive performance ratings that are, on average, 0.48 points higher than the sample average. Similarly, but in the opposite direction, when a significantly lower number of students are enrolled, schools receive average performance ratings approximately 0.58 points below the sample average.

Interestingly, consistent with prior research on negativity bias (e.g. Nielsen & Moynihan, 2017; Olsen, 2017), schools appear to face more severe penalties when any of the three outcome goals fall below the reference point compared to the gains in performance received when exceeding expectations (*socioeconomic mobility*) or national average (*wellbeing* and *transition to higher education*).

In summary, there is substantial evidence supporting that teachers align with lawmakers on all three formal reform goals.

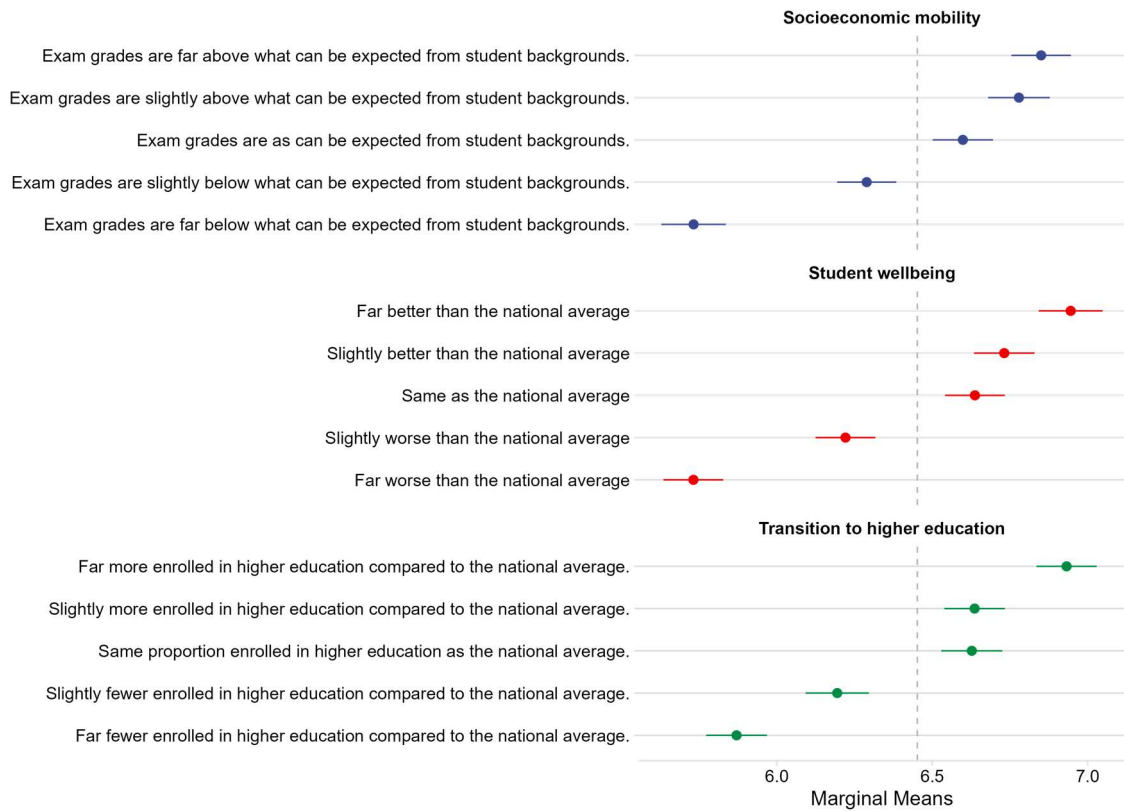


Figure 1. Marginal mean performance rating for the three performance dimensions concerning high school outcomes. Dashed line marks the sample average rating.

Production Attributes Cause Teachers to Inflate/Deflate Performance Ratings

Figure 2 shows the results for the remaining dimensions, all concerning service production. We draw three major conclusions. First, teachers put emphasis on peer effort. Concerning *past* effort, teachers tend to reward schools that have implemented many initiatives to improve the quality in the past years (marginal mean 6.87) but punish schools who have not implemented any initiatives (marginal mean 6.05). Similarly, teachers value *future* efforts: when schools have a significant focus on developing teaching in the future, they are rewarded by about 0.45 points. In contrast,

having no focus on developing teaching in the future, results in an average penalty of 0.59 points.

Second, teachers also emphasize clientele needs when evaluating performance. For student origin, schools are rewarded about 0.16 points when they have a high proportion of students with a non-Western background. Moreover, schools receive a penalty in their performance rating when having a low proportion of students with non-Western backgrounds (marginal mean is 0.16 points lower than the sample average).

Third, the impact of students' socioeconomic backgrounds is less clear. Although a high proportion of parents without higher education causes higher performance ratings (0.13 above the sample mean), a low proportion does not cause any penalty—at least not to a degree that is statistically significant. However, the overall pattern does follow the expected dose-response relationship.

To summarize, we believe our findings provide clear evidence in favour of the hypotheses stating that peer effort will lead to higher performance ratings. Concerning clientele need, results are more mixed. Our findings show robustness across both forms of production attributes, demonstrating a clear dose-response relationship with consecutively higher levels of deservingness corresponding to elevated performance ratings.

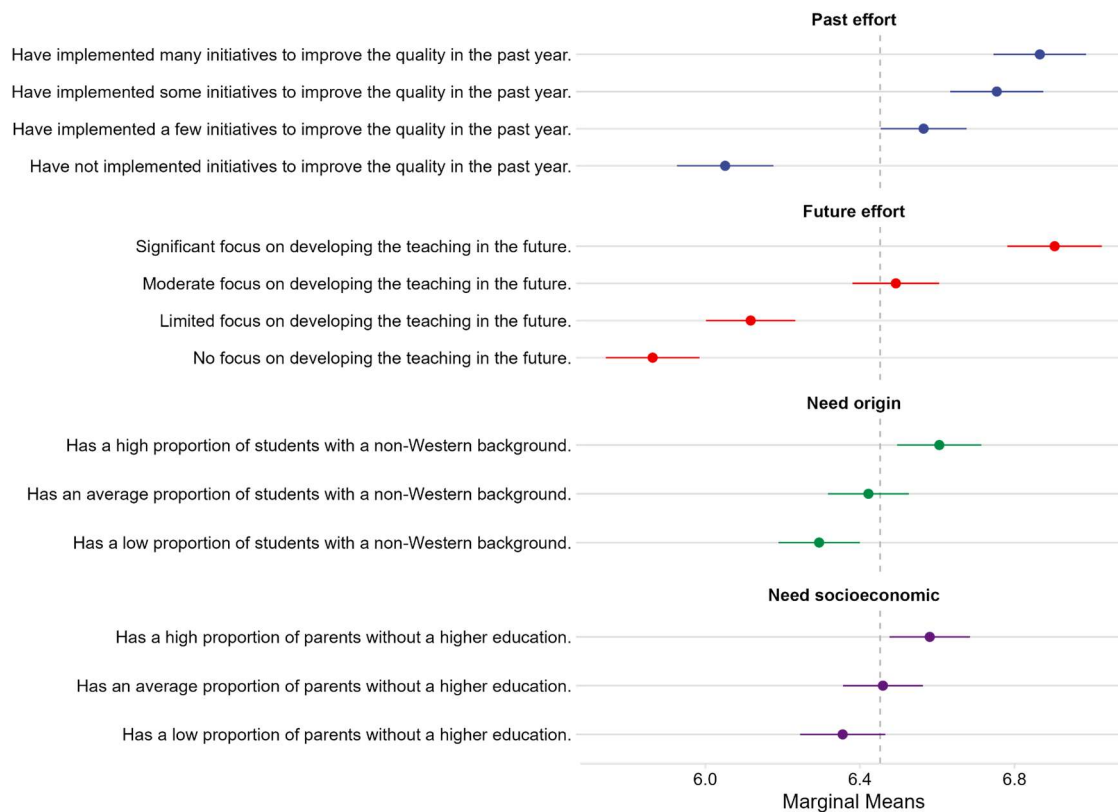


Figure 2. Marginal mean performance rating for the performance dimensions concerning organizational deservingness. Dashed line marks the sample average rating.

Discussion and Conclusion

More than forty years ago, Lipsky (2010) noted that lawmakers and frontline workers do not always see eye to eye on what should be factored in when assessing performance. Building on this premise, we have argued that because the performance assessments centre on the frontline workers themselves, workers will consider production attributes in their performance evaluation, irrespective of whether these attributes are included in lawmakers' formal goals. Using a conjoint experiment embedded in a large-scale survey among Danish high school teachers, we find evidence that frontline workers do factor in production attributes considerations when assessing organizational performance.

First, we find that organizations demonstrating peer effort tend to receive elevated performance ratings. When teachers consider information regarding formal goal attainment, they allow details about peer effort to influence their overall perceptions of high schools' performance. For instance, schools implementing 'numerous initiatives' to enhance performance received higher ratings.

Second, we find that teachers also factor in information about high schools' external pressures, systematically attributing higher performance ratings to schools with challenging clienteles such as students from minority backgrounds or parents with limited higher education.

Using a conjoint experiment design allowed us to empirically disentangle the individual effect of several pieces of information about organizational performance. Further, leveraging a real-world sample of frontline workers bolsters the credibility of these findings. However, our study should be interpreted in light of certain limitations. One limitation is our use of a restricted set of information. While we incorporated data on the attainment of the three reform-identified essential goals and two hypothesized service production attributes, our survey did not include other potentially important pieces of information. For instance, we refrained from asking respondents about other implicit goals that frontline workers might deem significant, such as legality or transparency (Lipsky, 2010). Omitting these aspects does not bias the effects of the production attributes estimated in our study but it does leave in the dark what other types of considerations frontline workers may rely when assessing performance. Future studies should try to untangle these issues further. Still, given the limited existing evidence on the influence of service production attributes in performance assessments, we believe our study stands as a timely and valuable contribution pushing the field forward.

Second, to what extent do our conclusions apply to other settings—frontline workers, organizational contexts, countries, and so on? Naturally, our study is limited in that we only test our theoretical propositions on a particular sample. Echoing the advice of Bryan et al., (2021), we call for scholars to map out the heterogeneity inherent in the effects we have proposed here. One starting point for studying heterogeneity is the type and degree of goal ambiguity. As noted by Lipsky (2010), frontline workers will fill in the blanks when lawmakers refrain from xxx. When more is left, more is to be filled in. Our case of the Danish high school reform comprised three clear outcome goals, which were widely communicated and concretized using specific metrics for each. But this may not always be the case. We can make the prediction that for cases with more goal ambiguity, production attributes will have more leverage in frontline workers' overall performance assessments.

While acknowledging its limitations, our study yields three significant implications for public managers and future research. Firstly, it underscores a critical aspect of democratic accountability. Lawmakers set goals that serve as promises to the electorate, articulating desired achievements. However, if frontline workers hold implicit notions of goal attainment, the alignment between lawmakers' intentions and actual implementation at the ground level becomes ambiguous.

Secondly, the effectiveness of frontline work delivery relies on some level of coherence between formal goals set by policymakers and those perceived by frontline workers. Yet, the integration of implicit performance criteria by frontline workers poses a challenge to organizational learning and improvement (Behn, 2003).

Finally, our findings highlight focal points for public managers. While acknowledging that the divergence in perspectives between policymakers and frontline workers may represent a Gordian knot, our study identifies two pivotal points—peer

efforts and clientele needs—that managers must address when communicating about goal attainment. Although lawmakers' might tend to overlook production attributes in formal goal-setting (Andersen et al., 2016; Moynihan et al., 2011), recognizing and integrating these attributes into daily interactions with workers becomes imperative for public managers.

In summary, our study sheds light on the considerations frontline workers prioritize when evaluating the fulfillment of formal goals. Notably, our findings reveal a 'halo effect' among the sampled teachers, where service production attributes significantly influenced their perception of achieving formal outcome goals. Despite these production attributes not being explicitly part of performance metrics, teachers factor in peer efforts and clientele needs. We believe further research in this realm holds substantial promise for scholars and practitioners seeking deeper insights into this crucial intersection.

References

- Andersen, L. B., Boesen, A., & Pedersen, L. H. (2016). Performance in Public Organizations: Clarifying the Conceptual Space. *Public Administration Review*, 76(6), 852–862. <https://doi.org/10.1111/puar.12578>
- Baekgaard, M., & Serritzlew, S. (2016). Interpreting Performance Information: Motivated Reasoning or Unbiased Comprehension. *Public Administration Review*, 76(1), 73–82. <https://doi.org/https://doi.org/10.1111/puar.12406>
- Bar-Eli, M., Azar, O. H., Ritov, I., Keidar-Levin, Y., & Schein, G. (2007). Action bias among elite soccer goalkeepers: The case of penalty kicks. *Journal of Economic Psychology*, 28(5), 606–621.
- Behn, R. D. (2003). Why Measure Performance? Different Purposes Require Different Measures. *Public Administration Review*, 63(5), 586–606. <https://doi.org/10.1111/1540-6210.00322>
- Belle, N., Cantarelli, P., & Belardinelli, P. (2017). Cognitive Biases in Performance Appraisal: Experimental Evidence on Anchoring and Halo Effects With Public Sector Managers and Employees. *Review of Public Personnel Administration*, 37(3), 275–294. <https://doi.org/10.1177/0734371X17704891>
- Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour*, 5(8), 980–989.
- Christensen, J., Dahlmann, C. M., Mathiasen, A. H., 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/10.1093/jopart/muy001>
- Clayton, K., Horiuchi, Y., Kaufman, A. R., King, G., Mayya, ¶, || K., Arias, S., Burden, B., Egami, N., Heinrich, T., Hopkins, D., Imai, K., Kalla, J., Kubinec, B., Rhee, K., Teele, D., Tingley, D., & Yang, J. (2023). *Correcting Measurement Error Bias in Conjoint Survey Experiments* *.
- Destler, K. N. (2017). A matter of trust: Street level bureaucrats, organizational climate and performance management reform. *Journal of Public Administration Research and Theory*, 27(3), 517–534. <https://doi.org/10.1093/jopart/muw055>
- Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? The effect of student–teacher demographic match on teacher expectations. *Economics of Education Review*, 52, 209–224.
- Hood, C. (2012). Public Management by Numbers as a Performance-Enhancing Drug: Two Hypotheses. *Public Administration Review*, 72(s1), S85–S92. <https://doi.org/https://doi.org/10.1111/j.1540-6210.2012.02634.x>
- Hvidman, U., & Andersen, S. C. (2016). Perceptions of public and private performance: Evidence from a survey experiment. *Public Administration Review*, 76(1), 111–120.
- James, O. (2011). Managing Citizens’ Expectations of Public Service Performance: Evidence from Observation and Experimentation in Local Government. *Public Administration*, 89(4), 1419–1435.
- James, O., & Moseley, A. (2014). Does Performance Information About Public Services Affect Citizens’ Perceptions, Satisfaction, and Voice Behavior? Field Experiments

- with Absolute and Relative Performance Information. *Public Administration*, 92(2), 493–511. <https://doi.org/10.1111/padm.12066>
- James, O., & Olsen, A. L. (2017). Citizens and Public Performance Measures: Making Sense of Performance Information. In G. G. Van Ryzin, O. James, & S. R. Jilke (Eds.), *Experiments in Public Management Research: Challenges and Contributions* (pp. 270–290). Cambridge University Press. [https://doi.org/DOI: 10.1017/9781316676912.013](https://doi.org/10.1017/9781316676912.013)
- Jilke, S., & Tummers, L. (2018). Which clients are deserving of help? A theoretical model and experimental test. *Journal of Public Administration Research and Theory*, 28(2), 226–238. <https://doi.org/10.1093/jopart/muy002>
- Kahneman, D. (2011). *Thinking, fast and slow*. macmillan.
- Lipsky, M. (2010). *Street-level bureaucracy: Dilemmas of the individual in public service*. Russell Sage Foundation.
- May, P. J., & Winter, S. C. (2009). Politicians, managers, and street-level bureaucrats: Influences on policy implementation. *Journal of Public Administration Research and Theory*, 19(3), 453–476. <https://doi.org/10.1093/jopart/mum030>
- Meyers, M. K., & Nielsen, V. L. (2012). Street-level bureaucrats and the implementation of public policy. *The SAGE Handbook of Public Administration*, 305–318.
- Ministry of Children and Education. (2019). *Status report on the implementation of the upper secondary school reform [Statusredegørelse for implementering af gymnasiereformen]*. <https://www.uvm.dk/-/media/filer/uvm/publikationer/2021/211018-statusredegørelse-for-implementering-af-gymnasiereformen-for-skoleaaret-2019-2020.pdf>
- Moynihan, D. P., Fernandez, S., Kim, S., Leroux, K. M., Piotrowski, S. J., Wright, B. E., & Yang, K. (2011). Performance regimes amidst governance complexity. *Journal of Public Administration Research and Theory*, 21(SUPPL. 1), 141–155. <https://doi.org/10.1093/jopart/muq059>
- Moynihan, D. P., Nielsen, P. A., & Kroll, A. (2017). Managerial Use of Performance Data by Bureaucrats and Politicians. In G. G. Van Ryzin, O. James, & S. R. Jilke (Eds.), *Experiments in Public Management Research: Challenges and Contributions* (pp. 244–269). Cambridge University Press. [https://doi.org/DOI: 10.1017/9781316676912.012](https://doi.org/10.1017/9781316676912.012)
- Nielsen, P. A., & Moynihan, D. P. (2017). How do politicians attribute bureaucratic responsibility for performance? Negativity bias and interest group advocacy. *Journal of Public Administration Research and Theory*, 27(2), 269–283. <https://doi.org/10.1093/jopart/muw060>
- 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. (2017). Compared to what? How social and historical reference points affect citizens' performance evaluations. *Journal of Public Administration Research and Theory*, 27(4), 562–580. <https://doi.org/10.1093/jopart/mux023>
- Patt, A., & Zeckhauser, R. (2000). Action Bias and Environmental Decisions. *Journal of Risk and Uncertainty*, 21(1), 45–72. <https://doi.org/10.1023/A:1026517309871>
- Petersen, M. B. (2012). Social welfare as small-scale help: Evolutionary psychology and the deservingness heuristic. *American Journal of Political Science*, 56(1), 1–

16. <https://doi.org/10.1111/j.1540-5907.2011.00545.x>
- Petersen, M. B. (2015). Evolutionary political psychology: On the origin and structure of heuristics and biases in politics. *Political Psychology*, 36(S1), 45–78. <https://doi.org/10.1111/pops.12237>
- Petersen, M. B., Sznycer, D., Cosmides, L., & Tooby, J. (2012). Who Deserves Help? Evolutionary Psychology, Social Emotions, and Public Opinion about Welfare. *Political Psychology*, 33(3), 395–418. <https://doi.org/10.1111/j.1467-9221.2012.00883.x>
- Petersen, N. B. G. (2020). How the source of performance information matters to learning on the front-lines: evidence from a survey experiment. *International Public Management Journal*, 23(2), 276–291. <https://doi.org/10.1080/10967494.2019.1699214>
- Petersen, N. B. G., Laumann, T. V., & Jakobsen, M. (2019). Acceptance or disapproval: Performance information in the eyes of public frontline employees. *Journal of Public Administration Research and Theory*, 29(1), 101–117. <https://doi.org/10.1093/jopart/muy035>
- Sinclair, R. C. (1988). Mood, categorization breadth, and performance appraisal: The effects of order of information acquisition and affective state on halo, accuracy, information retrieval, and evaluations. *Organizational Behavior and Human Decision Processes*, 42(1), 22–46. [https://doi.org/https://doi.org/10.1016/0749-5978\(88\)90018-0](https://doi.org/https://doi.org/10.1016/0749-5978(88)90018-0)
- Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European Journal of Operational Research*, 177(3), 1333–1352. <https://doi.org/10.1016/j.ejor.2005.04.006>
- Thorpe, A., Sirota, M., Juanchich, M., & Orbell, S. (2020). Action bias in the public’s clinically inappropriate expectations for antibiotics. *Journal of Experimental Psychology: Applied*, 26(3), 422–431. <https://doi.org/10.1037/xap0000269>

Online Supplementary Appendices

A1. Political Agreement on Guiding Objectives and Strengthened Quality Development²

Guiding Objectives and Strengthened Quality Development

The negotiating team agrees that the work related to quality assurance and development in the upper secondary education programs should adhere to guiding objectives. The work should be supported to achieve uniform, high educational quality and effective implementation of the reform of the upper secondary education programs.

Guiding Objectives

In collaboration with teachers, the management at individual institutions should have better opportunities to develop schools and teaching in line with local conditions. In the implementation of the agreement, emphasis should be placed on removing unnecessary requirements and process regulations, allowing schools greater flexibility in organizing their teaching.

Guided by the purposes of upper secondary education programs, guiding objectives for the programs are introduced. These objectives should be unburdened by bureaucracy and provide guidance for quality development while promoting a general reduction in bureaucracy within the programs. This includes simplifying the requirements for schools' quality work and obliging schools to work toward the guiding objectives. The reduction in bureaucracy will be carried out in a process that includes input from the sector.

The guiding objectives should set the direction for students' learning and academic level. The aim is for schools to challenge all students to reach their full

² <https://dpt.dk/wp-content/uploads/2020/01/160603-Styrkede-gymnasiale-uddannelser.pdf>

potential, ensuring that all students, regardless of their social background, realize their potential to the fullest, including through high-quality educational offerings.

Building upon the current well-being surveys conducted in accordance with the Education Environment Act, the objectives should also guide efforts to enhance students' well-being with the goal of improving student well-being in upper secondary education programs.

In accordance with the purpose of upper secondary education programs, the guiding objectives should also encompass the use of upper secondary education programs as a basis for higher education, including the involvement of a panel of stakeholders from higher education to assess the quality of the programs. The goal is to enhance the quality of upper secondary education programs in relation to higher education and increase the number of students who use their upper secondary education as a basis for higher education.

As a starting point, the concretization of the guiding objectives should be based on existing data, except for new well-being surveys and data collection from the stakeholder panel.

An investigation will be launched into the development of the academic level of students in the subjects of Danish and mathematics, with a focus on the past approximately 30 years, with the aim of increasing knowledge for the work of ensuring that all students reach their full potential.

The negotiating team will be involved in shaping the guiding objectives, with ongoing monitoring through measurable specific goals.

Continuous Monitoring through Measurable Specific Goals

Concrete goals will be linked to each of the guiding objectives, allowing for national and institutional-level tracking of progress. Each institution will determine its specific

goals, primarily based on data already used for school quality assessment. Goals related to student well-being will involve the introduction of annual well-being measurements.

These specific goals will be delineated based on factors such as dropout rates, educational potential, academic performance, examination results, transition to higher education, and well-being. The concretization of the goal to reduce the impact of students' social background on their results will be achieved by using the same data categorized by socioeconomic background and academic prerequisites.

Furthermore, a stakeholder panel, consisting of representatives from higher education and other relevant parties, will be established to monitor progress with the aim of increasing higher education and providing recommendations to enhance goal fulfillment.

Table A2. Full List of Performance Information Cues

Performance dimension	Performance attribute	Exact wording in Danish
Socioeconomic mobility	<ul style="list-style-type: none"> • Exam grades are far above what can be expected from student backgrounds • Exam grades are slightly above what can be expected from student backgrounds • Exam grades are as can be expected from student backgrounds • Exam grades are slightly below what can be expected from student backgrounds • Exam grades are far below what can be expected from student backgrounds 	<ul style="list-style-type: none"> • Prøvekarakterer langt bedre end det forventede ud fra elevernes baggrund • Prøvekarakterer lidt bedre end det forventede ud fra elevernes baggrund • Prøvekarakterer samme som det forventede ud fra elevernes baggrund • Prøvekarakterer lidt dårligere end forventede ud fra elevernes baggrund • Prøvekarakterer langt dårligere end det forventede ud fra elevernes baggrund
Transition to higher education	<ul style="list-style-type: none"> • Far more enrolled in higher education compared to the national average. • Slightly more enrolled in higher education compared to the national average. • Same proportion enrolled in higher education as the national average. • Slightly fewer enrolled in higher education compared to the national average. • Far fewer enrolled in higher education compared to the national average. 	<ul style="list-style-type: none"> • Langt flere i gang med en uddannelse end på landsplan • Lidt flere i gang med en uddannelse end på landsplan • Samme andel i gang med en uddannelse som på landsplan • Lidt færre i gang med en uddannelse end på landsplan • Langt færre i gang med en uddannelse end på landsplan
Student wellbeing	<ul style="list-style-type: none"> • Far better than the national average • Slightly better than the national average 	<ul style="list-style-type: none"> • Langt bedre end landsgennemsnittet • Lidt bedre end landsgennemsnittet

	<ul style="list-style-type: none"> • Same as the national average • Slightly worse than the national average • Far worse than the national average 	<ul style="list-style-type: none"> • Samme som landsgennemsnittet • Lidt dårligere end landsgennemsnittet • Langt dårligere end landsgennemsnittet
Organizational effort	<p><i>Future effort</i></p> <ul style="list-style-type: none"> • Significant focus on developing the teaching in the future. • Moderate focus on developing the teaching in the future. • Limited focus on developing the teaching in the future. • No focus on developing the teaching in the future. <p><i>Past effort</i></p> <ul style="list-style-type: none"> • Have implemented many initiatives to improve the quality in the past year. • Have implemented some initiatives to improve the quality in the past year. • Have implemented a few initiatives to improve the quality in the past year. • Have not implemented initiatives to improve the quality in the past year. 	<p><i>Future effort</i></p> <ul style="list-style-type: none"> • Stort fokus på at udvikle undervisningen fremadrettet • Moderat fokus på at udvikle undervisningen fremadrettet • Begrænset fokus på at udvikle undervisningen fremadrettet • Intet fokus på at udvikle undervisningen fremadrettet <p><i>Past effort</i></p> <ul style="list-style-type: none"> • Har gennemført mange initiativer for at forbedre kvaliteten det seneste år • Har gennemført nogle initiativer for at forbedre kvaliteten det seneste år • Har gennemført enkelte initiativer for at forbedre kvaliteten det seneste år • Har ikke gennemført initiativer for at forbedre kvaliteten det seneste år
Organizational need	<p><i>Need - origin</i></p> <ul style="list-style-type: none"> • Has a high proportion of students with a non-Western background. 	<p><i>Need - origin</i></p> <ul style="list-style-type: none"> • Har en høj andel af elever med ikke-vestlig baggrund

	<ul style="list-style-type: none"> • Has an average proportion of students with a non-Western background. • Has a low proportion of students with a non-Western background. <p><i>Need – socioeconomic</i></p> <ul style="list-style-type: none"> • Has a high proportion of parents without a higher education. • Has an average proportion of parents without a higher education. • Has a low proportion of parents without a higher education. 	<ul style="list-style-type: none"> • Har en gennemsnitlig andel af elever med ikke-vestlig baggrund • Har en lav andel af elever med ikke-vestlig baggrund <p><i>Need - socioeconomic</i></p> <ul style="list-style-type: none"> • Har en høj andel af forældre uden en videregående uddannelse • Har en gennemsnitlig andel af forældre uden en videregående uddannelse • Har en lav andel af forældre uden en videregående uddannelse
--	--	--

Table A3. Full list of maginal means (as reported in Figures 1 and 2)

Attribute	Level	Marginal mean	CI lower	CI upper
Student wellbeing	Far worse than the national average	5.73	5.64	5.83
	Slightly worse than the national average	6.22	6.12	6.32
	Same as the national average	6.64	6.54	6.73
	Slightly better than the national average	6.73	6.63	6.83
	Far better than the national average	6.95	6.84	7.05
Transition to higher education	Far fewer enrolled in higher education compared to the national average.	5.87	5.77	5.97
	Slightly fewer enrolled in higher education compared to the national average.	6.19	6.09	6.3
	Same proportion enrolled in higher education as the national average.	6.63	6.53	6.73
	Slightly more enrolled in higher education compared to the national average.	6.64	6.54	6.73
	Far more enrolled in higher education compared to the national average.	6.93	6.84	7.03
Socioeconomic mobility	Exam grades are far below what can be expected from student backgrounds.	5.73	5.63	5.84
	Exam grades are slightly below what can be expected from student backgrounds.	6.29	6.19	6.38
	Exam grades are as can be expected from student backgrounds.	6.6	6.5	6.7
	Exam grades are slightly above what can be expected from student backgrounds.	6.78	6.68	6.88
	Exam grades are far above what can be expected from student backgrounds.	6.85	6.75	6.95
Past effort	Have implemented a few initiatives to improve the quality in the past year.	6.56	6.45	6.68
	Have implemented many initiatives to improve the quality in the past year.	6.87	6.75	6.99
	Have implemented some initiatives to improve the quality in the past year.	6.75	6.63	6.87
	Have not implemented initiatives to improve the quality in the past year.	6.05	5.93	6.18
Future effort	Limited focus on developing the teaching in the future.	6.12	6	6.23
	Moderate focus on developing the teaching in the future.	6.49	6.38	6.61
	No focus on developing the teaching in the future.	5.86	5.74	5.98

	Significant focus on developing the teaching in the future.	6.9	6.78	7.03
Need socioeconomic	Has a high proportion of parents without a higher education.	6.58	6.48	6.68
	Has a low proportion of parents without a higher education.	6.36	6.25	6.47
	Has an average proportion of parents without a higher education.	6.46	6.36	6.56
Need origin	Has a high proportion of students with a non-Western background.	6.61	6.5	6.71
	Has a low proportion of students with a non-Western background.	6.29	6.19	6.4
	Has an average proportion of students with a non-Western background.	6.42	6.32	6.53

ⁱ Are peer efforts not performance? Because efforts relates to the activity undertaken to reach a desired output or outcome, they are the means to and end and thus, by definition, not performance (Andersen et al., 2016).