

(Dys)functionality of intentions or outcomes? Performance funding of Danish schools

Performance
funding of
Danish schools

Morten Lund Poulsen

*Department of Business Development and Technology, Aarhus Universitet,
Aarhus, Denmark*

Per Nikolaj Bukh

Aalborg Universitet, Aalborg, Denmark, and

Karina Skovvang Christensen

Department of Economics and Management, Aarhus Universitet, Aarhus, Denmark

Received 1 December 2020

Revised 26 July 2021

25 January 2022

11 April 2022

Accepted 12 April 2022

Abstract

Purpose – This paper studies how performance funding of education is perceived by principals, teachers and administrative staff and management. The dysfunctionality of performance measures often reflects how the measures prevent an organisation from achieving its goals. This paper proposes that perceptions of dysfunctionality can be analysed by separating the perceptions of the programme's intentions, of the school-level actions and of the outcomes for students.

Design/methodology/approach – Following a qualitative methodology, semi-structured interviews were conducted with teachers, school management, staff specialists and top management in a large Danish municipality when outcome-based funding was introduced.

Findings – The performance-funding programme affected teaching by changing educational priorities. Different perceptions of the (dys)functionality of intentions, actions and outcomes fuelled diverging responses. Although the performance measure was generally considered incomplete, interviewees' perceptions of the financial incentivisation and the dysfunctionality of actions depended on interpretations of the incentivisation and student-related outcomes of the programme.

Research limitations/implications – Dysfunctionality can be contested; the interpretations of the intention of a performance-funding programme affect the perceived dysfunctionality of reactions. Both technical characteristics of funding schemes and administrators' and principals' mediating roles are essential for the consequences of performance funding.

Originality/value – The paper examines conditions for dysfunctionality of performance measures. We demonstrate that actions can be perceived as dysfunctional because of a measurement's intentions, actions themselves and the actions' outcomes. Further, the paper illustrates how the reception of performance funding depends on how consequences are enacted based on educators' interpretations of the (dys)functionality of intentions, actions and outcomes.

Keywords Performance funding, Unintended consequences, Dysfunctionality, Incentives, Intentions and actions, Public sector

Paper type Research paper

1. Introduction

Performance measurement has become widespread in the public sector as an instrument to increase accountability (Van Hengel *et al.*, 2014; Steccolini *et al.*, 2020). When performance measures are used to hold publicly funded educational institutions accountable, student

The authors wish to acknowledge comments received at a research seminar at Örebro University School of Business, 15 January 2021, at the International EIASM Public Sector Conference, 27–28 September 2021, at the 11th EIASM Conference on Performance Measurement and Management Control, 15–17 September 2021 and at the Nordic Accounting Conference, Copenhagen Business School, 11–12 November 2021. The authors are also grateful for the valuable comments from Hans Englund and three anonymous referees.



testing is often used to assess not only students' learning outcomes but also schools' performance (Figlio and Ladd, 2015; Hasselbladh and Bejerot, 2020). Although potentially useful, scholars (e.g. Arnaboldi *et al.*, 2015; Muller, 2018; Siverbo *et al.*, 2019; van Thiel and Leeuw, 2002) have debated whether performance measurement in the public sector could be counterproductive if its adverse unintended consequences outweigh its benefits.

Following Parsons (1951), the consequences of using performance measures are termed "functional" when they "contribute to the goal attainment of organisations, and 'dysfunctional' when they obstruct an organisation from achieving its goals" (Vakkuri and Meklin, 2006, p. 242). When performance measurement is studied from a functionalistic perspective, dysfunctional consequences are unintended and often unanticipated (e.g. Ortagus *et al.*, 2020; Smith, 1993, 1995; Vakkuri and Meklin, 2006). However, unforeseen consequences should not be automatically identified with consequences that are necessarily undesirable (Merton, 1936), as unintended consequences can be functional (Van Helden *et al.*, 2012).

Numerous factors influence the consequences of tests, but testing's stakes play an important role (Luxia, 2005). Although the concept of stakes is relatively vague, testing's stakes are generally described as low if the results have little to no significance and high if they entail important consequences for students or educators (Au, 2007). Stakes can be raised by, for example, influencing students' possibilities for further education, firing teachers or closing schools (Andersen and Nielsen, 2020), or through rewards for performance improvements. This paper focuses on performance funding (Dougherty *et al.*, 2016), where funding depends on specific performance measures, thus raising their stakes (Li and Ortagus, 2019).

In education, performance funding is often based on student count, e.g. the number of students enrolled (Agyemang, 2010; Bjørnenak, 2000) or graduating (Aliabadi *et al.*, 2021). The measures can also be outcome-based, e.g. graduates' average wages (Ortagus *et al.*, 2020) or students' test results (Al-Samarrai *et al.*, 2018). However, performance funding, similar to performance measurement in general (Muller, 2018), often fails to achieve its intentions or has unintended consequences (Bell *et al.*, 2018; Hillman *et al.*, 2015; Ortagus *et al.*, 2020; Umbricht *et al.*, 2017).

To affect teaching practice, the financial pressures aligned with a performance measure must trickle down to individual educators through intervening or mediating layers (Jongbloed and Vossensteyn, 2016). Performance measurement can be incomplete, failing to completely capture organisational performance (Siverbo *et al.*, 2019), e.g. with difficulties in identifying cause-and-effect relationships (Hyndman and Eden, 2000). Incompleteness, along with disagreements on organisational missions or uncertainty regarding the means to achieve goals, can create ambiguous situations (Vakkuri and Meklin, 2006), where an excessive reliance on quantitative measures of performance and in turn mechanistic responses to the information provided may be inappropriate (Earl and Hopwood, 1980).

This paper focuses on the consequences of raising the stakes for schools directly via performance funding related to exam results. Within basic education, Al-Samarrai *et al.* (2018) found that performance funding increased students' test results. However, student testing is an incomplete measure of school performance (e.g. Hasselbladh and Bejerot, 2020). The literature on the unintended consequences of high-stakes testing in basic education (Au, 2007; Amrein-Beardsley *et al.*, 2010; Deming *et al.*, 2016) indicates that performance funding can have dysfunctional consequences. Accordingly, we aim to address two research questions: (1) How and why is outcome-based performance funding responded to? (2) What influences the (dys)functionality of responses?

To answer the questions, we investigate performance funding in a case study of the "Financing to Enhance Academically Low-Performing Students" (FEAWS), a Danish performance-funding programme. The programme introduced an exit exam-based

performance-funding scheme to incentivise enhancing outcomes for students who were performing poor academically (DME, 2017e). Thus, the FEAWS presents an opportunity to explore (dys)functionality and intentionality when performance funding trickles down to teachers. The case study focuses on a large Danish municipality with multiple schools participating in the programme.

We find that the performance-funding programme affected teaching by changing educational priorities. Different perceptions of the consequences and (dys)functionality of specific initiatives led to two diverging responses to the programme. Although it was generally agreed that the performance measure was incomplete, interviewees' perceptions of the financial incentivisation and actions' (dys)functionality differed. The divergence depended on interpretations of incentivisation and the student-related outcomes of the programme.

The paper makes two overall contributions to extant literature. First, the paper contributes to the literature on performance measurement and dysfunctionality (Siverbo *et al.*, 2019; Smith, 1995; Vakkuri and Meklin, 2006) by examining conditions for performance measures to be dysfunctional. Intentionality and (dys)functionality are often conflated in the literature. Even when they are not (e.g. Van Helden *et al.*, 2012), (dys)functionality typically concerns actions or their outcomes, but not actions' relationship to intention. We propose that actions can also be perceived as dysfunctional because of how and why they are responses to intentions. We demonstrate the three aspects of actions' (dys)functionality and highlight how interpretations of the performance funding scheme affect prioritisations, the resulting action and the perceptions of consequences.

Second, the paper contributes to the literature on performance funding in education (Bell *et al.*, 2018; Ortagus *et al.*, 2020; Umbricht *et al.*, 2017). Research is divided on whether performance funding fails to achieve its intentions and causes unintended consequences because of design (Dougherty and Reddy, 2013; Jongbloed and Vossensteyn, 2016; Ortagus *et al.*, 2020) or inherent issues (Bell *et al.*, 2018; Levačić, 2008; Umbricht *et al.*, 2017). We illustrate how the reception of performance funding depends on both perceived incompleteness and perceived dysfunctionality of financial incentivisation. Further, we show how consequences are enacted based on educators' interpretations of intentions, actions and outcomes' (dys)functionality.

The remainder of this paper is organised as follows. Section 2 presents the literature on high-stakes testing, performance funding, performance measurement and dysfunctionality, developing our conceptual framework. Section 3 describes the research setting and methodology, and Section 4 contextualises the research and presents its two interpretations. Subsequently, Section 5 presents how the interpretations perceived dysfunctionality. Finally, Section 6 concludes the paper.

2. Literature review and theoretical framing

Although other educational goals are important, academic achievements often take the highest priority (e.g. Hasselbladh and Bejerot, 2020). Academic performance is typically represented by test results, expressing the performance of students, teachers, schools and countries (Hasselbladh and Bejerot, 2020; Ratner, 2020). However, testing is subject to Goodhart's Law, originally posited in a paper discussing control over money supply (Goodhart, 1975, cf. Pidd, 2005) as "any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes". This leads to the so-called performance paradox (Meyer and Gupta, 1994), where performance indicators no longer measure actual performance (van Thiel and Leeuw, 2002).

Within the education literature (Pearson, 1988; Alderson and Wall, 1993), the consequences of tests and exams on teaching and learning at the classroom level are

usually called washback. Washback distinguishes consequences of tests from the impact of tests and exams on “individuals, policies, or practices, within the classroom, the school the educational system, or the society as a whole” (Wall, 1997, p. 291). Numerous factors influence washback, but testing’s stakes play an important role (Luxia, 2005). Although the concept of stakes is relatively vague, stakes are generally described as low if they have little to no significance and high if they entail important consequences for students or educators (Au, 2007).

2.1 Performance funding

Performance-based funding implies that funding depends on performance measures (e.g. Aliabadi *et al.*, 2021), thereby making educational institutions financially accountable to measured results either as additional funding (i.e. a bonus) or as part of ordinary funding (Ortagus *et al.*, 2020). Performance funding is now commonly used across countries as a policy mechanism, although the specific designs differ (Jongbloed and Vossensteyn, 2016). The performance measures used in performance funding are often cost drivers (Bjørnenak, 2000; Li and Ortagus, 2019) or outputs (Umbricht *et al.*, 2017); however, examples of outcomes also exist (Ortagus *et al.*, 2020).

Performance funding can depend directly or indirectly on test results. For example, student count can be an indirect performance measure if the funding mechanism creates competitive pressure to retain or obtain students (Levačić, 2008) by focussing on students’ academic results (e.g. Hasselbladh and Bejerot, 2020). It is less common to use performance funding based on test results directly to fund basic education. To the best of our knowledge, the only paper examining such a funding mechanism in basic education is Al-Samarrai *et al.* (2018). These authors studied the Jarkatan school system and found that a funding scheme with a bonus conditional on increases to schools’ performance relative to average performance one and two years previously positively impacted students’ test performance compared to other schools.

Performance funding often has a modest effect or does not achieve its intended results (Bell *et al.*, 2018). Furthermore, performance funding can have unintended consequences, e.g. a focus on short-term certificate production at the expense of degrees with higher labour market returns (Hillman *et al.*, 2015, 2018; Li and Ortagus, 2019), increasingly selective admission (Bell *et al.*, 2018; Gándara and Rutherford, 2020; Umbricht *et al.*, 2017), reduced degree production among weaker student groups (Ortagus *et al.*, 2021), funding disparities among institutions (Favero and Rutherford, 2020; Hagood, 2019) and artificial grade inflation (Hasselbladh and Bejerot, 2020).

The reason for this failure is contested. One explanation is policy design, which concerns models having various surmountable “obstacles” (Dougherty and Reddy, 2013, p. 13) that prevent performance funding from functioning as intended, such as a mismatch between funding and other policies affecting practice (Jongbloed and Vossensteyn, 2016, pp. 590–591). Hillman *et al.* (2015, pp. 515–516) noted that institutions might not know of causal mechanisms or have the capacity to respond to incentives. Similarly, Ortagus *et al.* (2020) argued that intended consequences could be achieved and unintended ones avoided if funding schemes present straightforward, well-designed incentives as part of ordinary funding.

Moreover, several studies (Kelchen, 2018; Gándara and Rutherford, 2018; Li and Ortagus, 2019; Ortagus *et al.*, 2021) have found that schemes based on additional funding for specific focus areas, e.g. attracting student groups, could counteract unintended consequences, although the particular design of the performance scheme seems to be important. Creating more sophisticated funding models can be difficult. Only recently have researchers (e.g. Rosinger *et al.*, 2021) examined how specific design features, such as the share of funding being dependent on performance, differ among funding schemes, meaning that not much

research-based guidance on design specifics is available for policy makers. However, even when present, the development of funding schemes often ignores research evidence (Gándara, 2019).

Alternatively, performance funding might fail to achieve its intended goals because performance measurement has inherent problems that are not easily overcome. Because of the intrinsic complexity of the public sector (Lapsley and Skærbæk, 2012), quantitative measures of performance and mechanistic responses to the information provided may be inappropriate (Earl and Hopwood, 1980). Additionally, ambiguous objectives and lack of clear cause-and-effect relationships can mean that a tight coupling between mission, objectives and targets may be neither possible nor appropriate (Hyndman and Eden, 2000).

In the performance funding literature, Bell *et al.* (2018), Umbricht *et al.* (2017, p. 667) and Levačić (2008, pp. 230–231) argued that the implicit assumption in performance funding that institutions need incentivisation to achieve better results might miss the point that factors outside of institution-level control determine performance. When performance funding based on enrolment is used to promote competition among universities to improve quality, the policy initiative rests, e.g. on the assumption that less attractive universities can raise their attractiveness. Research by Pischedda and Marinò (2021) suggests otherwise, indicating that attractiveness is highly influenced by factors unrelated to teaching quality, e.g. the institution's year of foundation or geographical location.

2.2 Performance measurement, (dys)functionality and intentionality

Performance measurement can be incomplete, particularly in the public sector (Siverbo *et al.*, 2019). Incompleteness means that a measure's meaning is “created through coherent and convincing histories to be told to members of the organisation, other organisations, constituencies and stakeholders” (Vakkuri and Meklin, 2006, p. 241) rather than being solely represented by its design or intentions. Vakkuri and Meklin further explained that performance measurement systems' dysfunctionality is typically defined from a functionalist perspective that assumes “a clear, unambiguous interpretation of organisational mission, goals and the mechanism through which the achievement of goals could be assessed” (2006, p. 242). When the linkage between goals and their attainment mechanisms is uncertain, dysfunctionality is conditioned on interpretations of organisational missions, goals and mechanisms' appropriateness, making for an ambiguous situation wherein decision-makers must manage uncertainty (Vakkuri and Meklin, 2006). Such interpretations become particularly important when more goals are formulated broadly rather than directly related to employees' specific work.

Although numerous forms of dysfunctionality can be listed (e.g. Smith, 1993, 1995; van Thiel and Leeuw, 2002), three overall categories of dysfunctional actions have been emphasised (Siverbo *et al.*, 2019; Vakkuri and Meklin, 2006). First, in *behavioural displacement* (Siverbo *et al.*, 2019, p. 1804), the design of performance measurement comes to mislead action. If measures condition a situation wherein actors must adapt, they can decrease performance despite their intentions to improve it. It can happen, for example, via “tunnel vision” (Smith, 1995), where important actions are unintentionally avoided because they do not improve what is measured. Second, in the intentional *strategic manipulation of incompleteness*, often termed “gaming” (Bevan and Hood, 2006), the incompleteness of the measurement system is used to increase the measured performance at the expense of actual performance (Roberts, 2018). For example, it can distort organisational output by focussing on goals that are more accessible rather than those that are more challenging to achieve but more worthwhile (Ortagus *et al.*, 2020). Third, affective reactions as *negative attitudes* can adversely affect employees' relational connection to performance measurement. Such attitudes may negatively affect other activities, e.g. through demotivation and undermining motivation

and morale (Arnaboldi *et al.*, 2015). Further, negative attitudes as reactions to performance measurement can lead to “job tension, conflict, frustration and resistance” (Siverbo *et al.*, 2019, p. 1804).

Measuring performance alone does not necessarily affect action (e.g. Ter Bogt *et al.*, 2015; Van Hengel *et al.*, 2014). However, Siverbo *et al.* (2019) suggested that as control tightens, the likelihood of dysfunctional consequences increases. Siverbo *et al.* (2019) also suggested that using results for reimbursement purposes would tighten control. Jongbloed and Vossensteyn (2016, p. 591) argued that financial incentives must trickle down from an organisational level to that of individual educators who decide how to react in an ambiguous situation.

Research on performance measurement within the public sector has demonstrated that performance measurement implementation can have unintended and dysfunctional consequences (e.g. Bevan and Hood, 2006; Broadbent and Laughlin, 1998; Smith, 1995). The literature on high-stakes testing (Amrein-Beardsley *et al.*, 2010; Au, 2007; Jennings and Bearak, 2014) finds comparable results, emphasising unintended consequences, including gaming, e.g. cheating when grades are intentionally inflated artificially (Hasselbladh and Bejerot, 2020), and behavioural displacement. The basic education literature often identifies three major groupings of unintended consequences. First, *teaching-to-the-test* (Amrein-Beardsley *et al.*, 2010) concerns classroom practice wherein students are taught test-taking strategies to inflate scores without improving actual abilities. Second, *narrowing the curriculum* (Au, 2007) concerns the relative weight of tested material in teaching, crowding out worthwhile but untested subjects. Finally, *focussing on marginal students* (Deming *et al.*, 2016) concerns the relative effort spent on students or grades wherein some students acquire more educational resources because their improvement affect targets the most.

However, these three groups’ (dys)functionality and intentionality are complex. For example, Shirrell (2016) demonstrated inconsistent understandings of accountability and, thereby, different school-level interpretations of policy intentions and their alignment with organisational mission. Jennings and Bearak (2014) argued that teaching-to-the-test could increase testing validity by enabling students to demonstrate their abilities better. Similarly, Van Helden *et al.* (2012) and Speklé and Verbeeten (2014) suggested that performance measurement could introduce functional changes irrespective of intentionality. Narrowing the curriculum or additional focus on marginal students can also be intended if a policy is meant to achieve a more significant focus on neglected subject areas or student groups such as minorities (e.g. Umbricht *et al.*, 2017). Such efforts only become dysfunctional when harming other subjects or student groups.

Because actions can be perceived as dysfunctional regardless of intentionality, the intentionality–(dys)functionality relationship is ambiguous. However, intentionality and dysfunctionality are often conflated, e.g. in Ortagus *et al.* (2020), Smith (1993, 1995) and Vakkuri and Meklin (2006), because unintended consequences are often dysfunctional. Even when the distinction is made explicit (e.g. in Van Helden *et al.*, 2012), intentions and dysfunctions are conceptualised as related to the outcomes of actions. Smith (1993, 1995), Siverbo *et al.* (2019) and Vakkuri and Meklin (2006) showed how actions themselves could also be dysfunctional. We introduce a third possible type of dysfunctionality.

We propose, as shown in Figure 1, that actions can be perceived as dysfunctional in three different ways: First, they can be dysfunctional if they do not achieve *outcomes* that are aligned with the organisational mission, e.g. by narrowing the curriculum at the expense of actual learning, which concerns the perceived action–consequence relation. Second, the *actions* by themselves can be dysfunctional even though they do not necessarily impair outcomes, e.g. by inflating grades or gaming to gain additional resources to a school. Third, and extending the earlier literature, we suggest that actions can be perceived as dysfunctional because of a rejection of the *intention*, e.g. by being critical of a policy intention regardless of the outcome of the action. The framework is summarised in Table 1.

3. Research setting and methodology

3.1 Schooling and testing in Denmark

Per the Danish constitution, all children aged 6–16 are entitled to free education in public basic education schools, the *Folkeskole*, encompassing primary and lower secondary education. The *Folkeskole* aims to develop students’ academic abilities for further education and enhance students’ understanding of and appreciation for intellectual freedom, equality and democratic values (DME, 2018a). In 2018, around 540,000 (76%) of the 710,000 Danish basic education students attended 1,082 public schools, averaging about 500 students per school. The remaining students were enrolled in various forms of private schools, partially publicly financed. The Ministry of Education sets the overall objectives and directions of the system; the 98 Danish municipalities maintain political and financial responsibility with a high degree of freedom in organising and funding the schools.

In Denmark, 10 years of basic education is mandatory. Education starts with grade zero, named so for historical reasons, and ends with compulsory exit exams in the ninth grade. The exams are graded according to academic objectives, which align with course objectives and skills generally considered important by teachers (Beuchert and Nandrup, 2018). The compulsory exit exams have traditionally had low stakes (Egelund, 2005), holding few consequences for educators or students. However, poor results in international comparisons in the early 2000s called the “PISA shock” (PISA being the Program for International Student Assessment) in the Danish educational debate meant that policymakers “raised questions about the quality of education, given the well-funded Danish school [system]” (Ratner, 2020, p. 217). Successive governments introduced reforms and national testing to increase students’ academic performance (DME, 2013). However, although schools performing poorly in the exit exams can come under stricter municipal or ministerial supervision, the exams only present low-powered incentives to educators relative to Anglo-Saxon systems (Andersen and Nielsen, 2020).

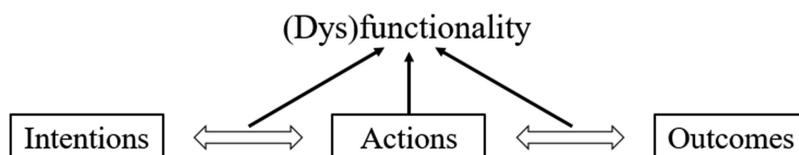


Figure 1.
Perceived (dys)
functionality of actions
and its relation to
intentions and
outcomes

| | Description | Example |
|------------|--|--|
| Intentions | The action is dysfunctional because of how it relates to an intention: the intention and thus acting upon it conflicts with the organisational mission | Proposed additional dimension to (dys) functionality, e.g. suggested by negative attitudes towards performance measurement, where acting on such a negatively understood measure could be dysfunctional irrespective of the resultant action or its outcomes |
| Actions | The action is dysfunctional because of how it itself conflicts with the organisational mission | Strategic reactions by abusing measures could on their own conflict with organisational mission, e.g. teaching-to-the-test could itself conflict with organisational mission without achieving poorer student outcomes |
| Outcomes | The action is dysfunctional because of its related outcomes: the outcomes that the action has conflict with organisational mission | Strategic reactions or behavioural displacement could be dysfunctional because of their outcomes, e.g. narrowing the curriculum could increase student test scores while negatively affecting actual student learning |

Table 1.
Types of
dysfunctionality

Exams are graded using the Danish grading scale. It is comparable to the European Credit-Transfer and Accumulation System with five passing grades of “02” (E), “4” (D), “7” (C) “10” (B) and “12” (A), and two failing grades, “-3” (Fx) and “00” (F). It is an absolute scale wherein grades are given relative to an individual student’s abilities concerning educational goals rather than to other students’ performance (DME, 2020). The grade “02” indicates a minimally acceptable performance; “4” is given for an adequate exam performance that shows some command over the relevant materials but has significant flaws; “7” is a good performance with some weaknesses (DME, 2007). In 2018, 1.3% of students achieved a grade point average (GPA) below 02 (DME, 2021). The national GPA for graduating students in 2018 was 7.3, with 6.3 and 7.9 in written and oral Danish and 7.0 in written mathematics (DME, 2018b).

3.2 FEAWS and its incentive

Shortly after the Danish Liberal Party won the election in 2015, the newly formed government announced the FEAWS to improve outcomes for students performing poorly (DME, 2017e), arguing that academically weak or low-performing students, operationalised as those achieving a GPA below 4, generally have poor educational outcomes such as a lower transition rate to further education (DME, 2017d).

The programme aimed to “raise the incentive to develop new educational methods and create academic progress” (DME, 2017c, own translation) through a bonus conditional on specific performance increases. The potential bonus was DKK 1.3M–1.5M (DME, 2017c) or around €175,000–€200,000 (DME, 2017a), which could finance three teachers for a school year and represented 5–8% of an ordinary school budget. Participant selection was based on the number of students achieving below a grade of 4 at the compulsory exit exams in Danish (reading, spelling, a written presentation and an oral examination) and mathematics (two exams: 1 h without aids and 3 h with aids). The results were averaged over three school years preceding the programme to create a baseline ranging from 28% to 62% (mean 40%). The bonus was conditional on improvements of 5, 10 and 15 percentage points or above against the baseline over the following three years. The programme selected 127 schools from the entire country. Participation was voluntary, and 104 schools chose to participate. The incentive model and an example of a hypothetical representative school are shown in Table 2.

As Table 2 shows, considerable differences might exist in “weak students” between years, implying that the selection criteria, i.e. having a high number of weak students, is not a precise measure. Furthermore, achieving the targets compared to the baseline might not accurately measure schools’ increased performance. Thus, the target difficulty has an element of arbitrariness. Moreover, the difficulty of the target increases annually, while the annual bonus remains the same.

Concurrently with the incentive, the Ministry also established a voluntary supplementary development initiative, the “Program for Enhancing Students” (PES) (DME, 2017b). Participation in the PES was free and included courses, workshops, networking and aid from the Ministry’s staff specialists focussing on paren–teacher collaboration, continuous evaluation and feedback, intensive learning programmes and student-to-student learning (DME, 2017b). Schools had free rein in choosing whether to participate and apply the tools included in the programme (DME, 2017e). Of the 104 schools participating in the FEAWS, 91 took part in the PES.

3.3 Case selection

To purposefully sample reactions to the programme, we chose a single municipality (cf. Miles *et al.*, 2014, pp. 31–32). Keeping contextual parameters such as the ordinary funding scheme, hierarchical relations and municipal projects constant, we focus on the consequences of the performance-funding scheme.

| | |
|------------|--|
| Baseline | The relative number of students achieving below a grade of 4 in the compulsory exams in Danish and mathematics: the baseline is calculated as the average of results in 2014, 2015 and 2016 |
| Bonus size | The size depends on the number of students in the 9th grade – <50 students: €175,000 – 50–100 students: €182,500 – ≥100 students: €200,000 The bonus is paid out yearly a few months after exams |
| Target | The target is the improvement compared to the baseline of 5pp the first year, 10pp the second year and 15pp the third year |
| Example | <i>School:</i> two classes, totalling 45 students in the 9th grade in all three years <i>Weak students,</i> i.e. those achieving below 4: 19 in 2014 (41%), 11 in 2015 (26%) and 16 in 2016 (37%) <i>Bonus size:</i> €175,000 <i>Baseline:</i> $(42\% + 26\% + 37\%)/3 = 35\%$ <i>Target:</i> Achieving the bonus would depend on 13 students (30%) or below achieving 4 the first year, 11 students (25%) the second and 9 students (20%) the third |

Table 2.
The performance-funding scheme and an example

The case municipality has over 300,000 inhabitants, among the largest in Denmark. Approximately 25,000 of the 30,000 municipality’s school-aged children attend one of the 46 public basic education schools. The municipality is organised such that the Children and Youth Administration is responsible for public schools and day care, setting the overall strategy and focus. Five municipal directors are each responsible for institutions in specific districts, and each school has a local management team headed by a principal. The programme selected seven of the municipality’s schools; all decided to participate.

The municipality funds all schools via a scheme that allocates funds through a variable component based on student count, some fixed components, e.g. for management, and a statistical model’s estimates of each school’s special education needs. The schools must fund each student they segregate to special needs education at other schools. Each segregated student costs the schools approximately the amount allocated for four students in the variable component in the funding scheme.

3.4 Data collection

The empirical data comprised semi-structured interviews and documents related to the FEAWS. Moreover, we followed the public debate on the programme, marking different stakeholders’ positions and rationales. To understand the municipal actors’ interpretations, we conducted two rounds of interviews. The first round comprised seven interviews with school management or managers (“SM”) and seven with top management and staff specialists (“members of the administration” or “MA”). The second round comprised six interviews with SM, one with an MA and 11 with teachers (“T”) at five participating schools. The documents consisted of formal information at the policy level issued by the Ministry of Education and were used to analyse the structure and intentions of the programme. Table 3 shows the timing of interviews relative to the cancellation of the scheme.

| Interviewees | Before cancellation | After cancellation | Total |
|-------------------|---------------------|--------------------|-------|
| Administration | 7 | 1 | 8 |
| School management | 7 | 6 | 13 |
| Teachers | | 11 | 11 |
| <i>Total</i> | 14 | 18 | 32 |

Table 3.
Interviewee position and interview timing relative to the cancellation

After a parliament election in 2019, a new Social Democrat minority government was formed. Subsequently, the funding scheme was terminated a year before its scheduled end. However, the PES continued. The programme's termination was not related to the specific results obtained but based on a general resistance to performance-funding among a majority in the new Parliament. Because the second round of interviews was done after the cancellation, the FEAWS unexpectedly no longer affected schools. Although this was not part of the original research design, it enabled a before-and-after view of the reactions to the performance-funding scheme.

SM and MA were contacted via e-mail. All schools agreed to participate. However, one school dropped out due to a restructuring during the second round of interviews. We interviewed the SM member in charge of the programme to focus on the mediating role of this person. We selected teachers for interviews through information gathered in interviews and snowballing (Miles *et al.*, 2014) among interviewees. The first round of 14 interviews took place at the school premises, with one interview conducted by phone. Because of COVID-19 concerns during the programme's third year, interviews were conducted by phone, Skype or Microsoft Teams; a single interview was conducted the following school year.

The interview guides were semi-structured, enabling comparisons and providing flexibility to explore interviewees' interpretations of the programme and its presence in everyday practice (Brinkmann and Kvale, 2015). The first guide was more inductive, with the second becoming more theory-informed (Hall and Messner, 2018). The first guide's four themes included the first impressions of the programme, the decision to participate, implementation and selection of initiatives and consequences. The second's themes included changes since the first interview, the programme in everyday practice, reflections on implementation and long-term consequences. Interviews lasted 45–120 min, averaging approximately 90 min. All interviews were taped and transcribed, totalling around 1,000 pages.

While interviewing, we paid attention to reliability by avoiding leading questions (Brinkmann and Kvale, 2015), transcribing the interviews word-for-word, having the coding reviewed by all authors and including questions and entire interview sequences in the data presented. Validity as gaining an authentic and credible understanding (Parker and Northcott, 2016) was addressed by allowing the interviewees an opportunity to comment on interpretations made at the end of interviews and during the recapitulations of their statements.

3.5 Data analysis

Analyses were performed based on systematic coding using NVivo. We used concept codes (Saldaña, 2016) for statements concerning intentions, actions, outcomes and relationships between them. We also used codes for unintended consequences identified in the educational literature. Additionally, we used time coding and evaluation coding (Saldaña, 2016) to identify "earlier", "now" and "later" as well as positive and negative judgements. This was done to explore changes over time, code the outcomes of actions and understand how the (dys)functionality of intentions, actions and outcomes was perceived.

The analysis was conducted as an iterative process (Eisenhardt, 1989), reviewing the literature, the coded sequences and parts of the transcripts repeatedly. We realised that a pattern emerged during this coding process, where two types of perceptions, interpretations and reactions could be identified. Some interviewees were mainly critical or sceptical or expressed negative attitudes when talking about the FEAWS, even if they appreciated the changes initiated following the programme's introduction. Other interviewees were more neutral in their statements, saw opportunities in the potential for a bonus or expressed a directly positive attitude, expressing willingness to change

teaching practice because of the FEAWS. The main difference between the two groups was whether they actively sought to achieve the bonus. In the analysis, we label the two groups of respondents “sceptical implementers”, shortened to “sceptics”, and “active users”, shortened to “actives”. Although not all interviewees were easily assigned to the grouping, most were, and we coded the interviews according to expressions to form the two subgroups used in the analysis.

In the subsequent section, statements represent general sentiments unless stated otherwise. To ensure anonymity, we avoid statements and actions attributable to specific persons or schools. We found no systematic connection between the different interpretations of the performance-funding scheme and identity. Interview timing is shown using “a” and “b” for the first and second interview periods.

4. Findings

Shortly after the Danish Liberal Party won the election in 2015, the newly formed minority government announced a performance-funding programme to enhance academically weak students. The government allocated a sum of DKK 500M (€67.5M) to a three-year project aimed at “encouraging schools to improve efforts for the academically weakest students”, declaring that “the funds from the scheme will only be paid out to schools that can document raising the level of achievement for the academically weakest students” (Regeringen, 2016, p. 17, own translation).

In the presentation of the programme, it was explicitly stated that it was inspired by the activity-based funding used in the Danish healthcare sector, where “a similar scheme . . . meant that hospital departments developed new methods to treat patients faster and better” (Regeringen, 2016, p. 17, own translation). However, the healthcare funding scheme was based on patient contacts, i.e. an activity measure, rather than outcomes. Similar activity-based funding of basic education based on student count are the norm within most Danish municipalities with roots back to the 1980s (Bjerg, 1991).

Around nine months after the announcement of the initiative, the Minister of Education, Merethe Riisager, and the Prime Minister, Lars Løkke Rasmussen, presented the performance-funding scheme. Intense debates followed its announcement in media; both the opposition and educators’ unions were critical. As a representative example of the critical comments, a politician from the Socialist People’s Party, Jacob Mark, stated that “you can’t run the *Folkeskole* like a business . . . How are the schools supposed to raise the level of achievement when they only get funded afterwards? . . . It results in a narrow-minded focus on grades” (*Folkeskolen.dk*, 25th April 2017, own translation).

After the parliament election in 2019, a new Social Democrat minority government was formed. Subsequently, the funding scheme was terminated a year before its scheduled end. The Social Democrats had not criticised the programme heavily but depended on openly critical socialist supporting parties to form a government. The Social Democrat Minister of Education, Pernille Rosenkrantz-Theil, stated that the programme

. . . only rewards final results and not the prior initiatives. Thus, individual students and teachers end up accountable for whether a school achieves a bonus or not. Consequently, the government is not going to continue such lump sum management of the *Folkeskole* (DME, 2019, own translation).

4.1 Interpretation of the incentive scheme

All the interviewees were aware of the debate surrounding the FEAWS and frequently referred to the political context of the programme. An administration member stated, e.g. that the programme was the product of “a liberal government that wanted to brand itself” (MA3). Other interviewees emphasised that the programme rested on a particular “ideology”.

However, school participation was voluntary, and a principal remarked that “you are employed as a loyal civil servant. You must translate ideas into what is best for the school. You can’t let your political conviction be in the way” (SM3b).

Interviewees often emphasised that an explicit financial incentive was a new funding policy in the Danish educational sector. Nonetheless, they acknowledged that financial incentives were already present because funds were allocated to schools based on student count. Further, an interviewee noted how “running a school well requires controlling the budget, and that is a prerequisite for a pedagogically and didactically strong school” (SM9b). Many interviewees also observed that the public schools already competed with each other and private schools for the academically strongest students. Such competition often focused on schools’ “branding” based on academic competence and specific focus areas, e.g. advanced music classes. Although being selected for the programme made additional funding possible, some schools expressed concern that selection could hurt their reputation. However, as one interviewee explained, the programme “focused on a group of students that really needs a boost that we do not accomplish at the moment” (MA6).

Immediately after the schools selected for the programme were announced, the Children and Youth Administration held a meeting for administrators and principals. It was decided that the programme was to be handled locally at the schools without interference from the municipal administration. An administrator explained how the FEAWS was perceived from the municipal perspective:

We see it as an opportunity to strengthen local efforts underway and . . . a welcome economic helping hand to the schools—if they achieved the targets . . . [but] to avoid putting unnecessary pressure on the schools, we did not announce expectations whether the schools would achieve the targets (MA7).

Shortly after the meeting, all the schools launched various initiatives focussed on enhancing academic abilities. Some initiatives encompassed all the year groups; most were aimed at the ninth-grade students. Following the literature on unintended consequences in education (Amrein-Beardsley *et al.*, 2010), most of the initiatives could be categorised as dysfunctional at first glance. One initiative was to focus on exam preparation. Students traditionally spent time preparing for exams at home; now, this was done part-time at the school, aided by teachers. Guided preparations focused on individual and group coaching for exams, where teachers explicitly mentioned that it was a teaching-to-the-test strategy (cf. Amrein-Beardsley *et al.*, 2010).

Moreover, another prevailing initiative was to increase attention to Danish language and mathematics, including additional lessons at the cost to other subjects, thus narrowing the curriculum (cf. Au, 2007). A third type of response was teacher conferences, where students’ academic progress in Danish and mathematics were analysed and specific initiatives were discussed. Among such initiatives were new student groupings and intensive learning programmes centred on low-performing students, thus focussing on marginal students (Deming *et al.*, 2016).

Most interviewees were critical of using direct financial incentives in education, showing negative attitudes (Siverbo *et al.*, 2019). However, there were differences in how interviewees perceived and responded to the programme. Overall, we identified two diverging interpretations: *sceptical implementers* and *active users* (shortened to “sceptics” and “actives” in the following). Both groups included teachers, principals and administration members. Sceptical implementers doubted the scheme’s usefulness but implemented changes, especially with inspiration from the PES’ supplementary aid. Active users focused directly on and prioritised the incentive scheme. Although the specific initiatives were often similar between groups, sceptics’ efforts were often aimed at all students at the schools, while actives’ efforts were of greater intensity and more targeted.

4.2 Sceptical implementers

To sceptics, the scheme was problematic owing to its “rhetoric that the principal now had to concentrate on the weakest students . . . and that money was a motivation” (SM7a). One principal explained that low-performing students already received attention, potentially at the expense of other students. Further, it was emphasised that addressing the needs of low-performing students requires attention in all grades, not only in the last year of schooling. Neither was a financial incentive perceived as helpful. The interviewees were instead motivated by a desire to achieve the best outcomes for students. One principal explained, “[we] felt a bit insulted by the belief that a carrot could make someone a better teacher” (SM6a). Sceptical teachers agreed, noting that the scheme was often the target of collegial jokes.

Sceptics stated that improvements in student learning only weakly affected the measure. Numerous other factors that were perceived as non-controllable should have been considered, e.g. the performance measure did not account for the substantial differences between the classes from year to year or the possibility of students transferring between schools. Both factors were considered to have a primarily negative effect on scoring without reflecting school-level efforts. Moreover, at some schools, the targets were perceived as challenging; at others, no improvements were required. One principal stated, upon achieving the bonus, “no matter what we had done, [the year group] would have been above the baseline . . . I see it as a lottery ticket” (SM6b). Following these interpretations, the performance measure was perceived to be incomplete (c.f. [Siverbo et al., 2019](#)) and the target arbitrary.

Furthermore, the measure could be affected by actions that had nothing to do with improved learning for low-performing students. A principal mentioned how a manager hypothetically,

could have spent enormous amounts of effort on [the project], which is what we are supposed to do, but if I had to reach the target, I could have lobbied my neighbouring school for their best students (SM9b).

Thus, the principal reasoned that transferring in high-performing or transferring out low-performing students was the easiest way to affect the measure. The principal emphasised that such strategies were not implemented yet remained a possibility. While interpreting the scheme as vulnerable to gaming (cf. [Bevan and Hood, 2006](#)), sceptics maintained that the incentives did not influence them.

Although sceptics doubted the scheme’s usefulness, its introduction was instrumental for changes. New initiatives were implemented to improve academic performance, focussing particularly on the two subjects included in the FEAWS’ performance measure and typically inspired by the PES. One such initiative was the increased attention to and structuring of exam preparation. Initially, a manager “had no great expectations” (SM6a) of the PES but became positive after experiencing the effects of changes implemented in the first year. Teachers mentioned how it helped with “those easy things that actually enhance [student results]” (T11), e.g. using student testing in everyday practice to strengthen exam preparations. The initiatives mainly focused on low-performing students but were “not just useful to the weakest students. It’s just as much for the most talented” (T4); they helped students become more comfortable with and knowledgeable of exam scenarios.

Sceptical managers afforded the scheme limited attention and indicated that they did not change educational practices to achieve targets. Because of not having paid attention to the school’s progress towards the target, one management team was even surprised when informed that their school got the bonus. This inattention was also evident when asked about how the bonus was spent. The same school’s SM noted that the additional funding would just be carried forwards as a buffer to shield against future budgetary risk. Teachers at schools with sceptical managers often did not know whether new initiatives were related to the

FEAWS or how the bonus was spent. Some did not even know of the scheme's early termination. In turn, participating in the FEAWS seemed not to have raised exam-related stakes (Au, 2007) for sceptics.

However, the additional funding was perceived to be important. A principal mentioned that "additional [funding] would be great. Amazing, actually. But we have not focused on that" (SM3a). Rather, sceptical managers often expressed how the scheme could affect teachers by potentially raising their perceived stakes. One principal recounted,

We briefed the teachers on the scheme . . . and [the PES] and its focus on the ninth grade. But that did not mean that [teachers] under any circumstances should feel pressure towards the school getting the bonus. We made it about learning how to improve as a school (SM3a).

SMs believed that how the scheme affected teachers depended on their mediation. Although the bonus did not seem to be in focus, sceptical managers often allocated additional resources to the teaching in ninth grade. Management did so at one school during the first year but not the second because it was not perceived as needed, conversely achieving the bonus in the second year but not the first. Being asked whether similar allocations were to achieve the bonus, another principal argued that if "we see a need, then we do something . . . We do not say: 'This [bonus]. How do we obtain it?'. . . We do not use it as our rationale" (SM3b). Thus, although the actions could be interpreted as responding to the FEAWS, the sceptical managers rationalised them differently.

Sceptics considered the scheme's cancellation inconsequential. One principal reasoned that it acquired "no attention from the administration or the ministry anymore, so we have moved on" (SM3b). They no longer focused on implementing new initiatives, and although teachers kept participating in the PES, it was given less attention. The implemented initiatives, e.g. changes to the reading period before exams, continued because of their perceived beneficial effects. Further, some sceptics hoped that the termination meant that no more outcome-based performance funding would be implemented in education.

4.3 Active users

The actives generally acknowledge that the FEAWS' performance measure could be incomplete (cf. Siverbo *et al.*, 2019) and gamed (Bevan and Hood, 2006). However, they expressed that incompleteness was an exception, e.g. a student group that differed from a more typical year. They further emphasised that means other than gaming, especially targeting resources and improving instruction-level practices, could affect actual performance and targets, making the programme an opportunity to improve and achieve additional funding. Consequently, the actives differed from sceptics in their interpretation of the target difficulty. Sceptics often interpreted targets as too challenging or easy; actives interpreted targets as challenging but achievable. Thus, the funding scheme was perceived to raise the stakes (Au, 2007), and it motivated effort.

The actives did not find the scheme's financial incentivisation problematic. The scheme had issues, e.g. teachers could be critical of it or come to feel pressured by it. However, as one principal explained, "I know that some really dislike scheme. I think that it's interesting to try an [economic] carrot, right? Something out of the ordinary" (SM4a). Further, a vice principal noted that "it's not the money that makes a difference, but the approach where you want schools to do better . . . and then actually ask the question: Are we getting something for those funds?" (SM5b).

A distinct aspect characterising managers and teachers classified as actives was their explicit aim to improve students' performance to achieve the bonus. One vice principal stated that it was a "hunt to achieve the bonus to make life sweeter" (SM5b) for staff and students. Another vice principal explained how initiatives depended on specific assessments to improve grades: "we could see: if we were to achieve it this year, they are weakest in math.

That is a purely strategic prioritisation” (SM1b). Similarly, a third manager summarised how the school had approached the project:

We must openly address that we executed a project aimed at the ninth grade that could have been aimed at the sixth or fourth grade. And that would have had an effect, too. Now we are focused on the ninth because we must succeed this year with this year group. Is the sixth or fourth grade any less important? Such a prize obviously steers your attention (SM2a).

Assessing possibilities for and challenges to achieving the bonus was an ongoing issue. If students did not improve as planned, additional resources would be allocated if available. One principal explained how the incentive scheme was actively used to direct attention and prioritise:

Throughout the year, we have specifically looked at: “What are our chances? How are the percentages looking? . . . Does this specific student have what it takes to get a 4?” And we did that often because we wanted the bonus (SM2b).

Thus, active SMs acted according to the programme’s intention because of the incentive.

The sceptics emphasised that many initiatives that could improve the chances of achieving targets had been initiated before the FEAWS. The actives retrospectively connected similar initiatives to the FEAWS. They explained how pre-existing practices were adjusted and aimed at achieving the bonus: “I do not feel it was: ‘Now we are starting something up! We had already begun . . . [the program] became something extra: ‘What can we do to make them even more ready for the exams?’” (T3). One vice principal argued that the programme “suited both our work to continue improving and to shake up school activities with some money that was low-hanging fruit” (SM5a).

The active managers engaged in dialogue with teachers. Some initially believed that teachers would be sceptical of incentivisation, but gradually, teachers became more positively inclined. Teachers concurred: “first, [the program] primarily concerned management and the teachers directly involved. But as time progressed, others’ interest grew” (T1) owing to specific team meetings that shared knowledge from the PES between staff. The management prioritised the meetings to make the most of the programme and achieve the bonus. During the second year, efforts intensified as new teachers were involved with the following year group and through additionally allocated resources.

Furthermore, teachers used the scheme to argue for continuing or intensifying activities. One school had used the ability grouping of students and flexible organisation of learning groups before the FEAWS. This was believed to improve students’ results but increased cost because it was based on three teachers sharing two classes, thus requiring an additional teacher. One teacher noted that ability grouping “easily becomes controversial . . . it can sound like you split people into A and B groups” (T3) instead of treating students as equals. When the FEAWS was introduced, it was seen as an opportunity to intensify the work with ability groupings:

You can tell [management] that you need something [to accomplish the targets] . . . “Do you need some more weekly Danish lessons? Do you need another teacher for some lessons? Or are there conflicts between students that is too much for a new teacher?” You know . . . something (T8).

Thus, the scheme was “an opportunity to argue for [resources] with management” (T2). Consequently, actions were also initiated for the active teachers because of the FEAWS’ intentions.

Nonetheless, it was emphasised that the bonus was a side benefit. One vice principal explained that “that is not what drives us . . . What’s most important is the academic results” (SM1b). Similarly, others stated that the initiatives positively affected student outcomes irrespective of the bonus. The vice principal above mentioned the initial uncertainty about

“whether [the target] was realistic. But we worked intensively . . . and did something right because they improved quite markedly” (SM1a). Thus, achieving the bonus showed actives that they had “chosen the right path . . . I do not think we have been lucky. It was hard work” (T4). When the schools with active managers achieved the target and obtained the bonus, managers and teachers saw it as a success.

Compared to the sceptics, awareness of the scheme’s cancellation among the actives was higher. However, because the FEAWS had been instrumental in achieving the beneficial outcomes, one vice principal explained how the lack of a bonus would affect them:

We are going to continue the initiatives. . . . But I must admit that I have not looked as much at the specific students . . . The money suddenly disappearing had that effect. But that does not mean that the project will disappear (SM2b).

Following the cancellation of the incentive scheme, the exam-related stakes lowered again. Moreover, some new initiatives that explicitly aimed at obtaining the bonus were concluded. Other initiatives that were less focused on the student-specific outcomes continued. For example, the school that had increased attention regarding ability grouping continued the project owing to its perceived positive effects. The actives who believed they would have attained the bonus were indignant over its cancellation. Although it was acknowledged that it had worked well in this instance, most sceptics did not want more outcome-based performance funding.

5. Intentions and dysfunctionality

The interpretations of the FEAWS by sceptical implementers and active users differed in how dysfunctionality was perceived. Overall, the sceptics maintained that incentivisation was problematic, did not rationalise their actions based on the incentives and did not perceive their actions to be because of the scheme. One sceptical principal elaborated that it was important that “it isn’t the money that comes to control us! We have to focus on ongoing processes and accomplishing better results for all children” (SM7). It was further stated that using the bonus as an argument would entail problematic prioritisations. One sceptical MA explained how,

the downside . . . is that it becomes an entirely financial incentive: you let economic thinking rather than the individual child’s needs take control . . . The upside is that something very complex becomes tangible. But it can become so concrete that you overlook how complicated the world is (MA4).

Thus, focussing on the bonus was perceived to entail behavioural displacement (Siverbo *et al.*, 2019) by misguiding actions. For the sceptics, reacting to the incentive was a sufficient condition for action being dysfunctional because such action would conflict with their educational values and perceived organisational mission. Consequently, the sceptics rationalised their actions as rejecting the FEAWS even if they achieved the intended outcomes of increased student learning.

Conversely, active users emphasised functionality when rationalising their actions. One vice principal reasoned that “the Danish tradition isn’t centred on student results. But if you want improvements? That is what we must concentrate on . . . And the programme helped strengthen that focus” (SM5b). Further, an active principal exemplified how the intentions of the FEAWS was perceived to achieve the FEAWS’ intended outcomes:

The goal is not that our students must get a 4, and that is it. We must form democratic citizens capable of taking responsibility for their own lives . . . But those are not opposites . . . I believe that the student who gets the experience of “Wow! I thought I was bad at reading, but I got a 4!” leaves the school with more confidence. That is not contradictory to our other goals (SM2b).

The actives acknowledged the possible tensions between money and student needs, stressing the importance of handling potential conflicts. Focussing on the bonus could be dysfunctional but was not a sufficient condition. It depended on the beneficial or detrimental outcomes of actions, given how the tension was handled. Thus, the interpretations of how the outcomes aligned with the overall goals of the education were important for determining whether the actions were functional or dysfunctional.

Many actions initiated by sceptical and active managers were similar if implemented with greater intensity by actives. However, the interpretation of the actions' dysfunctionality differed between the two groups depending on their interpretations of the relationship between actions, intentions and outcomes. The sceptics perceived it as functional when additional funds were allocated to the ninth grade if the students' needs required it—but dysfunctional if done to achieve the bonus. For the actives, dysfunctionality was a question of whether the actions achieved beneficial outcomes. Targeting resources and changing instruction-level practice helped achieve targets and increase learning regardless of the reason behind the implementation. However, gaming, e.g. lobbying for transfers of students, would not benefit students but solely earn the bonus. It would not necessarily have worse outcomes but still represented dysfunctional action.

Given these ambiguous interpretations of the initiatives to improve exam outcomes, SMs actively attempted to handle the scheme's trickling down to affect teachers. Generally, managers of both types stated that the bonus should not affect individual teachers. One teacher explained that the scheme could create a high-stakes situation with "an unpleasant pressure on the 9th-grade teachers . . . [because it was] loathsome to be responsible for the school's financial condition" (T5), thus increasing the perceived stakes (cf. [Au, 2007](#)). In such situations, managers had an important role when mediating between organisational priorities and teaching practices: sceptical managers rejecting the scheme and actives aiming for the bonus by initiating improvements.

Management's mediation attempts did not always work as intended. At a school with sceptical managers, a teacher explained how some colleagues at the school still perceived the stakes as being raised:

They still felt the pressure even though we had agreed that there was not! And then it is hard to put your finger on, you know, does it come from yourself? . . . [If your students do not achieve the target] does that mean you are a "bad" teacher? Or that we have not done as much as we should? Or all those other questions that, of course, occur to all teachers (T11).

It could be perceived as "a failure not to achieve the bonus" (T2). To these teachers, it would represent a personal rather than an organisational failure, i.e. a "bad" teacher rather than school, which meant the measure was internalised.

The teachers who felt the pressure did not wish to be interviewed, but we interviewed other teachers at the school. They characterised the aforementioned teachers as what we term active users given the belief that the targets could and should be achieved. Furthermore, many managers and teachers reflected on reactions to such raised the perceived stakes. One teacher explained how

Suppose I had an entire grade and had to boost the academically challenged [to achieve the bonus]. In that case, I'm sure that . . . if I wanted to or not, I would concentrate on them . . . You have to economise your time, and if there is not time for everyone . . . logically, you would focus on those around the 4. That cannot be a surprise to anyone! (T2).

In this situation, where the stakes were perceived to be high and the targets challenging, it was tempting to focus on marginal students. The pressure was also believed to lead to excessive teaching-to-the-test ([Amrein-Beardsley et al., 2010](#)). The teacher added, "if you just looked at it from the perspective of an Excel sheet, a 10? That is good enough . . .

“The students with 10s or 12s, they’ll be fine!” . . . [But] we have an obligation to reach all students” (T2). Efforts had to be prioritised in the classroom. Raised stakes in this way affecting which students were prioritised could be interpreted as either gaming or behavioural displacement, but it would not necessarily influence students’ outcomes. However, prioritising efforts due to exam-related stakes, owing to incentivisation and thereby the intention behind the scheme, would diverge from the organisational mission and norms to help students equally. Thus, responding to the intention could make actions dysfunctional to actives even if neither the action, prioritising it itself nor its outcomes were dysfunctional.

Sceptics and actives agreed that some test-taking practice, self-described teaching-to-the-test, was necessary: “exams have to be something that [the students] feel comfortable with” (T7). However, if students learnt to “structure their responses as well as possible, but the content is rubbish . . . it becomes just cheap points” (T7). Thus, teaching-to-the-test became dysfunctional when it concerned “points” over the content, i.e. measured results over learning.

Individual teachers taking personal responsibility for the performance measure and its consequences were considered dysfunctional by all interviewees that mentioned such effects. Sceptical managers believed that this would depend on managerial mediation. For example, the principal whose teachers internalised targets believed that the teachers would resist the pressure independently, only perceiving increased stakes if their management made achieving the target into a big deal. This belief contrasted with teachers’ experiences from schools with active managers: “money is always important here, and [the scheme]’s been mentioned a lot. There have been clear expectations that we as teachers and school would perform well and deliver results” (T9). This teacher mentioned that the school’s management had clearly prioritised achieving the bonus and often asked for grade projections. However, the teacher clarified that “if we do not get it? It’s a bonus! It is not in the [ordinary] budget” (T9). Similarly,

I think that it could have adverse effects . . . if you budget with [the bonus]. That would make it really important, which would make it influence teaching . . . If management reminded me daily, “remember to improve those results!” . . . I would be stressed about whether my teaching was good enough (T8).

The managers and teachers were used to not having a fixed budget. The ordinary funding scheme implied that students transferring to other schools or referred to special education schools meant a loss of funding. Students transferring in represented a gain. Budgeting with the bonus by spending the funds prior to achieving the target would change its status to a possible loss, raising the stakes. This could affect “your way of thinking so much that you risk missing other important concerns” (SM2a), thereby inducing behavioural displacement. Both actives and sceptics mentioned that such dysfunctional consequences were possible. They diverged on whether the consequences depended on reacting to the intention as a sufficient condition or because of how such a reaction enacted the bonus’ status to affect outcomes.

Active managers believed that teachers could internalise the incentives but emphasised that it was avoidable. One active manager explained that improving results for low-performing students required additional resources, which made it a managerial responsibility. The manager stated that “if organising a response was up to individual teachers? They would not have the time for it” (SM1b). Similarly, a teacher with active management said, “without management support, you cannot do these things” (T7).

Because the sceptical managers did not focus on achieving the bonus, they were less involved in prioritising activities aimed at improving results for the low-performing students. One teacher with sceptical management noted that “I am missing a collegial space to share knowledge [from the supplementary aid] . . . achieving better results is not just up to individual teachers. We discuss it, but that alone does not achieve anything” (T6).

Active management created such a collegial space by explicitly prioritising the scheme. These actions, e.g. mentioning the programme and how the school progressed at internal meeting, and the targeted allocations made the project a collective response. For example, teachers splitting classes through ability grouping meant that they shared responsibility for students. One active principal stated the following:

Numbers can be dangerous, but the fact that these things have been measured means that it can become a tool for teachers to approach management to communicate that they need something extra (SM2b).

The targets allowed for discussions of what was to be achieved, how to achieve it and different roles in the response. Teachers could affect it through their interactions with management on the prioritisations of specific efforts. Instead of being individual or organisational, the financial incentive's level of incentivisation was enacted through interactions between management and teachers.

Some teachers stated that financial circumstances differed by schools. The teachers argued that the ordinary funding scheme could allocate insufficient funds to some schools relative to their financial needs, thereby offering different opportunities. Inequitable funding could "twist it for the schools that can't afford to do [special projects]" (T5). Favero and Rutherford (2020) also emphasised that relatively well-off institutions respond more easily to performance funding to more easily achieve additional funding. However, the teachers were usually uninvolved in budgeting and financial decisions and lacked detailed knowledge regarding how management prioritised funding. The principals acknowledged that funding differed between schools but did not connect the differences to the demands and reactions to the FEAWS. Expressing a general sentiment, one principal explained that "there are many prioritisations when planning your budget, but I can only spend the funds I have" (SM9b). Given different obligatory requirements, such as a ministerial minimum of hours spent on subjects and municipal projects, some "90–95% of [schools'] resources are already budgeted" (MA4). The school management had to prioritise what remained, which made the (dys) functionality of a response relative to its opportunity cost in what was not prioritised.

6. Discussion and conclusion

This paper examines a Danish performance-funding programme that introduced an exam-based performance-funding scheme to select basic-education schools and incentivised enhancing outcomes for low-performing students. Focussing on a large Danish municipality with multiple schools participating in the programme, we studied how and why performance funding was responded to and what influenced the dysfunctionality of actions, outcomes and intentions.

Overall, we identified two diverging interpretations, which we conceptualised as *sceptical implementers* and *active users*. The former doubted the scheme's usefulness but implemented changes, especially owing to the PES' supplementary aid; the latter focused directly on the incentive scheme when prioritising initiatives and allocating resources. Interpretations at the managerial level were consistent over time, but some sceptical teachers gradually became more positively inclined because of positive experiences based on how active managers handled the programme.

We show how the interpretations differed on the scheme's performance measure's perceived incompleteness (Siverbo *et al.*, 2019) and alignment with overall educational goals. Both groups found that the performance measure used in the FEAWS was incomplete. Further, it was ambiguous (Vakkuri and Meklin, 2006) because the interpretations of the incompleteness' implications differed. Sceptical implementers perceived the targets as arbitrary, which made target achievement depend on chance or gaming (Bevan and Hood, 2006). To active users, the performance measure could be circumstantially arbitrary and

affected by gaming; however, the actives mostly perceived the targets as challenging but achievable. The actives' interpretations of the target difficulty affected exam-related stakes (Au, 2007), i.e. they perceived that the scheme increased the stakes, which did not happen to sceptics. Even if most interviewees were critical of performance funding based on student results, the interpretations differed on how responding to financial incentivisation aligned with the organisational mission.

Managements prioritised the programme against other ongoing efforts. The sceptical managers only implemented changes because of the supplementary aid programme, as they perceived the scheme's intention to incentivise as dysfunctional. The active managers explicitly sought to achieve targets and implemented changes to obtain the bonus. Upon the unexpected termination of the programme, the financial incentive and along with it the initiatives to achieve the bonus disappeared, further indicating that the actives' actions responded to the programme's intention, i.e. for schools to be financially incentivised to achieve better results. Although the specific initiatives were often similar between the two groups, the sceptics focused more on all students. In contrast, actives' efforts were of greater intensity and more targeted on the ninth grade and students close to achieving the goal. The sceptics considered it a result of chance when achieving the bonus; the actives saw it as a success achieved by their targeted efforts.

Furthermore, we illustrate how the trickling down of the performance-funding programme depended on interactions between management and teachers. When active managers responded to the programme's intentions, the bonus was enacted as a collective goal: the managers prioritised initiatives to achieve the targets, where individual teachers could interact with management to affect such prioritisations, and the responsibility for achieving the targets was shared between different teachers and management. Generally, the interviewees from both groups indicated that achieving the bonus should not be regarded as a personal responsibility or an indication of the inadequacy of the teachers. However, some active teachers took personal responsibility for the targets, internalising them, even though their managers were sceptics. These teachers did not have the opportunity to affect managerial prioritisations, leaving the target achievement to these individual teachers.

6.1 (Dys)functionality, interpretations and intentions

In addressing the research questions, the paper makes two overall contributions to the literature. First, the paper contributes to the literature on performance measurement and dysfunctionality (Siverbo *et al.*, 2019; Smith, 1993, 1995; Vakkuri and Meklin, 2006) by examining conditions for performance measures to be dysfunctional and demonstrating how (dys)functionality is relative to interpretations and can concern actions, outcomes and intentions. Vakkuri and Meklin (2006) argued that the dysfunctionality of performance measurement concerns whether it helps or hinders achieving the organisational mission. However, Vakkuri and Meklin (2006) emphasised that mission, goals and the means to achieve them complicate dysfunctionality by neither necessarily being agreed upon nor fixed. Further, Siverbo *et al.* (2019) argued that dysfunctional consequences become more likely as control over incomplete measures tightens. Unintended consequences are often confounded with dysfunctionality (e.g. Ortagus *et al.*, 2020; Vakkuri and Meklin, 2006) but can be functional (e.g. Speklé and Verbeeten, 2014; Van Helden *et al.*, 2012). The literature has focused on how actions themselves and their outcomes are dysfunctional. In this paper, we show that actions can also be perceived as dysfunctional because of how and why they respond to intentions. Thus, we contend that actions can be dysfunctional in and of themselves, based on their outcomes and when responding to intentions that are perceived to be dysfunctional.

Our findings demonstrate actions perceived as dysfunctional on their own, e.g. cheating and lobbying for high-performing students to achieve the bonus. However, these actions do

not necessarily worsen outcomes. Our findings also show actions perceived as dysfunctional because of their outcomes. At an organisational level, not prioritising the FEAWS could seem dysfunctional because of its potential benefits. However, its activities had to be weighed against other projects, and dysfunctionality was a matter of interpretation. This conditionality was also apparent at the instructional level where teachers interpreted similar perceived initiatives differently depending on whether the intentions of the programme were considered dysfunctional. The interviewees openly explained actions that could be interpreted as teaching-to-the-test (Amrein-Beardsley *et al.*, 2010), narrowing the curriculum (Au, 2007) and focussing on marginal students (Deming *et al.*, 2016). Such activities are often portrayed as either behavioural displacement or gaming in the literature. However, the interviewees generally perceived the activities as functional because they increased learning to achieve beneficial outcomes. Thus, the dysfunctionality of actions was conditioned on the perception of the outcomes. For example, teaching-to-the-test was functional if used to teach students how to approach exams. It was also emphasised that teaching-to-the-test could become excessive and harm outcomes by narrowing the curriculum. Such dysfunctional reactions become more likely as the stakes are raised, e.g. if teachers took personal responsibility for achieving the targets.

Finally, the findings illustrate how actions can be perceived as dysfunctional when responding to intentions. Although sceptics and actives generally agreed on the two former types of dysfunctionality, they differed in their interpretation of the intentions. Consequently, they differed in how they perceived the (dys)functionality of the action initiated to achieve the bonus and how such actions aligned with the organisational mission to increase student learning. Sceptics emphasised that actions initiated because of the scheme's intention, i.e. in response to incentivisation, contrasted with educational values. Actives maintained that incentivisation aligned with or did not conflict with the educational values. Thus, the actives reconciled the potential tension between acting to achieve the bonus, i.e. reacting to the intention, and increasing learning, i.e. improve outcomes. However, the responses depended on how the incentivisation trickled down to teachers. If the stakes were raised by budgeting with the bonus or if teachers internalised targets, our findings indicate in line with the arguments by Siverbo *et al.* (2019) that the likelihood of dysfunctional response increased.

6.2 Performance funding and its enactment

As its second contribution, this paper adds to the literature on performance funding in education (Bell *et al.*, 2018; Ortagus *et al.*, 2020; Umbricht *et al.*, 2017) by demonstrating how the reception of performance funding depends on the perceived dysfunctionality of the financial incentivisation. Within basic education, Al-Samarrai *et al.* (2018) found that funding measures based on student test scores increased school performance. However, existing research on performance within education often finds that performance funding fails to achieve its intended and even has unintended consequences (Bell *et al.*, 2018; Hillman *et al.*, 2018; Ortagus *et al.*, 2020; Umbricht *et al.*, 2017). Some argue that this is because of design issues, which act as obstacles that can be overcome by more sophisticated models (Dougherty and Reddy, 2013; Jongbloed and Vossensteyn, 2016; Ortagus *et al.*, 2020). Other researchers have argued that quantitative performance measures may be inappropriate (Earl and Hopwood, 1980) and that a clear cause-and-effect relationship may be lacking, thus complicating the tight coupling between mission and objectives (c.f., Hyndman and Eden, 2000) implied by performance funding. Therefore, using funding-based incentives on outcome measures may have inherent problems that are not easily overcome (Bell *et al.*, 2018; Umbricht *et al.*, 2017; Levaić, 2008). Our findings show how design issues and inherent problems were constituted by interpretations based on perceptions of values and organisational mission. This, in turn, has implications for how a performance-funding

scheme is enacted based on educators' interpretations of the (dys)functionality of intentions, actions and outcomes.

The FEAWS's scheme was not technically sophisticated, and the target difficulty had an element of arbitrariness. Neither the criteria for being selected for the programme nor the targets were accurate measures of schools' performance, and the scheme did not include measures to deflect possible unintended consequences. The interviewees acknowledged its incompleteness and noted how the scheme could be gamed, i.e. that the bonus could be obtained without any real improvements in student learning. Further, the programme was highly political, as made visible by its early termination. As [Gándara \(2019\)](#) emphasised, the performance-funding design was both a political and a technical process.

Although numerous interviewees disagreed with the intentions of the FEAWS and most criticised the use of financial incentives in basic education, the scheme still affected initiatives at the schools. Interpretations mediated the specific actions at an organisational and instructional level. To the sceptics, the scheme's incompleteness and incentivisation were perceived as dysfunctional and as inherent issues. Although responding to indirect financial incentives as part of the ordinary funding scheme was everyday practice, responding to a politicised and direct financial incentive was, by the sceptics, perceived to be dysfunctional. It did not align with their perceptions of educational values, i.e. organisational mission, and related action entailed behavioural displacement. The actives emphasised that although the scheme could be gamed, it was not dysfunctional to prioritise initiatives to attain the bonus if the actions increased student learning. However, if the actions harmed educational outcomes, or if the actions were solely meant to affect the likelihood of target achievement rather, the actives also perceived those actions as dysfunctional. Thus, actions with perceived dysfunctional outcomes were rejected across all interviews, while only the sceptics rejected actions solely due to the incentives. Furthermore, gaming only became a self-perceived issue if the performance measure was perceived as the teachers' personal responsibility.

Unintended, dysfunctional consequences depended on interpretations and how reactions to the programme were enacted at specific schools rather than on design obstacles or inherent issues related to its design. If the programme had obstacles, these could be overcome by enacting the scheme as an organisational incentive for actives. If the programme had inherent issues, it was because the incentivisation conflicted with educational values, which was not solvable.

The bonus from the FEAWS and the ordinary educational activities at the schools differed with respect to the timing of the financing: While the ordinary funding was available in the school year where the expenses were borne, the bonus would be paid in the following school year. In principle, the schools could budget with a deficit the first year and realise a surplus when obtaining the bonus in the next year. [Ortagus et al. \(2020\)](#) argued that performance-based funding measures work better as a part of ordinary funding rather than as a potential bonus. Although the effects of different funding schemes are outside of the scope of this paper, it seems likely that when expenses are increased to obtain the bonus, the stakes were raised because not getting the bonus represented a potential loss. In our case, enacting the bonus as a potential loss would require taking on financial risk, which the interviewees were reluctant to do. Therefore, ordinary funding might work better than a potential bonus owing to raising the stakes irrespective of how the funding-based incentive is mediated in its trickling down.

Stakes depend on interpretations. In situations where additional expenses raise the stakes in reality because of the risk of a deficit, sceptical managers, who viewed the funding scheme as dysfunctional, could lower the perceived stakes by enacting the initiatives as unrelated to the budget and the bonus scheme. The reactions from schools with active managers partly depended on target difficulty, which could differ annually because of the incompleteness of the measure. Thus, the annual stakes depend on the target difficulty. However, if the bonus was enacted with the status of a potential loss instead of a potential gain, the stakes would be raised

even further. Although, for example, funding according to graduates would concern a general institutional performance, such as retention, which is affected by many educators, the FEAWS focused solely on graduating classes. It concerned specific students taught by specific teachers. As a bonus, it could more easily be ignored or rejected. Still, given its technical design, it allowed for the possibility of becoming an individual incentive as the measure became internalised by some teachers. Consequently, the effects of the incentive scheme depended on the scheme's technical design but also on the interpretations and managers' mediations.

Generally, how the scheme came to trickle down depended on the supplementary aid programme, the PES. Performance-funding schemes are often stand-alone (Jongbloed and Vossensteyn, 2016) and only potentially aligned with other policies. The PES was generally perceived as positive at the schools and helped define the improvement initiatives. Thus, our findings are in line with Hillman *et al.* (2015), who argued that institution-level capabilities are important for the effect of performance-funding programs.

6.3 Limitations and future research

The present study was conducted in one municipality, thus representing a single case. By keeping contextual parameters such as the ordinary funding scheme, hierarchical relations and ongoing municipal projects constant, we focused specifically on the effect of the performance-funding scheme. Interviews were conducted at all the municipality's schools participating in the programme, and we analysed how perceptions of the programme were mediated by managerial attention and priorities at the school level. Although not statistically generalisable, the case study has allowed for an in-depth investigation of how the programme trickled down to teachers and how tensions between educational aims and financial considerations were coped with.

Following Parker and Northcott's (2016) understanding of theoretical generalisations, we expect that the nuanced understanding of dysfunctionality that we propose will be theoretically generalisable across different settings and that the insights will extend beyond the boundaries of our case study. Our focus was on performance funding, where research has shown that performance funding often fails to achieve its intended results (e.g. Bell *et al.*, 2018) or has dysfunctional consequences (e.g. Ortagus *et al.*, 2020; Favero and Rutherford, 2020; Hasselbladh and Bejerot, 2020). While the policy design, including the funding scheme's technical aspects and sophistication, is important (Jongbloed and Vossensteyn, 2016; Ortagus *et al.*, 2020), further research could examine how dysfunctionality is also related to actors' perceptions of intentions, actions and outcomes. Additional focus should be given to how middle managers' mediation influences whether performance funding achieves its intended results. Such studies should not be limited to educational settings, as the mechanisms are likely to be generalisable to funding schemes more broadly. Much research on performance management and measurement within the public sector has focused on the dysfunctionality of measures and quantifications and resistance towards qualifications. Further research could similarly examine how perceptions of (dys)functionality with respect to intentions, actions and outcomes influence the consequences of performance management more generally.

References

- Agyemang, G. (2010), "Accounting for needs? Formula funding in UK schools sector", *Accounting, Auditing and Accountability Journal*, Vol. 23 No. 1, pp. 111-132, doi: [10.1108/09513571011010619](https://doi.org/10.1108/09513571011010619).
- Al-Samarrai, S., Shrestha, U., Hasan, A., Nakajima, N., Santoso, S. and Wijoyo, W.H.A. (2018), "Introducing a performance-based component into Jakarta's school grants: what do we know about its impact after three years?", *Economics of Education Review*, Vol. 67, pp. 110-136, doi: [10.1016/j.econedurev.2018.10.005](https://doi.org/10.1016/j.econedurev.2018.10.005).

- Alderson, J.C. and Wall, D. (1993), "Does washback exist?", *Applied Linguistics*, Vol. 14 No. 2, pp. 115-129, doi: [10.1093/applin/14.2.115](https://doi.org/10.1093/applin/14.2.115).
- Aliabadi, F.J., Farooq, M.B., Sharma, U. and Mihret, D.G. (2021), "Institutional work and the interplay of stability and change in public budgeting reform: the case of public universities in Iran", *Accounting, Auditing an Accountability Journal*, Vol. 34 No. 4, pp. 786-818, doi: [10.1108/AAAJ-11-2019-4261](https://doi.org/10.1108/AAAJ-11-2019-4261).
- Amrein-Beardsley, A., Berliner, D.C. and Rideau, S. (2010), "Cheating in the first, second, and third degree: educators' responses to high-stakes testing", *Education Policy Analysis Archives*, Vol. 18 No. 14, pp. 1-36, doi: [10.14507/epaa.v18n14.2010](https://doi.org/10.14507/epaa.v18n14.2010).
- Andersen, S.C. and Nielsen, H.S. (2020), "Learning from performance information", *Journal of Public Administration Research and Theory*, Vol. 30 No. 3, pp. 415-431, doi: [10.1093/jopart/muz036](https://doi.org/10.1093/jopart/muz036).
- Arnaboldi, M., Lapsley, I. and Steccolini, I. (2015), "Performance management in the public sector: the ultimate challenge", *Financial Accountability and Management*, Vol. 31 No. 1, pp. 1-22, doi: [10.1111/faam.12049](https://doi.org/10.1111/faam.12049).
- Au, W. (2007), "High-stakes testing and curricular control: a qualitative metasynthesis", *Educational Researcher*, Vol. 36 No. 5, pp. 258-267, doi: [10.3102/0013189x07306523](https://doi.org/10.3102/0013189x07306523).
- Bell, E., Fryar, A.H. and Hillman, N. (2018), "When intuition misfires: a meta-analysis of research on performance-based funding in higher education", in Hazelkorn, E., Coates, H. and McCormick, A.C. (Eds), *Research Handbook on Quality, Performance and Accountability in Higher Education*, Edward Elgar Publishing, Cheltenham, pp. 108-124, doi: [10.4337/9781785369759.00017](https://doi.org/10.4337/9781785369759.00017).
- Beuchert, L.V. and Nandrup, A.B. (2018), "The Danish national tests at a glance", *Danish Journal of Economics, SSRN Electronic Journal* No. 1, pp. 1-37.
- Bevan, G. and Hood, C. (2006), "What's measured is what matters: targets and gaming in the English public health care system", *Public Administration*, Vol. 84 No. 3, pp. 517-538, doi: [10.1111/j.1467-9299.2006.00600.x](https://doi.org/10.1111/j.1467-9299.2006.00600.x).
- Bjerg, J. (1991), "Reflections on Danish comprehensive education, 1903-1990", *European Journal of Education*, Vol. 26 No. 2, pp. 133-141, doi: [10.2307/1502799](https://doi.org/10.2307/1502799).
- Bjørnenak, T. (2000), "Understanding cost differences in the public sector—a cost drivers approach", *Management Accounting Research*, Vol. 11 No. 2, pp. 193-211, doi: [10.1006/mare.2000.0128](https://doi.org/10.1006/mare.2000.0128).
- Brinkmann, S. and Kvale, S. (2015), *InterViews: An Introduction to Qualitative Research Interviewing*, 3rd ed., Sage, Thousand Oaks, CA.
- Broadbent, J. and Laughlin, R. (1998), "Resisting the 'new public management': absorption and absorbing groups in schools and GP practices in the UK", *Accounting, Auditing and Accountability Journal*, Vol. 11 No. 4, pp. 403-435, doi: [10.1108/09513579810231439](https://doi.org/10.1108/09513579810231439).
- Deming, D.J., Cohodes, S., Jennings, J. and Jencks, C. (2016), "School accountability, postsecondary attainment, and earnings", *Review of Economics and Statistics*, Vol. 98 No. 5, pp. 848-862, doi: [10.1162/REST_a_00598](https://doi.org/10.1162/REST_a_00598).
- DME (2007), *Relevant Extract from Ministerial Order No. 262 of 20 March 2006 on the Grading Scale and Other Forms of Assessment*, Danish Ministry of Education, Copenhagen, available at: https://ufm.dk/en/education/the-danish-education-system/grading-system/ministry_of_education_order_262_2007_grading_scale.pdf (accessed 8 June 2021).
- DME (2013), *Agreement on an Improvement of Standards in the Danish Public School (Primary and Lower Secondary Education)*, Danish Ministry of Education, Copenhagen, available at: <https://eng.uvm.dk/-/media/filer/uvm-eng/pdf/13/131007-folkeskolereformaftale-eng-red-2-.pdf> (accessed 22 June 2021).
- DME (2017a), *Model for the School Funding (In Danish: Model for Skolepuljen)*, Danish Ministry of Education, Copenhagen, available at: <https://www.uvm.dk/-/media/filer/uvm/udd/folke/pdf/17/apr/170425-model-for-skolepuljen.pdf> (accessed 8 June 2021).
- DME (2017b), *Program to Enhance Academically Low-Performing Students (In Danish: Program for Loft Af Fagligt Svageste Elever)*, Danish Ministry of Education, Copenhagen, available at:

- <https://www.uvm.dk/-/media/filer/uvm/udd/folke/pdf17/apr/170424-program-for-loeft-af-de-fagligt-svageste-elever.pdf> (accessed 8 June 2021).
- DME (2017c), *Funding to Enhance Academically Low-Performing Students (In Danish: Pulje Til Løft Af Fagligt Svage Elever I Folkeskolen (Skolepuljen))*, Danish Ministry of Education, Copenhagen, available at: <https://www.uvm.dk/puljer-udbud-og-prisuddelinger/puljer/puljeoversigt/tidligere-udmeldte-puljer/grundskole/pulje-til-loeft-af-fagligt-svage-elever-i-folkeskolen-skolepuljen> (accessed 8 June 2021).
- DME (2017d), *Handout: The Challenge (In Danish:Udleveringsnotits: Udfordringsbilledet)*, Danish Ministry of Education, Copenhagen, available at: <https://www.uvm.dk/-/media/filer/uvm/udd/folke/pdf17/juli/170703-udleveringsnotits-udfordringsbilledet.pdf> (accessed 8 June 2021).
- DME (2017e), *500 Million DKK to Enhance Academically Low-Performing Students (In Danish: 500 Mio. Kr Til Løft Af Fagligt Svage Elever)*, Danish Ministry of Education, Copenhagen, available at: <https://www.uvm.dk/aktuelt/nyheder/uvm/2017/april/170425-500-mio-kr-til-loeft-af-fagligt-svage-elever> (accessed 8 June 2021).
- DME (2018a), *Gradings at Folkeskole the exit exams 2017/2018 (in Danish: Karakterer fra folkeskolens afgangseksamen 2017/2018)*, Danish Ministry of Education, Copenhagen, available at: <https://www.uvm.dk/-/media/filer/uvm/stat/pdf18/181025-tabelnotat-karakterer-fra-folkeskolens-afgangseksamen-2017-2018.pdf> (accessed 8 June 2021).
- DME (2018b), *The Aims of the Folkeskole*, Danish Ministry of Education, Copenhagen, available at: <https://eng.uvm.dk/primary-and-lower-secondary-education/the-folkeskole/the-aims-of-the-folkeskole> (accessed 22 June 2021).
- DME (2019), *The Government terminated the school funding program from school-year 2019/2020 (In Danish: Regeringen lukker skolepuljen fra og med skoleåret 2019/2020)*, Danish Ministry of Education, Copenhagen, available at: <https://www.uvm.dk/aktuelt/nyheder/uvm/2019/okt/191002-regeringen-lukker-skolepuljen-fra-og-med-skoleaaret-1920> (accessed 8 June 2021).
- DME (2020), *The Use of the 7-point Grading Scale (In Danish: Anvendelse Af 7-Trins-Skalaen)*, Danish Ministry of Education, Copenhagen, available at: <https://www.uvm.dk/uddannelsessystemet/7-trins-skalaen/anvendelse-af-7-trins-skalaen> (accessed 8 June 2021).
- DME (2021), *Gradings at Folkeskole the Exit Exams 2019/2020 (In Danish: Karakterer for Folkeskolens Afgangseksamen 2019/20)*, Danish Ministry of Education, Copenhagen, available at: <https://www.uvm.dk/-/media/filer/uvm/udd/folke/pdf20/okt/201021-tabelnotat-grundskolekarakterer-2019-20-ua.pdf> (accessed 8 June 2021).
- Dougherty, K.J. and Reddy, V. (2013), "Performance funding in higher education: what are the mechanisms? What are the impacts?", *ASHE Higher Education Report*, Vol. 39 No. 2, pp. 1-152, doi: [10.1002/aehe.20008](https://doi.org/10.1002/aehe.20008).
- Dougherty, K.J., Jones, S.M., Lahr, H., Natow, R.S., Pheatt, L. and Reddy, V. (2016), *Performance Funding for Higher Education*, John Hopkins University, Baltimore.
- Earl, M.J. and Hopwood, A.G. (1980), "From management information to information management", in Lucas, H.C. (Ed.), *The Information Systems Environment*, North-Holland, New York, pp. 3-13.
- Egelund, N. (2005), "Educational assessment in Danish schools", *Assessment in Education*, Vol. 12 No. 2, pp. 203-212, doi: [10.1080/09695940500143886](https://doi.org/10.1080/09695940500143886).
- Eisenhardt, K.M. (1989), "Building theories from case study research", *Academy of Management Review*, Vol. 14 No. 4, pp. 532-550, doi: [10.2307/258557](https://doi.org/10.2307/258557).
- Favero, N. and Rutherford, A. (2020), "Will the tide lift all boats? Examining the equity effects of performance funding policies in US higher education", *Research in Higher Education*, Vol. 61, pp. 1-25, doi: [10.1007/s11162-019-09551-1](https://doi.org/10.1007/s11162-019-09551-1).
- Figlio, D.N. and Ladd, H.F. (2015), "School accountability and student achievement", in Ladd, H.F. and Goertz, M.E. (Eds), *Handbook of Research in Education Finance and Policy*, Routledge, New York, NY, pp. 194-210.

-
- Gándara, D. (2019), "Does evidence matter? An analysis of evidence use in performance-funding policy design", *The Review of Higher Education*, Vol. 42 No. 3, pp. 991-1022, doi: [10.1353/rhe.2019.0027](https://doi.org/10.1353/rhe.2019.0027).
- Gándara, D. and Rutherford, A. (2018), "Mitigating unintended impacts? The effects of premiums for underserved populations in performance-funding policies for higher education", *Research in Higher Education*, Vol. 59 No. 6, pp. 681-703, doi: [10.1007/s11162-017-9483-x](https://doi.org/10.1007/s11162-017-9483-x).
- Gándara, D. and Rutherford, A. (2020), "Completion at the expense of access? The relationship between performance-funding policies and access to public 4-year universities", *Educational Researcher*, Vol. 49 No. 5, pp. 321-334, doi: [10.3102/0013189X20927386](https://doi.org/10.3102/0013189X20927386).
- Goodhart, C.A.E. (1975), "Monetary relationships: a view from Threadneedle street", *Papers in Monetary Economics*, Vol. 1, Reserve Bank of Australia.
- Hagood, L.P. (2019), "The financial benefits and burdens of performance funding in higher education", *Educational Evaluation and Policy Analysis*, Vol. 41 No. 2, pp. 189-213, doi: [10.3102/0162373719837318](https://doi.org/10.3102/0162373719837318).
- Hall, M. and Messner, M. (2018), "The field research method as applied to behavioral accounting research: interviews and observations", in Libby, L. and Thorne, L. (Eds), *Routledge Companion to Behavioral Accounting Research*, Routledge, New York, NY, pp. 225-237.
- Hasselbladh, H. and Bejerot, E. (2020), "Representing and intervening in Swedish education—Mediating and adjudicating by grading numbers", *Financial Accountability and Management*, Vol. 36 No. 1, pp. 20-32, doi: [10.1111/faam.12213](https://doi.org/10.1111/faam.12213).
- Hillman, N.W., Tandberg, D.A. and Fryar, A.H. (2015), "Evaluating the impacts of 'new' performance funding in higher education", *Educational Evaluation and Policy Analysis*, Vol. 37 No. 4, pp. 501-519, doi: [10.3102/0162373714560224](https://doi.org/10.3102/0162373714560224).
- Hillman, N.W., Fryar, A.H. and Crespín-Trujillo, V. (2018), "Evaluating the impact of performance funding in Ohio and Tennessee", *American Educational Research Journal*, Vol. 55 No. 1, pp. 144-170, doi: [10.3102/0002831217732951](https://doi.org/10.3102/0002831217732951).
- Hyndman, N. and Eden, R. (2000), "A study of the co-ordination of mission, objectives and targets in UK executive agencies", *Management Accounting Research*, Vol. 11 No. 2, pp. 175-191, doi: [10.1006/mare.2000.0127](https://doi.org/10.1006/mare.2000.0127).
- Jennings, J.L. and Bearak, J.M. (2014), "'Teaching to the test' in the NCLB era: how test predictability affects our understanding of student performance", *Educational Researcher*, Vol. 43 No. 8, pp. 381-389, doi: [10.3102/0013189X14554449](https://doi.org/10.3102/0013189X14554449).
- Jongbloed, B. and Vossensteyn, H. (2016), "University funding and student funding: international comparisons", *Oxford Review of Economic Policy*, Vol. 32 No. 4, pp. 576-595, doi: [10.1093/oxrep/grw029](https://doi.org/10.1093/oxrep/grw029).
- Kelchen, R. (2018), "Do performance-based funding policies affect underrepresented student enrolment?", *Journal of Higher Education*, Vol. 89 No. 5, pp. 702-727, doi: [10.1080/00221546.2018.1434282](https://doi.org/10.1080/00221546.2018.1434282).
- Lapsley, I. and Skærbæk, P. (2012), "Why the public sector matters", *Financial Accountability and Management*, Vol. 28 No. 4, pp. 355-358, doi: [10.1111/j.1468-0408.2012.00550.x](https://doi.org/10.1111/j.1468-0408.2012.00550.x).
- Levačić, R. (2008), "Funding schools by formula", in Soguel, N.C. and Jaccard, P. (Eds), *Governance and Performance of Education Systems*, Springer, Dordrecht, pp. 205-245, doi: [10.1007/978-1-4020-6446-3_10](https://doi.org/10.1007/978-1-4020-6446-3_10).
- Li, A.Y. and Ortagus, J.C. (2019), "Raising the stakes: impacts of the complete college Tennessee act on underserved student enrollment and sub-baccalaureate credentials", *The Review of Higher Education*, Vol. 43 No. 1, pp. 295-333, doi: [10.1353/rhe.2019.0097](https://doi.org/10.1353/rhe.2019.0097).
- Luxia, Q. (2005), "Stakeholders' conflicting aims undermine the washback function of a high-stakes test", *Language Testing*, Vol. 22 No. 2, pp. 142-173, doi: [10.1191/0265532205lt300oa](https://doi.org/10.1191/0265532205lt300oa).
- Merton, R. (1936), "The unanticipated consequences of purposive social action", *American Sociological Review*, Vol. 1 No. 6, pp. 895-904.

- Meyer, M.W. and Gupta, V. (1994), "The performance paradox", *Research in Organizational Behavior*, Vol. 16, pp. 309-369.
- Miles, M.B., Huberman, A.M. and Saldaña, J. (2014), *Qualitative Data Analysis: A Methods Sourcebook*, 3rd ed., Sage, Thousand Oaks, CA.
- Muller, J. (2018), *The Tyranny of Metrics*, Princeton University Press, Princeton.
- Ortagus, J.C., Kelchen, R., Rosinger, K. and Voorhees, N. (2020), "Performance-based funding in American higher education: a systematic synthesis of intended and unintended consequences", *Educational Evaluation and Policy Analysis*, Vol. 42 No. 4, pp. 520-550, doi: [10.3102/0162373720953128](https://doi.org/10.3102/0162373720953128).
- Ortagus, J., Rosinger, K., Kelchen, R., Voorhees, N. and Chu, G. (2021), "A national analysis of the impact of performance-based funding on completion outcomes among underserved students", Working Paper, InformEd, available at: https://policy-lab.squarespace.com/s/InformEd-States-Working-Paper_PBFSuccess-22hz.pdf.
- Parker, L.D. and Northcott, D. (2016), "Qualitative generalising in accounting research: concepts and strategies", *Accounting, Auditing and Accountability Journal*, Vol. 29 No. 6, pp. 1100-1131, doi: [10.1108/AAAJ-04-2015-2026](https://doi.org/10.1108/AAAJ-04-2015-2026).
- Parsons, T. (1951), *The Social System*, Free Press, New York, NY.
- Pearson, I. (1988), "Tests as levers for change", in Chamberlain, D. and Baumgardner, R.J. (Eds), *ESPM in the Classroom: Practice and Evaluation*, Modern English, London, pp. 98-107.
- Pidd, M. (2005), "Perversity in public service performance measurement", *International Journal of Productivity and Performance Management*, Vol. 54 Nos 5/6, pp. 482-493, doi: [10.1108/17410400510604601](https://doi.org/10.1108/17410400510604601).
- Pischedda, G. and Marinò, L. (2021), "Children of a lesser god? Demand-driven mechanism and the potential rise of unequal competition in IHES", *Higher Education Policy*. doi: [10.1057/s41307-021-00230-4](https://doi.org/10.1057/s41307-021-00230-4).
- Ratner, H. (2020), "Europeanizing the Danish school through national testing: standardized assessment scales and the anticipation of risky populations", *Science, Technology, and Human Values*, Vol. 45 No. 1, pp. 232-234, doi: [10.1177/0162243919835031](https://doi.org/10.1177/0162243919835031).
- Regeringen (2016), *Regeringsgrundlag Marienborgaftalen 2016: for et friere, rigere og mere trygt Danmark*, The Danish Ministry of State, Copenhagen, available at: https://www.regeringen.dk/media/7628/final02_regeringsgrundlag2016_upload.pdf (accessed 8 June 2021).
- Roberts, J. (2018), "Managing *only with* transparency: the strategic functions of ignorance", *Critical Perspectives on Accounting*, Vol. 55, pp. 53-60, doi: [10.1016/j.cpa.2017.12.004](https://doi.org/10.1016/j.cpa.2017.12.004).
- Rosinger, K.O., Ortagus, J., Kelchen, R., Cassell, A. and Brown, L. (2021), "New evidence on the evolution and landscape of performance funding in higher education", Working Paper, InformEd.
- Saldaña, J. (2016), *The Coding Manual for Qualitative Researchers*, Sage, Los Angeles, CA.
- Shirrell, M. (2016), "New principals, accountability, and commitment in low-performing schools", *Journal of Educational Administration*, Vol. 54 No. 5, pp. 558-574, doi: [10.1108/JEA-08-2015-0069](https://doi.org/10.1108/JEA-08-2015-0069).
- Siverbo, S., Cäker, M. and Åkesson, J. (2019), "Conceptualizing dysfunctional consequences of performance measurement in the public sector", *Public Management Review*, Vol. 21 No. 12, pp. 1801-1823, doi: [10.1080/14719037.2019.1577906](https://doi.org/10.1080/14719037.2019.1577906).
- Smith, P. (1993), "Outcome-related performance indicators and organizational control in the public sector", *British Journal of Management*, Vol. 4 No. 3, pp. 135-151, doi: [10.1111/j.1467-8551.1993.tb00054.x](https://doi.org/10.1111/j.1467-8551.1993.tb00054.x).
- Smith, P. (1995), "On the unintended consequences of publishing performance data in the public sector", *International Journal of Public Administration*, Vol. 18 Nos 2-3, pp. 277-310.

-
- Speklé, R.F. and Verbeeten, F.H.M. (2014), "The use of performance measurement systems in the public sector: effects on performance", *Management Accounting Research*, Vol. 25 No. 2, pp. 131-146, doi: [10.1016/j.mar.2013.07.004](https://doi.org/10.1016/j.mar.2013.07.004).
- Steccolini, I., Saliterer, I. and Guthrie, J. (2020), "The role(s) of accounting and performance measurement systems in contemporary public administration", *Public Administration*, Vol. 98 No. 1, pp. 3-13, doi: [10.1111/padm.12642](https://doi.org/10.1111/padm.12642).
- Ter Bogt, H.J., Van Helden, G.J. and Van der Kolk, B. (2015), "Challenging the NPM ideas about performance management: selectivity and differentiation in outcome-oriented performance budgeting", *Financial Accountability and Management*, Vol. 31 No. 3, pp. 287-315, doi: [10.1111/faam.12058](https://doi.org/10.1111/faam.12058).
- Umbricht, M.R., Fernandez, F. and Ortagus, J.C. (2017), "An examination of the (un)intended consequences of performance funding in higher education", *Educational Policy*, Vol. 31 No. 5, pp. 643-673, doi: [10.1177/0895904815614398](https://doi.org/10.1177/0895904815614398).
- Vakkuri, J. and Meklin, P. (2006), "Ambiguity in performance measurement: a theoretical approach to organisational uses of performance measurement", *Financial Accountability and Management*, Vol. 22 No. 3, pp. 235-250, doi: [10.1111/j.0267-4424.2006.00401.x](https://doi.org/10.1111/j.0267-4424.2006.00401.x).
- Van Helden, G.J., Johnsen, Å. and Vakkuri, J. (2012), "The life-cycle approach to performance management: implications for public management and evaluation", *Evaluation*, Vol. 18 No. 2, pp. 159-175, doi: [10.1177/1356389012442978](https://doi.org/10.1177/1356389012442978).
- Van Hengel, H., Budding, T. and Groot, T. (2014), "Loosely coupled results control in Dutch municipalities", *Financial Accountability and Management*, Vol. 30 No. 1, pp. 49-74, doi: [10.1111/faam.12027](https://doi.org/10.1111/faam.12027).
- van Thiel, S. and Leeuw, F.L. (2002), "The performance paradox in the public sector", *Public Performance and Management Review*, Vol. 25 No. 3, pp. 267-281, doi: [10.1080/15309576.2002.11643661](https://doi.org/10.1080/15309576.2002.11643661).
- Wall, D. (1997), "Impact and washback in language testing", in Clapham, C. and Corson, D. (Eds), *Encyclopedia of Language and Education*, Kluwer Academic Publications, Dordrecht, pp. 291-302.

About the authors

Morten Lund Poulsen is a research assistant at the Department of Business Development and Technology at Aarhus University, Denmark. His research focuses on incentives and performance funding in Danish basic education.

Per Nikolaj Bukh is a Professor in Management and Control at Aalborg University, Denmark. He has published articles on management accounting, formula funding and performance funding. Furthermore, he often contributes to postgraduate programmes and conferences. Per Nikolaj Bukh is the corresponding author and can be contacted at: pnb@pnbukh.com

Karina Skovvang Christensen is an Associate Professor at the Department of Economics and Business Economics at Aarhus University, Denmark. She has published articles and books on several subjects, including knowledge management, performance funding and strategy.