AARHUS UNIVERSITY

Coversheet

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

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

How to cite this publication

Please cite the final published version:

Jakobsen, M., Jacobsen, C. B., & Serritzlew, S. (2019). Managing the Behavior of Public Frontline Employees through Change-Oriented Training: Evidence from a Randomized Field Experiment. *Journal of Public Administration Research and Theory*, 29(4), 556-571. <https://doi.org/10.1093/jopart/muy080>

Publication metadata

Title:	Managing the Behavior of Public Frontline Employees through Change-Oriented Training: Evidence from a Randomized Field Experiment
Author(s):	Morten Jakobsen, Christian Bøtcher Jacobsen, Søren Serritzlew
Journal:	<i>Journal of Public Administration Research and Theory</i> , 29(4), 556-571
DOI/Link:	https://doi.org/10.1093/jopart/muy080
Document version:	Accepted manuscript (post-print)

General Rights

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

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

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

Managing the Behavior of Public Frontline Employees through Change-Oriented Training: Evidence from a Randomized Field Experiment

ABSTRACT

Most public frontline employees work out of sight of management. Direct management through authority and incentives is therefore often challenging. We argue that managers can rely on change-oriented training as an effective management tool that can affect employee behavior and work as an alternative or supplement to commands and incentives. First, we outline the concept of change-oriented training and explain how it can direct the behavior of public employees towards organizational goals and political priorities. Second, we test the effect of change-oriented training on employee behavior in a large-scale field experiment of teachers in Danish public schools. We show that training has a substantial effect on teachers' behavior in the classroom and infer from this that training should be seen as a valuable management tool with the ability to influence the behavior of frontline employees.

INTRODUCTION

Many public employees are street-level bureaucrats (Lipsky 2010) whose work takes place close to citizens but out of sight of managers and decision-makers. These employees use their discretion when they make important choices about how to deliver services and to whom. This work context can challenge public managers who are accountable for the provision of services and for being responsive to politicians and service users. Especially in situations when public organizations are instructed to change their service delivery or goals, public managers have to come up with ways to ensure that their employees act according to those demands. Although much research points to the altruistic inclinations of public employees to respond to the public interest (Perry, Hondeghem, and Wise 2010), discretion provides them with opportunity to act in accordance with their own values and preferences, which are not necessarily aligned with those of their superiors (Gailmard 2010). Shaping the behavior of public employees can be attempted through direct management strategies such as authority and incentives, but these are oftentimes relatively ineffective in public organizations, with unobservable processes, low outcome measurability, and motivation crowding (Langbein 2010).

An alternative strategy for affecting employee behavior is employee training. In *Administrative Behavior*, Simon (1976 [1947]: 15) made the observation that employee training can function as an indirect management tool:

Like organizational loyalties and the efficiency criterion, and unlike the other modes of influence we have been discussing, training influences decisions “from the inside out”. That is, training prepares the organization member to reach satisfactory decisions himself, without the need for the constant exercise of authority or advice. In this sense,

training procedures are alternatives to the exercise of authority or advice as means of control over the subordinate's decisions.

Although this observation was made 70 years ago, it has been curiously neglected in the public administration literature. Public administration studies have focused on employee adaptation to organizations within the framework of organizational socialization (Moyson et al. 2018), but although training can be seen as an aspect of organizational socialization, we have very sparse knowledge about the potential for adapting public employees through targeted training processes. Although there is a vast literature on training as a human resource development strategy and the effectiveness of training on employee performance, skills, efficacy, and so on in generic literatures (Garavan 2007; Aguinis and Kraiger 2009) and studies on organizational socialization among public employees (Moyson et al. 2018), we have only sparse knowledge about the usefulness of training as a management tool to change the behavior of public employees.

This study provides two important contributions to the literature on employee training in public organizations. First, we provide a theoretical framework for understanding what we label *change-oriented training* as a management tool that can affect public frontline workers' behavior towards organizational and political priorities. Here, we point to two mechanisms: ability and motivation. First, training can increase the ability of employees and thereby reduce the cost of a specific type of behavior, and second, training can accentuate and clarify the value associated with the given behavior and thereby direct employee motivation and energy towards the trained behavior. Thus, we argue that training can be seen as a management tool that can be used to direct employee behavior.

Second, we test this hypothesis using a randomized field experiment, which is a rare example of an experimental test of training effectiveness. This field experiment was carried out in the Danish public school sector and examined whether inclusion in a training program could change the behavior of teachers to increase the involvement of students in learning activities. 265 school classes were included in the field experiment. These classes were randomly assigned to a treatment or control group. In the treatment group, teachers were enrolled in a training program about increased student involvement in teaching. We use panel data including baseline and post-treatment measures of the students' experiences of the teachers' practices to estimate the effects of the training program on the extent to which teachers involve students. We find that change-oriented training can be an effective tool that, under the right circumstances, can be used by public managers to direct employee behavior towards practices and goals promoted by the training program.

In the following, we first present a theoretical framework in which we combine arguments from public administration with human resource management (HRM) and psychological literatures in order to conceptualize change-oriented training and provide an understanding of how it can affect the behavior of public frontline workers. Then, we discuss the research design, data, and methods before presenting the results from the field experiment. Finally, we discuss the usefulness of employee training as a management tool.

EMPLOYEE DISCRETION IN PUBLIC SERVICE WORK AS AN AGENCY PROBLEM

In his seminal piece on street-level bureaucracy, Michael Lipsky demonstrated how public employees have considerable discretion in determining the nature, amount, and quality of

benefits and sanctions provided by government agencies (2010, 13). The work carried out by employees in core public services such as police, social services, and schools is for the most part complex, and although regulation is a defining feature of public organizations, most work tasks are so complicated that employees must base their judgments on discretionary choices and interpretations of any given problem that they are confronted with (Tummers et al. 2015). Hence, employees act as street-level bureaucrats with considerable impact on the realization of political decisions.

In the language of principal-agency (PA) theory, the behavior of most public employees is unobservable, which creates an agency problem for the principal – i.e. how to motivate the agent to act in accordance with the principal's preferences (Miller 2005). Principals in the public sector – management and, ultimately, politicians – with an interest in specific outcomes over other outcomes are challenged by the employees' ability to diverge from outcomes preferred by the principal (Dixit, 2002). This is particularly problematic in public service delivery, where performance is multi-dimensional (Walker and Andrews 2013) and often characterized by goal conflict or ambiguity (Chun & Rainey 2005).

Thus, an important question in public management is how political principals can ensure that frontline employees contribute to organizationally prioritized performance goals rather than to personally prioritized goals, shirking, or even undermining prioritized goals through sabotage behavior (Brehm and Gates 1999). Classical principal-agency theory recommends the use of direct management interventions such as authority and incentives, but for at least three reasons, these management tools are often ineffective in public service delivery. First, direct interventions target extrinsic motivation but risk crowding out intrinsic motivation and other types of autonomous motivation (Frey 1997; Jacobsen et al. 2014). Especially in public organizations,

many employees are characterized by public service motivation and the willingness to contribute to value creation for society and citizens (Perry, Hondeghem, and Wise 2010). At best direct interventions can prove to be ineffective towards such employees, but at worst they may supplant autonomous motivation and leave public employees demotivated (Weibel, Rost & Osterloh, 2010). Second, this is particularly problematic, since public organizations are characterized by multiple and conflicting goals, not least because of political struggles. Incentivizing some goals can lead to goal displacement and effort substitution towards incentivized and measured goals and, unintentionally, away from less explicit goals (Jacobides and Croson 2001). Finally, since goals and processes in public organizations are often vague and difficult to measure (Wilson 1989), measurement and monitoring are typically associated with considerable costs (Langbein 2010). These three challenges reduce the usefulness of direct attempts to manage the behavior and performance of public employees. We know from extant research that they simply do not work under all conditions (Moynihan 2010). Public managers therefore need alternative means for aligning employee behavior with organizational goals.

CHANGE-ORIENTED TRAINING AS A MANAGEMENT STRATEGY

As mentioned, the idea of using employee training as a mode of organizational influence was already suggested by Herbert Simon in *Administrative Behavior*. Simon also suggested that training can substitute direct management initiatives because training can “minimize, or even dispense with, certain review processes by giving the subordinates training that enables them to perform their work with less supervision” (Simon 1976, 15). Nonetheless, there is still today limited empirical knowledge on the effects of employee training in public organizations.

The organizational socialization literature focuses broadly on how newcomers in particular adapt to organizations (Van Maanen and Schein 1979), and the literature generally supports that employees adapt to factors such as structure, relations, and mentoring (Moyson 2018). However, although employee training is sometimes described as an organizational socialization tactic, few empirical studies have investigated training. However, these studies either investigate training effects on newcomers only (Antonacopoulou and Güttel 2010) or integrate training as a component of a general socialization processes (Oberfield 2014). Thus, we need more knowledge about how employee training affects organizational members broadly, including experienced employees whose norms and values may be more difficult to change. Training is here regarded as a targeted management tool directed at the adjustment of the skills and values of both newcoming and experienced employees to the organizational environment. Thus, employee training is an aspect of organizational socialization, but it relies on a specific management tactic – using training for targeted adaptations. It is important to address the role of employee training, because we lack knowledge about its effects on the behavior of public employees in a specific direction and towards different political priorities.

Employee training has mainly been studied in the HRM and organizational psychology literatures. In both fields, employee training is regarded as a key facet of strategic practices in organizations, which allows organizations to align practices with changing requirements and the need for new employee skills (Garavan 2007; Combs et al. 2006; Huselid 1995). According to the HRM literature, organizations can also adapt to needs for new competences through recruitment, but training allows organizations to maintain valuable employees while providing them with impetus to adapt to changing roles at work (Winterton 2007). Training is specifically defined by the ambition to directly increase the trainees' contributions to an organization,

making it different from general education and employee development, which are broader approaches with less specific ends (Kraiger 2002). In this study, we refer to change-oriented training because we focus specifically on training intended to direct employee behavior towards practices and goals prioritized by management and politicians. Therefore, this focus is different from training intended to optimize the quality of existing practices by developing the employees' skills, knowledge, and performance in relation to those existing practices. In other words, the "change" in change-oriented training signifies its ability to shift employee behavior towards different practices and political priorities. With this focus, change-oriented training is an alternative, or supplement, to commands, monitoring, and economic incentives as tools for securing that public employees follow the priorities of political principals.

The HRM literature on employee training has been preoccupied with understanding effective training design and delivery, but in recent years, attention has shifted to evaluating training effects (Tharenou 2009), which is also our focus here. An important question is how training can be evaluated and, particularly, which criteria are most relevant to assess. According to a classical HRM model, training effects can be regarded as a chain of effects from employee reactions, via learning and behavior to organizational results (Kirkpatrick 1979). Thus, training effects should ultimately be measured on performance, which, in private businesses, means economic terms such as profits. However, training effects on performance are often difficult to identify given the long causal chain, and assessments of performance are particularly challenging in public organizations. As noted earlier, performance in public organizations is not only difficult to measure; it is typically multi-dimensional and also related to processes. Returning to the Kirkpatrick model, training effects should therefore also be assessed by actual work behavior

following training. Such training effects are important because they can serve as antecedents of job performance more generally (Kraiger 2002).

Empirical studies generally support the notion that training is important for job-related behavior and performance, although effect sizes vary substantially across studies from highly positive to even negative findings (Arthur Jr. et al. 2003; Aguinis and Kraiger 2009). However, most studies have focused on the organizational level and general training policies, used self-reported training activities, or studied mechanisms for transferring training to organizational outcomes (Grossman and Salas 2011), while there is a lack of studies of how specific employee training programs are actually converted to valued individual behavior in organizations (Tharenou 2009). Furthermore, most studies of training impact are cross-sectional and therefore not necessarily reporting causal effects of training. The few existing experimental studies show positive results on skill adaptation and behavior (Davis and Yi 2004), but most of these studies have been carried out in relatively technical and simplistic production settings, which can have little resemblance with public service provision, where employees work in professionalized communities and with substantial discretion. Finally, the educational training literature has investigated how training affects teacher behavior (see Desimone 2009 for an overview), and these studies generally support that professional development of teachers can affect teaching and also offer some best practice guidelines as to how teaching can be designed (Kennedy, 2016). A recent study also uses experimental evidence to show that training can improve very specific teacher behavior such as the opportunity to learn mathematics (e.g. Urick, Wilson, Ford, Frick, and Wronowski 2018), and although such studies are encouraging for the use of training for adapting employee behavior, we still need to understand how specific change-oriented training programs can lead to changed behavior.

WHY SHOULD CHANGE-ORIENTED TRAINING AFFECT THE BEHAVIOR OF PUBLIC EMPLOYEES?

Simon (1976, 16) sees training as a key mode of influence because it can be used to affect employee behavior through two mechanisms: “Training may supply the trainee with the facts necessary in dealing with these decisions; it may provide him a frame of reference for his thinking; it may teach him ‘approved’ solutions; or it may indoctrinate him with the values in terms of which his decisions are to be made”. The first mechanism, the supplying of facts, has to do with the knowledge and competences of the employee. The other mechanism, frames of references, approved solutions, and “indoctrination”, has to do with the values and motivation of the employees. This mechanism is closely related to the adaptation processes described as organizational socialization, and it has been taken up in HRM research, where the AMO (abilities, motivation, and opportunities) framework argues that training mainly works through abilities and motivation (Appelbaum et al. 2000). Particularly in a school setting, Desimone (2009) has presented a model suggesting that training affects the knowledge/skills and attitudes/beliefs of trained professionals.

The first mechanism, supplying facts or providing abilities, involves the obtainment of new knowledge and skills through training, and according to much of the HRM literature, this is the main mechanism for training (Jiang et al. 2012). More specifically, this may involve both cognitive learning, which is the acquisition and understanding of new concepts, and behavioral learning, which is related to the physical ability to act through the obtainment of skills (Winterton 2007). According to learning theory, effective training involves not only the transfer of knowledge but also opportunities to reflect on the concrete relevance of those practices and

practicing actions (Kolb 2014; Kennedy 2016). For example, in relation to student involvement, teachers can be taught what involvement is, about different forms of involvement, or about situational cues for when and how to involve students, but we expect training to be most effective when teachers are also given the opportunity to practice involvement and receive feedback on practiced behavior. A review of teacher training programs found that more effective programs had features such as follow-up steps in teachers' schools, active learning methods, and alignment with broader school policies (Blank and de las Alas 2009). In a principal-agency framework, we can expect that the provision of knowledge and skills allows employees to exert new types of behavior or reduces the costs (in terms of time and effort) of an already exerted type of behavior. Thus, from an employee perspective, training increases the likelihood that an individual will engage in behaviors involving those skills because they are more easily available. If a training program, for example, provides a teacher with skills and knowledge on how to involve students in class, the teacher is more likely to actually involve students.

A second mechanism of training addresses the employees' frames of reference, which are tied to their values, attitudes, and motivation. Whereas direct interventions risk crowding out motivation, training is as an indirect management intervention expected to strengthen motivation. The AMO framework suggests that motivation can be a mechanism for training, and one way to think about this mechanism is that effective training entails communication about the value of the trained skills and knowledge, which might spur willingness to devote energy to reach that particular end. Following Simon's description, this may take the form of indoctrination, but on a more positive note, it may simply be the accentuation of the importance and meaning of a given type of behavior. Thus, training has been argued to clarify the importance of the trained behavior for organization, society and citizens (Parlberg & Perry 2007), and this may both increase the

employees' motivation and align their values more with those of their principals. First, motivation can be affected because training can provide trained individuals with clear goals, which are a well-known antecedent of work motivation (Wright 2004). For example Frayne and Geringer (2000) showed that a training program of general managers resulted in increased self-efficacy and outcome expectancy. Second, training can incur shifts in public values among employees, and this can provide better alignment between individual and organization (Paarlberg & Perry 2007). In public organizations, public service motivation is an important force that ensures employee effort despite discretion, but for managers it also entails particular agency problems, because public service motivation is tied to particular value premises, which guide employee effort (Gailmard 2010). For example, training can increase the value teachers place on involving students in the class-room and thereby ensure that already public service motivated teachers direct their efforts towards increased student involvement. Based on the expectation that employee training can affect the employees' competences and values in relation to a specific type of behavior, we therefore expect that employee participation in a change-oriented training program pushes the employees' work behavior towards the practices and goals promoted by the program.

DESIGN: RANDOMIZED FIELD EXPERIMENT

Studying the effect of management initiatives aimed at changing the behavior of employees is notoriously difficult since such initiatives are rarely randomly assigned but rather chosen with a clear purpose. This implies that observational studies of effects of management strategies are beset by endogeneity problems. Management tools such as assigning employees to training programs are used in certain situations, potentially in response to concrete problems. A

management decision to offer training to specific employees might be a function of the manager's perception of the need for better qualifications, of a management desire to affect the behavior of these employees, or wanting to reward specific employees, perhaps for good performance. All these examples entail that the independent variable of interest, a management decision to include employees in a training program, is affected by factors closely related to the dependent variable of interest in this study, employee behavior. In other words, any correlation between training and behavior may be due to an effect of training on behavior (which is what we are interested in) or by an effect of employee behavior on the likelihood of being included in a training program (which is entirely irrelevant to our research question). It is impossible to distinguish between these two effects in statistical analyses of observational data. Hence, it is difficult to establish the causal effect of inclusion in training from the correlation between inclusion and behavior. We solve this problem by randomizing the inclusion in training programs in a large-scale field experiment (Blom-Hansen, Morton and Serritzlew 2015; Angrist and Pischke 2009, 15-22).

The experiment, which we conducted in cooperation with the Danish Ministry of Education¹ and the Association of Danish Pupils (NGO), focuses on student-centered instruction in Danish primary schools. In 2013, the Danish parliament passed a law to reform the primary schools. According to the Minister of Education, an important element of the reform was that schools should focus more on student-centered instruction in their daily teaching (Antorini, 2014). Student-centered instruction includes teaching practices that emphasize student responsibility and activity rather than content or what teachers are doing (Cannon and Newble

¹ A report with the empirical results of the experiment has been delivered to the Ministry of Education.

2000, 16). Specifically, our case focuses on getting teachers to involve students in the planning, execution, and evaluation of the teaching of the class, such as involving students in deciding which books to read, which topics a given course should cover, and the procedure for how to cover a specific topic in class. This entails a change in the daily work procedures and decisions made by the teachers who do not already practice this approach to teaching. In Denmark, 98 municipalities administer public primary schools in accordance with the laws issued by the national parliament and the guidelines issued by the Ministry of Education. Thus, the Ministry of Education and local municipalities are responsible for implementing the reform.

In the experiment, we examine the effect of a training program financed by the Ministry of Education on the degree to which eighth grade Danish and Math teachers use a student-centered approach in their class teaching. The teachers assigned to the training program followed a workshop course on student-centered instruction. The training program began early September 2014 and ended in the beginning of March 2015. By focusing on training of teachers in public schools, we are able to study the effect of change-oriented training on a large group of employees with wide discretion (classroom teaching is not routinely observed by school leaders). It also enables us to study a case where the change in behavior can be observed (we use the experiences of students to measure the teaching practice). If the change-oriented training has an effect on behavior, we should be able to detect it. However, it is important to stress that the results cannot be readily generalized to other types of tasks, for example when discretion is limited. We return to this in the discussion.

We estimate the effects of the training program by comparing the extent of student-centered instruction in classes assigned to the training treatment with the student-centered instruction in classes from two types of control groups in which the teachers were not enrolled in

the training programs. For both the treatment and control groups, we use pre-intervention measures of teachers' behavior to account for the baseline level of student-centered instruction, and post-intervention measures to capture the outcome level. First, the training treatment is compared to a control group consisting of parallel classes of the classes included in training treatment. We term this the parallel control group. Since the parallel classes are located at the same school and grade level as the treatment classes, they consist of the same types of students, teachers, administrative procedures, and so on, which make them ideal for comparisons. However, location at the same school also entails a risk of experimental contamination, since the teachers from the control classes might apply student-centered techniques after talking to teachers from the training treatment (see Gerber and Green 2012, 253-83). Therefore, we added another type of control group. This control group is "clean" in the sense that the classes were drawn from schools where no classes were included in the training programs. We call this the clean control group. The treatment group is compared to each of the two types of control groups.

Furthermore, two versions of the treatment is tested; one with lower intensity, which we denote the "basic treatment", and one with higher intensity, which we denote the "extended treatment" (the contents of the two treatment versions are detailed below). Since two versions of the treatment are applied, the classes in the experiment were assigned to one of the following five experimental groups: 1. the basic treatment, 2. the extended treatment, 3. the parallel control group for the basic treatment, 4. the parallel control group for the extended treatment, or 5. the clean control group.

Assignment to Experimental Groups

Since teaching takes place in classes, we assigned school classes instead of individual students to the treatment and control groups (i.e., a cluster-level random assignment, see Gerber and Green 2012, 80-85). The experiment included school classes from forty-eight municipalities from which we obtained permission to include any of their public schools in the experiment. Thus, we were able to investigate the effects of implementing the training programs as a normal management decision, where the public managers of a municipality decide that their schools should follow the training program.

We needed classes from three different schools in order to assign one class to each of the five experimental groups: one school for each version of the training treatment and their parallel control classes, and one school for the clean control group. Therefore, we first sampled 60 strata that each contained three schools. In order to generate homogenous experimental groups, each stratum contained schools from the same municipality and with similar student performance (measured by student grades). The number of strata sampled from each municipality depended on the municipality size (more strata from larger municipalities). In order to simulate a management decision that did not only include schools volunteering to participate, we used a random draw to decide which strata of schools to include from each of the municipalities.

Within each stratum, we randomly drew a school to provide classes for the basic treatment and the parallel control to the basic treatment, a school to provide classes for the extended treatment and the parallel control to the extended treatment, and a school to provide the clean control class. Then, random assignment was also used within the schools to decide which classes should be in the treatment group, the parallel control group, and the clean control group. However, our design requires that the schools included in the experiment have at least two eighth grade classes in the beginning of the 2014/15 school year. Therefore, we excluded from the

experiment all schools that only registered one eighth-grade class in August 2014. In total, that produced a sample of 265 school classes distributed across the experimental groups, as shown in Figure 1.

[Figure 1 about here]

Some of the classes that were assigned to the treatment did not actually receive the treatment. Seven out of 106 classes in the treatment group (or, 6.6 pct. of the treatment group classes) declined to participate in the training program for various reasons. Moreover, one class in the control group received the treatment even though it was not supposed to. As explained further in the presentation of our estimation strategy, we apply the intention-to-treat principle in the paper's main analysis, and provide analyses of the local average treatment effect (Imbens, Angrist, and Rubin 1996) in the supplementary material.

Treatment: A Change-Oriented Training Program on Student-Centered Instruction

The training treatment began early September 2014 and ended at the beginning of March 2015. To make the training program effective, the Ministry of Education collaborated with consultants knowledgeable about student-centered instruction employed by the Association of Danish Pupils. The consultants designed and implemented the training treatment, which was fully financed by the Ministry of Education. The training program itself was developed based on the pedagogical principals of action learning (see Beyer et al. 2004), which draw on Kurt Lewin's (1949) action

research and Chris Argyris' (1995) action science.² As detailed below, the training treatment is expected to impact the teachers' instruction practice through the two mechanisms described in the theory section (see *Why should change-oriented training affect the behavior of public employees?*). As mentioned, two versions of the training treatment were implemented: A "basic treatment" and an "extended treatment". We start by describing the basic treatment since the extended treatment consists of the basic treatment plus additional activities.

The basic treatment is a workshop course that includes three workshop sessions for the teachers, class exercises between the sessions, and a workshop session for the school principals. The workshop sessions for the teachers focused on different techniques and exercises related to student-centered instruction that teachers can use in their daily teaching. Hence, the training treatment aimed to provide the teachers with facts, knowledge, and skills that lower the costs of complying with the new political goal of increased student-centered instruction (c.f., the first mechanism of change-oriented training outlined in the theory section). The materials used in the training treatment include booklets, videos, board games, and LEGO bricks among other things.

The first round of workshop sessions was held in September 2014, the second sessions in November 2014, and the third sessions at the end of February and early March. The Danish and Math teachers from the treatment classes participated in the workshops together with two students from each treatment class. The workshop sessions were held simultaneously at different locations across the country. Teachers and students from about 10-15 treatment classes participated at each location. Two students from each teacher's class participated in the workshops in order to give them an idea of what student-centered instruction is and what their

² The pedagogical foundation and references used to develop the contents of the training program (including the specific techniques for student-centered instruction) are described in more detail by the Association of Danish Pupils in Andersen, Jeppesen, and Gloy (2014, 110-119).

teachers were introduced to. For each workshop session, the consultants were accompanied by young volunteers (10th grade students affiliated with the Association of Danish Pupils) who served as teaching assistants.

The first of the three workshop sessions for the teachers introduced them to the aim and purpose of the training program. Convincing the teachers of the idea and relevance of the program and student-centered instruction was important since the teachers were instructed by management to participate. This speaks directly to the second mechanism through which we expect the change-oriented training to impact employees; the employees' frames of reference tied to their values, attitudes, and motivation. The consultants specifically prepared for this task and devoted time to clarify the importance of student-centered instruction and explained how and why it can contribute to teaching and students' skill formation. In addition to this, the very enrollment of the teachers in the workshops sends a strong signal about the importance of student-centered instruction.

Prior to the first workshop session, the teachers were asked to do a reflection exercise on their own and with their class about their perception of student-centered instruction. This exercise was repeated in the first workshop session. In the first session, the teachers were introduced to ten specific techniques that relate to either the planning, execution, or evaluation of the class teaching. For example, one of the planning techniques focuses on involving the class in designing the procedure for peer-feedback in class. The teacher asks the students to write down ideas for "codes of conduct" when providing peer-feedback. Subsequently, the ideas are spread out in the classroom, and the students are divided into teams that collect what they believe are the most relevant "codes of conduct" for peer-feedback. Finally, a presentation and discussion of the various ideas result in a shared outline of "codes of conduct" for peer-feedback. After the

first workshop, the school principals of the treatment schools participated in a separate workshop introducing them to student-centered instruction. The purposes of the principals' workshop was to give them an idea of what their teachers were introduced to as well as securing their support for the training programs.

In the second workshop session, the teachers reflected on employing the techniques from the first workshop and were introduced to additional techniques as well. In the third workshop session, the teachers were given advice on how to improve on the techniques based on their experiences following the first workshop and were also introduced to various board games that can be utilized for student-centered instruction. Thus, the basic treatment consists of the workshop sessions and exercises between the sessions described above.

The extended treatment consists of the basic treatment plus an extra workshop session at the teachers' school, which made it possible to do in-class exercises. The teachers did not learn new techniques in the extra session, but fine-tuned their application of the techniques learned in sessions 1-3 with in-class exercises and feedback from the consultants.

Measures

We measure the work practice in the classrooms by surveying the students and using them as informants. One major advantage is that since a class contains a substantial number of students, we are able to obtain several measure points of the teaching practice in each classroom.

Furthermore, we expect that to a lesser degree than the teachers, students are influenced by social desirability and are less strategic when answering, since they are not directly accountable for the teacher's applied practice. We are not able to completely rule out that students, to some extent, indicate that the level of student-instructed practice is higher than it actually is. This problem has

two potential sources. The first source is a social desirability effect (student may perceive student involvement as socially desirable and hence inflate the level of student involvement). Such an effect would be similar in treatment and control groups, and is accounted for by use of a first-difference estimation strategy as detailed below. The second source is a contamination effect. Students in the treatment group may be aware that the teacher participates in training with the aim of increasing student involvement. This may increase the likelihood that the students rate student involvement as higher. We address this problem in three ways. First, we perform robustness analyses without students that participated in the training courses. We discuss the results in the section on estimation strategy. Second, the survey questions were formulated to measure teacher behavior (we discuss the survey questions further below). Third, a comparison of the results in the two subjects, Danish and Math, indicate that a contamination effect is unlikely to explain the results. We discuss this in the results section.

The extent of student-centered instruction was measured by surveying the students of the treatment and control groups at baseline (end of August 2014) and again in May 2015 (about 2-3 months after the last workshop session).³ Hence, we study the effects after the treatment has been completed. Studies of training designed to reduce bias in hiring decisions have indicated that effects can sometimes be very short-lived and may sometimes only last for days (Dobbin & Kalev (2016: 4). The 2-3 months period between the completion of the treatment and the post treatment survey ensures that we measure effects that have been lasting for some time. The timing of the post survey treatment does not allow us to study long-term effects, but we offer a discussion of the long-term potential of change-oriented training in the discussion section. We

³ The surveys were administered by the Association of Danish Pupils and the researchers using an online survey program.

measured the overall extent of student-centered instruction by a scale formed by two subscales measuring student-centered instruction in Danish and Math classes respectively. The Danish and math subscales each consists of six items, and they were developed by a quantitative pilot study and qualitative evaluation of how the students perceive and understand the items.⁴ The items are presented in the Supplementary Material. Factor analyses support uni-dimensionality for the Danish and the Math subscales respectively. Cronbach's Alpha for the Danish subscale is 0.85 (baseline) and 0.89 (post-intervention), and it is 0.89 (baseline) and 0.91 (post-intervention) for the math subscale. To avoid dropping too many respondents because of missing answers, we allow for missing values on two of the six items in each scale (in case of missing values on one or two items, the respondent's score on the subscale was calculated based on the remaining four or five items). Respondents with more than two missing answers on one of the subscales were dropped from the analysis. Thus, our main analysis includes students ($n=2,680$) who provided information for both the Danish and Math subscales at baseline as well as at the post-intervention measures. Each of the two subscales is formed by the six items as an additive index and rescaled to range from 0 (lowest student-centered instruction) to 10 (highest student-centered instruction). The two subscales are then added to construct an overall scale of student-centered instruction. This is also rescaled to range from 0 to 10. For example, if student i indicates 7 on the Danish subscale and 8 on the Math subscale, student i 's overall measure of student-centered instruction

⁴ In 2012, six items constructed to measure student-centered instruction were included in a survey of 3,475 students conducted by the Association of Danish Pupils (2012). The items were then adjusted based on the experiences, feedback, and statistical tests from this study (among other things, factor analysis was used to test whether the items measure one latent dimension). Then the wording of the items was subjected to a more qualitative evaluation in the spring of 2014. A class of eighth-grade students completed the survey, which was followed by a session where the students explained to the researchers how they interpreted the item wording. The qualitative evaluation lead to minor revisions of the wording.

is 7.5. Our analysis examines three dependent variables: 1) the overall measure of student-centered instruction, 2) the Danish subscale of student-centered instruction, and 3) the math subscale of student-centered instruction.

Survey Response and Balance Checks

As is often the case, it was easier to convince classes in the treatment group to participate in the surveys compared to classes from the two control groups (since the control group did not receive a placebo, teachers and students knew whether they were assigned to the treatment or control group). Thus, 87.7 percent of the classes in the training treatment are represented by student responses on all four measure points (Danish and Math at baseline and post-measures) compared to 76.4 percent in the parallel control group and 75.5 percent in the clean control group.⁵ Another potential source of bias is that not every student in each class provided information on all four measure points. 50.5 percent of the individual students from the treatment group provided information on all four measure points compared to 39.9 percent and 40.9 percent in the controls groups. Therefore, in Table 1 we provide a comprehensive check of whether the experimental groups are similar at baseline (i.e., a balance check) when we observe the students who are included in the analysis (i.e., the students who answered the surveys). This allows us to examine whether non-response distorted the balance of the experimental groups. The balance check uses detailed register data on students and their parents as well as the baseline survey of the students.

⁵ Table S6 in the Supplementary Material provides an analysis that only includes the post-intervention measure as dependent variable. Thus, when we give up the panel data structure, we gain higher response rates. This analysis includes 92.5 pct. of the treatment classes, 85.8 pct. of the parallel control classes, and 88.7 pct. of the in the clean control classes.

The register data are obtained through Statistics Denmark, which is a rich source of individual-level data on socioeconomic indicators.

Table 1 shows the student characteristics across the experimental groups. The characteristics are summarized by means or percentages (the latter is indicated by %). Stars indicate significant differences between the training treatment and the control groups. Since the balance check contains 55 comparisons, we would expect that some of them are statistically significant by chance. In general, the experimental groups are fairly balanced on the indicators presented in Table 1. Notably, there are no significant differences with regard to student-centered practices (see the top three rows of Table 1) or the students' test results at baseline. The students in the training treatment display a slightly better level of well-being but have lower income compared to the parallel control group. However, these differences are not statistically significant when we compare the treatment group with the clean control group. Furthermore, there are no statistically significant differences between the experimental groups when we examine the time spent on homework. If we observe the students' family background, 22 percent of the treatment students live in single-parent households compared to 18 percent in the parallel control group. Apart from that, we find statistically significant differences for father's education and material status, but these differences are also substantially very small.

In the Supplementary Material (Table S2) we present a comparison of the respondents and non-respondents within each experimental group. On average, non-response was more frequent among students with lower test score and low socio-economic status. Since this response pattern is observed in all of the experimental groups it should not disturb the balance of the groups (which is also indicated by Table 1).

[Table 1 about here]

Estimation Strategy

Our analysis consists of two parts. In the first part, we estimate the overall effect of the treatment on student-centered instruction. Here, we use the overall measure of student-centered instruction as dependent variable. Moreover, when estimating the overall treatment effect, we do not distinguish between which type of treatment (basic or extended) the class was allocated to. Here, the treatment variable indicates whether a student was in a class assigned to a treatment (=1) or not (=0). That is, we estimate the pooled effect of the two types of treatment. This requires two regressions; one where the training treatment is compared to the parallel control group and one where it is compared to the clean control group.

In the second part of the analysis, we examine each of the subscales of student-centered instruction (Danish and Math) as dependent variables, and we distinguish between whether the class was assigned to the basic or extended training treatment. This enables us to see whether any overall treatment effect is produced mainly by one of the two versions of the treatment and/or whether any treatment effect is located mainly in Danish or Math class.

The effects of the training treatments are estimated using first difference models that exploit the panel structure of our data. The model follows Wooldridge's (2013, 465-68) specification of the difference model to estimate policy effects in the case of two measure points where none of the observations was treated in the first wave (baseline). Since we have two waves in our dataset (baseline and the post-intervention measure, $t=1, 2$), the first difference approach is identical to estimation using fixed effects models as well as dummy variable regression where an intercept for each student is estimated using dummy variables (Wooldridge 2013, 490). In formula, our model used to estimate the overall treatment effect can be written as:

$$\Delta Y_i = \delta_0 + \beta_1(TREATMENT_i) + \beta_2(CLEAN_i) + \Delta u_i,$$

where ΔY_i is the change from baseline to the post-treatment in the level of student-centered instruction reported by student i . Since we have three experimental groups, we need two dummy variables to capture the treatment status of the students. Hence, the dummy variables $TREATMENT_i$ and $CLEAN_i$ indicate whether student i belongs to the treatment group, the clean control group, or the parallel control group (if both dummy variables equal 0). δ_0 is a constant that captures the change in student-centered instruction in the parallel control group, and Δu_i is change in the error term for student i . Since treatment was randomly assigned we can assume that Δu_i is uncorrelated with treatment status. The model presented above estimates the effect of the treatment compared to the parallel control group. In the analysis, we also estimate a second model that captures the treatment effect when compared to the clean control group. In this model we replace $CLEAN_i$ with a dummy variable indicating assignment to the parallel control group ($\Delta Y_i = \delta_0 + \beta_1(TREATMENT_i) + \beta_2(PARALLEL_i) + \Delta u_i$).

When, in the second part of the analysis, we analyze the subscales of student-centered instruction and distinguish between the different versions of the treatment, we extend the models accordingly. For example, this model explains the change in student-centered instruction in Danish class when we distinguish between the two treatment versions and use the parallel control for the basic treatment as reference category:

$$\begin{aligned} \Delta Y_i^{Danish} = & \delta_0 + \beta_1(BASIC_TREATMENT_i) + \beta_2(EXTENDED_TREATMENT_i) \\ & + \beta_3(EXTENDED_PARALLEL_i) + \beta_4(CLEAN_i) + \Delta u_i, \end{aligned}$$

All models are estimated using cluster robust standard errors at the class level in order to account for the hierarchical data structure, students within classrooms, when calculating standard errors (we obtain similar results when using hierarchical linear models, see below and the Table S3 in the Supplementary Material).

The analysis applies, as far as possible, the intention-to-treat principle, meaning that the students and classes are analyzed as belonging to the experimental group to which they were assigned at baseline. Of course, it is not possible to include students/classes that did not respond to the surveys. Of the 93 treatment classes that responded to the surveys (and therefore are included in the analysis), five declined to participate in the treatment program and one class assigned to the control group received the treatment. These are analyzed according to the group to which they were assigned at baseline. In addition to the intention-to-treat analysis, we also provide an analysis of the local average treatment effect (Imbens, Angrist, and Rubin 1996), which is presented in the supplementary material.

In addition to the estimation presented above, we made a series of tests in which the effects are estimated using different estimation procedures and models. These are presented in the Supplementary Material. Following Raudenbush and Bryk (2002), we provide estimations using hierarchical linear models, which are commonly applied within the education literature (e.g., Jacob et al. 2015). Using different specifications of hierarchical linear models, we find that the estimated effects are similar to those obtained by the first difference approach (see Table S4). We also find a significant effect in a model that excludes the students who participated in the workshop sessions along with the teachers (see Table S5). This is an indication that the results are not the product of a contamination effect. Finally, we find similar effects using a model that

does not include the baseline measures in order to include more students in the analysis (see Table S6).

RESULTS

The first part of the analysis consists of the overall effect of the training treatment on student-centered instruction in the classrooms. Table 2 presents the effect of being assigned to a training treatment on the overall measure of student-centered instruction. Here, there are three experimental groups: the training treatment (the basic and extended versions of the treatment are combined into one treatment group), the parallel control group, and the clean control group. Model I uses the parallel control groups as reference category. In this model, the overall effect of the training treatment is 0.39, which is statistically significant at the 0.01-level. The treatment effect is estimated to be 0.51 when the clean control group is used as reference category in Model II ($p < 0.01$). Furthermore, we see that there is no significant difference between the two control groups, which indicates that there is no serious contamination effect of the treatment classes on the parallel control classes.

[Table 2 here]

Hence, using two types of control groups, which are not related to each other, we find a positive effect, which provides very strong evidence of an overall effect of the training treatment, as measured three months after the treatment. This shows that change-oriented training has effects in the short term. The estimated effect is more or less the same size, regardless of the type of control group applied. The effect size of 0.39 corresponds to 0.24 standard deviations on the baseline measure of overall student-centered instruction, and the effect size of 0.51 corresponds

to 0.31 standard deviations. We can compare this to the effect size of another good predictor of student-centered instruction, which is the subject (Danish versus Math). The baseline difference between student-centered instruction in Danish and Math classes is 0.36 (across all the experimental groups) in favor of Danish. Thus, the effect of the training treatment is of about the same size as the difference between Danish and Math.

Estimations of the local average treatment effect (see Table S3 in the Supplementary Material) show marginally larger effect sizes. Here, the effect sizes are 0.41 and 0.53 when the training is compared to the parallel control group and the clean control group, respectively.

Moving on to the second part of the analysis, we break down the analysis by subject and version of the training treatment in Tables 3 and 4. Table 3 presents the effects of each version of the treatment (basic or extended) on student-centered instruction in Danish class (i.e., the Danish subscale is used as dependent variable). Table 4 presents the effects of each version of the treatment on student-centered instruction in Math class (i.e., the math subscale is used as dependent variables).

[Table 3 here]

Table 3 lists three estimated models of the treatment effects in Danish classes. Model I estimates the effect by comparing the basic treatment to the parallel control group of the basic treatment. Thus, the estimated effect of the basic treatment on student-centered instruction in Danish class is 0.42 ($p < 0.05$). In Model III, we can observe the estimated effect of the basic treatment to be 0.50 ($p < 0.05$) when using the clean control group as reference category.

In Model II, we see that the effect of the extended treatment on student-centered instruction in Danish class is 0.73 ($p < 0.001$), when the parallel control group of the extended treatment is applied. The effect of the extended treatment is 0.75 ($p < 0.001$) when the clean control group is applied (see Model III). This is somewhat larger than the overall effect from Table 2. Moreover, it is about twice as large as the difference between the student-centered instruction in Danish class and Math class at baseline. In sum, both the basic and the extended treatment lead to higher student involvement when compared to control groups where teachers were not exposed to training. We infer from this that there is a substantial effect of the training treatment in Danish classes. Table 3 also shows the effect is larger for the extended treatment in all three models. Hence, as expected, the effect of training is larger when the training is more intensive.

The large effects of the training treatment in Table 3 indicate that most of the overall effect on student-centered instruction can be attributed to effects in Danish class. This is supported when we observe Table 4, which examines the effects of the training treatment in Math class. Models I and III show that the basic treatment has no effect on the student-centered instruction in Math. This is important, though, when considering whether any effects of the training treatments could be caused by a contamination effect (students in the treatment groups may be aware that the teacher has participated in training and this may be reflected in their evaluations). If the results were produced by such a contamination effect, we should also expect to observe an effect of the basic treatment in Math class. Therefore, it seems unlikely that the found effects of the training treatment in Danish class are driven by a contamination effect.

[Table 4 here]

The effect of the extended treatment is 0.41 when we compare with the parallel control group in Model II. However, the p-value is 0.10 and thus over the conventional threshold of 0.05. When compared to the clean control group, the estimated effect is 0.55 ($p < 0.05$), which could indicate that there is a little experimental contamination that causes us to underestimate the effect when applying the parallel control group. In combination, the results of Models II and II indicate that the extended treatment has an effect on student-centered instruction in math. Yet, it does not change the fact that most of the overall effect of the training treatment can be attributed to the effect in Danish class.

DISCUSSION

Public management is never easy. Results are often hard to measure, public street-level employees have wide discretion and individual goals, and control and reward systems are frequently met with resistance or even hostility. Extant research in public management tells us that classical management tools can be effective. Public employees sometimes respond as intended to incentives, resulting in better performance (one clear example in public schools is Lavy 2015), and measurement and control systems can make public employees change their focus to reflect organizational goals. However, an even more obvious result is that these tools sometimes do not work as intended (Langbein, 2010). For example, incentives can crowd out motivation (Frey, 1997; Deci and Ryan 1985), or outputs can be so hard to measure that incentives of sufficient power are simply unattainable (Dixit 2002).

In practice, public managers use many tools to affect employee behavior. We take up Simon's (1976 [1947]) old but neglected idea that training is a key mode of influence in

organizations. To address the problem of endogeneity (change-oriented training is likely to be implemented in practice as a reaction to present employee behavior), we rely on a randomized field experiment to test the effect of change-oriented training on behavior. The field experiment shows that change-oriented training can in fact direct and change the behavior of Danish teachers in public schools. Teachers have wide discretion in the classroom, and it is difficult and costly for managers to observe teacher behavior. The training treatment was designed to motivate and qualify teachers to increase student involvement in learning. A training treatment of moderate intensity involving three workshops had a highly significant effect of 0.23-0.27 standard deviations on a scale of student-centered instruction in Danish. A similar, but more intensive, treatment had a stronger effect of 0.36-0.40 standard deviations on the same scale. This effect is comparable to the effect of subject, which is 0.20 at baseline in favor of Danish. Thus, we find substantial effect sizes of the treatment in Danish classes. The results could partly be produced by a contamination effect. If students are aware that the teacher has participated in training, they might adjust their evaluation of student involvement accordingly. Since we find similar effects when excluding students having participated in the training courses and since we do not find an effect of the basic treatment in Math, the effect of contamination is likely to be limited. We conclude from this that training can have a strong effect on employee behavior and that training should be considered a powerful management tool. Whether the change in behavior affects outcomes such as learning is of course another question. So is the question of the long-term effects of change-oriented training. An interesting question for future research is the longevity of the changes. Future studies can address this by supplementing by measurement of changes in behavior after longer periods.

We find more limited effects for Math teachers. This may indicate that the effectiveness of training as a management tool is contingent, just as the classical management tools. The study has not been designed to test the scope conditions for training to be effective in changing behavior, but the different results for teachers in Danish and Math lead us to speculate that the strength of professional norms governing employee behavior is important to how training works as a management tool. The most conspicuous difference between the field experiment of Math and Danish teachers is the different nature of the two subjects. A possible explanation of the difference in the effects of the training treatment in the two subjects is that training is more effective in affecting behavior when professional norms prescribe a wide variety of practices, which may be more prevalent in Danish than in Math. If professional public employees already have clear ideas about how the work must be done, a training treatment may not be very effective in changing behavior.

The implication for public managers is that training should be seen as a management tool complementary to the classical control and incentive systems. When management seeks to change behavior in ways consistent with existing professional norms, training may be a particularly effective tool, while classical tools can be treacherous. If they are perceived as controlling, they may have detrimental effects because they crowd out existing motivation. However, when management wants to change behavior in ways that run counter to existing professional norms, training is likely to be less effective. In such situations classical tools can force employees to change. Hence, concluding that training is always a useful management tool is too simplistic. It is more accurate to conclude that training can be an effective tool but that the effectiveness depends on factors such as the nature of the task, the type of employees, and the existence of professional norms.

Change-oriented training differs from classical management tools and has several potential strengths seen from the perspective of a public manager. First, training is attractive to employees. It affects employee productivity and has pay-offs in higher wages (Barron, Black, and Loewenstein 1989). Second, in addition to affecting employee behavior, it can add to human capital. Third, employees may see training as an investment in them, which, again, can have positive effects on loyalty and motivation. Under the right circumstances, training can result in the achievement of several goals simultaneously. Another important strength is that the effect of training may be of a more permanent nature. If training reduces the cost of a specific behavior, and if training transforms employees' values, the effect sticks. Classical control and incentive systems only work as long as they are in effect. We measure changes in teaching practice 2-3 months after the treatment, but it is quite possible that the effects of change-oriented training last longer than that, and longer than the effects of traditional systems. Finally, training is a tool that can reach where traditional tools cannot. Control and incentive systems require that output or outcomes are at least partially observable. Training can work even under complete information asymmetry.

We have argued that training can affect behavior through two mechanisms, which both imply that as a management tool, change-oriented training can be used to change employee behavior towards the goals set out by politicians and public managers.: 1) Training affects human capital; it imparts knowledge and skills on employees. This reduces the cost of using new techniques or work strategies for employees. 2) Training can shift values among employees towards those stressed in training. In the study we have been unable to address these mechanisms directly, but we encourage future studies to further target the mechanisms of employee training in order to also provide better understanding of how different aspects of training can successfully

direct employee behavior towards organizational goals in a way that preserves employee motivation and understandings of professional discretion. In the current study we found no indications of negative effects of training, but just like any other managerial intervention it is possible that, for some employees or under certain conditions, training may involve negative implications such as motivation crowding out or limited skill transfer. Looking further into employee responses to training is therefore an important avenue for future research.

Change-oriented training does, however, have certain weaknesses as a management tool. First and foremost, the results indicate that the effectiveness of training depends on whether the desired change in behavior is compatible with existing professional norms. Second, training is a substantial investment. It is costly, requires planning, and is time-consuming. Finally, the permanent nature of the effects may also be a liability. While traditional tools may easily be adapted or abandoned, it is harder to reverse the effects of training.

REFERENCES

- Aguinis, H., and K. Kraiger. 2009. "Benefits of Training and Development for Individuals, Teams, Organizations, and Society." *Annual Review of Psychology* 60: 451–74.
- Andersen, L. B., and M. L. Jakobsen. 2010. "Does ownership matter for the provision of professionalized services? Hip operations at publicly and privately owned clinics in Denmark." *Public Administration* 89: 956–74.
- Andersen, L. B. 2009. "What determines the behaviour and performance of health professionals? Public service motivation, professional norms and/or economic incentives." *International Review of Administrative Sciences* 75, no.1: 79–97. <https://doi.org/10.1177/0020852308099507>
- Andersen, L. B., A. Boesen, and L. H. Pedersen. 2016. "Performance in public organizations: Clarifying the conceptual space." *Public Administration Review* 76, no. 6: 852–62.
- Andersen, L. B., E. Heinesen, and L. H. Pedersen. 2014. "How does public service motivation among teachers affect student performance in schools?" *Journal of Public Administration Research and Theory* 24, no. 3: 651–71.
- Andersen, L. B., T. B. Jørgensen, A. M. Kjeldsen, L. H. Pedersen, and K. Vrangbaek. 2013. "Public values and public service motivation: Conceptual and empirical relationships." *The American Review of Public Administration* 43, no. 3: 292–311.
- Andersen, R. H., A. H. Jeppesen, and T. Gloy. 2014. *Elevinddragelse i undervisningen : en vej til øget faglighed, trivsel og samfundsengagement*. Randers: Danske Skoleelever.
- Antonacopoulou, E., & Güttel, W. 2010. "Staff induction practices and organizational socialization: A review and extension of the debate." *Society and Business Review* 5, no. 1: 22-47.

- Antorini, C. 2014. "Folkeskolereformen giver eleverne mere indflydelse." *Magasinet Politik* 26.
- Retrieved (01.22.2018) at <http://magasinetpolitik.dk/folkeskolereformen-giver-eleverne-mere-indflydelse/>.
- Appelbaum, E., T. Bailey, P. Berg, and A. L. Kalleberg. 2000. *Manufacturing Advantage: Why High Performance Work Systems Pay Off*. Ithaca: Cornell University Press.
- Argyris, Chris. 1995. "Action science and organizational learning." *Journal of Managerial Psychology* 10, no. 6: 20-26. <https://doi.org/10.1108/02683949510093849>
- Arthur Jr, W., W. Bennett Jr, P. S. Edens, and S. T. Bell. 2003. "Effectiveness of training in organizations: A meta-analysis of design and evaluation features." *Journal of Applied Psychology* 88, no.2: 234–45. <http://dx.doi.org/10.1037/0021-9010.88.2.234>
- Association of Danish Pupils. 2012. *Elevinddragelse: velfærdspotentiale ved øget samproduktion i skolen*. Randers: Danske Skoleelever.
- Barron, John M., Dan A. Black, and Mark A. Loewenstein. 1989. "Job matching and on-the-job training." *Journal of Labor Economics* 7, no. 1: 1-19. <http://dx.doi.org/10.1086/298196>
- Beyer, M., M. Buchardt, J. Bøndergaard, P. F. Lauersen, L. T. Nielsen, and H. Plauborg. 2004. *Læreres læring – aktionsforskning i folkeskolen*. Copenhagen: CVU København & Nordsjælland KLEO.
- Blom-Hansen, J., R. Morton, and S. Serritzlew. 2015. "Experiments in public management research." *International Public Management Journal* 18:151–70.
- Brehm, J., and S. Gates. 1999. *Working, Shirking, and Sabotage: Bureaucratic Response to a Democratic Public*. Ann Arbor, MI: University of Michigan Press.

- Cannon, R. A., and D. Newble. 2000. *A Handbook for Teachers in Universities and Colleges: A Guide to Improving Teaching Methods*. London: Kogan Pagens.
- Combs, J., Y. Liu, A. Hall and D. Ketchen. 2006. "How much do high-performance work practices matter? A meta-analysis of their effects on organizational performance." *Personnel Psychology* 59, no. 3: 501–28.
- Davis, F. D., and M. Y. Yi. 2004. "Improving computer skill training: Behavior modeling, symbolic mental rehearsal, and the role of knowledge structures." *Journal of Applied Psychology* 89, no.3: 509–23. <http://dx.doi.org/10.1037/0021-9010.89.3.509>
- Deci, E. L., and R. M. Ryan. 1985. "The general causality orientations scale: Self-determination in personality." *Journal of Research in Personality* 19: 109–34. DOI: [10.1016/0092-6566\(85\)90023-6](https://doi.org/10.1016/0092-6566(85)90023-6)
- Desimone, L. M. 2009. "Improving Impact Studies of Teachers' Professional Development: Toward Better Conceptualizations and Measures." *Educational Researcher* 38:181–99. <https://doi.org/10.3102/0013189X08331140>
- Dixit, A. 2002. Incentives and Organizations in the Public Sector: An Interpretative Review. *Journal of Human Resources* 37, no. 4: 696–727. DOI: [10.2307/3069614](https://doi.org/10.2307/3069614)
- Dobbin, Frank, & Alexandra Kalev. (2016). Why Diversity Programs Fail and What Works Better. *Harvard Business Review* 94, no. 7-8: 52-60
- Frayne, C. A. and J. M. Geringer. 2000. "Self-management training for improving job performance: A field experiment involving salespeople." *Journal of Applied Psychology* 85, no.3: 361-372. <http://dx.doi.org/10.1037/0021-9010.85.3.361>

- Frey, B. S. 1997. *Not Just for the Money*. Cheltenham: Edward Elgar.
- Gailmard, S. 2010. "Politics, principal–agent problems, and public service motivation." *International Public Management Journal* 13: 35–45.
- Garavan, T. N. 2007. "A Strategic Perspective on Human Resource Development." *Advances in Developing Human Resources* 9: 11–30. <https://doi.org/10.1177/1523422306294492>
- Grossman, R. and E. Salas. 2011. "The Transfer of Training: What Really Matters." *International Journal of Training and Development* 15, no.2: 103-120. <http://dx.doi.org/10.1111/j.1468-2419.2011.00373.x>
- Holten, A.-L., A. Bøllingtoft, and I. Wilms. 2015. Leadership in a changing world: developing managers through a teaching and learning programme. *Management Decision*. 53, no.5: 1107–24. <https://doi.org/10.1108/md-10-2014-0617>
- Huselid, M. A. 1995. "The Impact of Human Resource Management Practices on Turnover, Productivity, and Corporate Financial Performance." *Academy of Management Journal* 38, no.3: 635–72. <https://doi.org/10.5465/256741>
- Jacob, R., R. Goddard, M. Kim, R. Miller, and Y. Goddard. 2015. "Exploring the Causal Impact of the Mcrel Balanced Leadership Program on Leadership, Principal Efficacy, Instructional Climate, Educator Turnover, and Student Achievement." *Educational Evaluation and Policy Analysis* 37, no. 3: 314-32. <https://doi.org/10.3102/0162373714549620>
- Jacobides, M. G., and D. C. Croson. 2001. "Information Policy: Shaping the Value of Agency Relationships." *Academy of Management Review* 26: 202–23.

- Jacobsen, C. B., and L. E. Jensen. 2017. "Why Not 'Just for the Money'? An Experimental Vignette Study of the Cognitive Price Effects and Crowding Effects of Performance-Related Pay." *Public Performance & Management Review* 40: 551–80.
<https://doi.org/10.1080/15309576.2017.1289850>
- Jacobsen, C. B., J. Hvitved, and L. B. Andersen. 2014. "Command and motivation: How the perception of external interventions relates to intrinsic motivation and public service motivation." *Public Administration* 92: 790–806.
- Jiang, K., D. P. Lepak, J. Hu, and J. C. Baer. 2012. "How Does Human Resource Management Influence Organizational Outcomes? A Meta-analytic Investigation of Mediating Mechanisms." *Academy of Management Journal*, 55, no.6: 1264-94. <https://doi.org/10.5465/amj.2011.0088>
- Kirkpatrick, D. L. 1979. "Techniques for evaluating training programs." *Journal of American Society for Training and Development* 33: 78–92.
- Kolb, D. A. 2014. *Experiential Learning: Experience as the Source of Learning and Development*. New Jersey: Pearson Education.
- Kraiger, K. 2002. *Creating, Implementing, and Managing Effective Training and Development: State-of-the-Art lessons for Practice*. San Francisco: Jossey-Bass.
- Kroll, A., and D. P. Moynihan. 2015. "Does Training Matter? Evidence from Performance Management Reforms." *Public Administration Review* 75, no.3: 411–20. <https://doi.org/10.1111/puar.12331>

Langbein, L. 2010. "Economics, Public Service Motivation, and Pay for Performance: Complements or Substitutes?" *International Public Management Journal* 13: 9–23.

<https://doi.org/10.1080/10967490903547134>

Lavy, V. 2015. "Teachers' Pay for Performance in the Long-Run." NBER Working Paper No. 20983.

Lewin, K. 1946. "Action research and minority problems." *Journal of Social Issues* 2, no.4: 34–46.

<http://dx.doi.org/10.1111/j.1540-4560.1946.tb02295.x>

Lipsky, M. 2010. *Street-level Bureaucracy, 30th Ann. Ed.: Dilemmas of the Individual in Public Service*.

Russell Sage Foundation.

May, P. J., and S. C. Winter. 2007. "Politicians, Managers, and Street-level Bureaucrats: Influences on Policy Implementation." *Journal of Public Administration Research and Theory* 19: 453–76.

Miller, G. J. 2005. "The Political Evolution of Principal-agent Models." *Annual Review of Political Science* 8: 203–25.

Moyson, S., N. Raaphorst, S. Groeneveld, and S. Van de Walle. 2018. "Organizational Socialization in Public Administration Research: A Systematic Review and Directions for Future Research." *The American Review of Public Administration* 48, no.6: 610-27.

<https://doi.org/10.1177/0275074017696160>

Noe, R. A., and M. J. Tews. 2009. "Strategic training and development." In *The Routledge Companion to Strategic Human Resource Management*, eds. J. Storey, P.M. Wright and D. Ulrich, 262–84. London: Routledge.

- Oberfield, Z. 2014. "Motivation, Change, and Stability: Findings From an Urban Police Department." *American Review of Public Administration*, 44, no.2: 210-32.
<https://doi.org/10.1177/0275074012461297>
- Paarlberg, L. E., and J. L. Perry. 2007. "Values Management: Aligning Employee Values and Organization Goals." *The American Review of Public Administration* 37, no.4: 387–408.
<https://doi.org/10.1177/0275074006297238>
- Perry, J. L., and L. R. Wise. 1990. "The motivational bases of public service." *Public Administration Review* 50: 367–73. <https://doi.org/10.1080/23276665.2014.892272>
- Perry, J. L., A. Hondeghem, and L. R. Wise. 2010. "Revisiting the Motivational Bases of Public Service: Twenty Years of Research and an Agenda for the Future." *Public Administration Review* 70, no.5: 681–90. <https://doi.org/10.1111/j.1540-6210.2010.02196.x>
- Seidle, B., S. Fernandez, and J. L. Perry. 2016. "Do Leadership Training and Development Make a Difference in the Public Sector? A panel study." *Public Administration Review* 76, no.4: 603–13.
<https://doi.org/10.1111/puar.12531>
- Simon, H. A. 1976. *Administrative Behavior*, 3rd edition. New York: Free Press.
- Tharenou, P. 2009. "Training and Development in Organizations." In *The Sage Handbook of Human Resource Management*, eds. A. Wilkonson, N. Bacon, T. Redman, and S. Snell, 155–72.
London: Sage.
- Tummers, L. L., V. Bekkers, E. Vink, and M. Musheno. 2015. Coping during public service delivery: A conceptualization and systematic review of the literature. *Journal of Public Administration Research and Theory* 25, no. 4: 1099–1126.

- Van Maanen, J., & Schein, E. (1979). "Toward a Theory of Organizational Socialization." In *Research in Organizational Behavior* (Vol. 1), edited by B. Staw, 209-264. Greenwich, CT: JAI Press.
- Van Ryzin, G. G. 2011. "Outcomes, Process, and Trust of Civil Servants." *Journal of Public Administration Research and Theory* 21, no.4: 745–60. <https://doi.org/10.1093/jopart/muq092>
- Walker, R. M., and R. Andrews. 2013. "Local Government Management and Performance: A Review of Evidence." *Journal of Public Administration Research and Theory* 25: 101–33.
- Weibel, A., K. Rost, and M. Osterloh. 2009. "Pay for Performance in the Public Sector – Benefits and (Hidden) Costs." *Journal of Public Administration Research and Theory* 20, no.2: 387–412. <https://doi.org/10.1093/jopart/mup009>
- Wilson, J. Q. 1989. *Bureaucracy: What Government Agencies Do and Why They Do it*. Basic Books.
- Winterton, J. 2007. "Training, Development, and Competence." In *The Oxford Handbook of Human Resource Management*, edited by P. J. Boxall, and P. Wright, 324–43. Oxford: Oxford University Press.
- Wright, B. E. 2004. "The Role of Work Context in Work Motivation: A Public Sector Application of Goal and Social Cognitive Theories." *Journal of Public Administration* 14, no.1: 59-78. <https://doi.org/10.1093/jopart/muh004>
- Yelon, S. L., and J. K. Ford. 1999. "Pursuing a Multidimensional View of Transfer." *Performance Improvement Quarterly* 12, no.3: 58–78. <https://doi.org/10.1111/j.1937-8327.1999.tb00138.x>

FIGURES AND TABLES

Figure 1. Assignment to Experimental Groups

Table 1. Balance Check, Means or Percentages (%)

	Training treatment	Parallel control	Clean control	N (total)
Student-centered instruction				
Overall measure (0-10)	4.28	4.05	4.12	2680
Danish subscale (0-10)	4.47	4.29	4.15	2680
Math subscale (0-10)	4.08	3.82	4.08	2680
Student background characteristics				
Test score, Danish, 6th grade (std.)	-0.01	0.01	0.00	2481
Test score, Math, 6th grade (std.)	-0.03	0.04	0.01	2472
Student's own activity in Danish class (0-10)	7.13	7.03	6.95	2676
Hours spent on Danish homework per week	2.36	2.36	2.20	2490
Student's interest in Danish (0-10)	6.03	5.82	5.73 *	2680
Student's own activity in Math class (0-10)	6.87	6.76	6.89	2676
Hours spent on Math homework per week	1.95	1.99	2.17	2483
Student's interest in Math (0-10)	6.11	6.02	6.25	2679
Student's well-being (0-10)	7.91	7.73 *	7.74	2650
Girl (%)	50	49	53	2665
Single parent (%)	22	18 **	20	2664
Age (years)	14.22	14.24	14.20	2664
No. of children in household	2.29	2.26	2.29	2664
Descent				

Denmark (%)	93	92	91	2664
Immigrant (%)	2	2	2	2664
Descendent of immigrants (%)	6	6	7	2664
Income, (1,000 DKK)	1.56	2.10 *	1.55	2644
Mother's marital status				
Married (%)	72	72	74	2623
Divorced (%)	16	16	15	2623
Never been married (%)	12	12	11	2623
Mother's age (years)	44.2	44.3	44.6	2623
Mother's descent				
Denmark (%)	90	90	89	2623
Immigrant (%)	10	10	11	2623
Descendent of immigrants (%)	0	0	0	2623
Mother's highest education				
Primary school (%)	14	13	12	2599
Secondary school (%)	06	06	06	2599
Vocational school (%)	38	41	35	2599
Short post-secondary education (%)	05	05	05	2599
Medium length post-secondary education (%)	28	26	31	2599
Long post-secondary education (%)	09	10	11	2599
Mother's income, (1,000 DKK)	333.84	340.15	344.68	2623
Father's marital status				
Married (%)	72	73	72	2570
Divorced (%)	16	15	18	2570

Never been married (%)	12	12	10 *	2570
Father's age	46.61	46.76	47.32	2570
Father's descent				
Denmark (%)	90	90	90	2570
Immigrant (%)	10	09	10	2570
Descendent of immigrants (%)	00	01	00	2570
Father's highest education				
Primary school (%)	16	16	13 *	2539
Secondary school (%)	05	03 **	04	2539
Vocational school (%)	44	44	44	2539
Short post-secondary education (%)	09	09	07	2539
Medium length post-secondary education (%)	14	15	19 **	2539
Long post-secondary education (%)	11	13	13	2539
Father's income, (1,000 DKK)	487.12	492.88	497.32	2623
Class and school characteristics				
Class size	22.42	22.80	22.38	214
Region				
Capital (Copenhagen) (%)	22	20	22	214
Middle Jutland (%)	26	30	26	214
North Jutland (%)	11	10	11	214
Zealand (%)	18	10	14	214
Southern Denmark (%)	24	30	27	214

Note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$, two-sided significance tests of the differences between the treatment group and each of the two control groups. The significance tests include the stratification variables and cluster robust standard errors.

Table 2. Effects of Change-Oriented Training on Overall Student-Centered Instruction. First difference Models.

	Model I	Model II
Training Treatment	0.39 (0.12)**	0.51 (0.16)**
Parallel control	(ref.)	0.12 (0.15)
Clean control	-0.12 (0.15)	(ref.)
Constant	-0.29 (0.08)***	-0.40 (0.13)**
N	2680	2680

Note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$, two-sided significance tests. Cluster robust standard errors in parentheses (cluster: class). Dependent variable: change in overall measure of student-centered instruction.

Table 3. Effects of Change-Oriented Training on Student-Instructed Teaching Practice in Danish Class. First Difference Models.

	Model I	Model II	Model III
Basic Treatment	0.42 (0.19)*	0.48 (0.18)**	0.50 (0.19)*
Extended Treatment	0.67 (0.21)**	0.73 (0.20)***	0.75 (0.21)***
Parallel control to Basic Treatment	(ref.)	0.06 (0.19)	0.08 (0.20)
Parallel control to Extended Treatment	-0.06 (0.19)	(ref.)	0.02 (0.19)
Clean control group	-0.08 (0.20)	-0.02 (0.19)	(ref.)
Constant	-0.48 (0.14)***	-0.54 (0.13)***	-0.56 (0.14)***
N	2680	2680	2680

Note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$, two-sided significance tests. Cluster robust standard errors in parentheses (cluster: class). Dependent variable: change in Danish subscale measure of student-centered instruction.

Table 4. Effects of Change-Oriented Training on Student-Instructed Teaching Practice in Math Class. First Difference Models.

	Model I	Model II	Model III
Basic Treatment	-0.01 (0.24)	0.08 (0.24)	0.22 (0.28)
Extended Treatment	0.32 (0.24)	0.41 (0.25)	0.55 (0.28)*
Parallel control to Basic Treatment	(ref.)	0.09 (0.19)	0.23 (0.23)
Parallel control to Extended Treatment	-0.09 (0.19)	(ref.)	0.14 (0.24)
Clean control	-0.23 (0.23)	-0.14 (0.24)	(ref.)
Constant	-0.01 (0.13)	-0.10 (0.14)	-0.25 (0.19)
N	2680	2680	2680

Note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$, two-sided significance tests. Cluster robust standard errors in parentheses (cluster: class). Dependent variable: change in Danish subscale measure of student-centered instruction.