



AARHUS UNIVERSITY



Cover sheet

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

The content in the accepted manuscript version is identical to the final published version, although typography and layout may differ.

How to cite this publication

Please cite the final published version:

Jacobsen, C. B., Andersen, L. B., Bøllingtoft, A., & Eriksen, T. L. M. (2022). Can Leadership Training Improve Organizational Effectiveness? Evidence from a Randomized Field Experiment on Transformational and Transactional Leadership. *Public Administration Review*, 82(1), 117-131.
<https://doi.org/10.1111/puar.13356>

Publication metadata

Title: Can Leadership Training Improve Organizational Effectiveness? Evidence from a Randomized Field Experiment on Transformational and Transactional Leadership

Author(s): Christian Bøtcher Jacobsen, Lotte Bøgh Andersen, Anne Bøllingtoft, Tine Louise Mundbjerg Eriksen

Journal: *Public Administration Review*, 82(1), 117-131

DOI/Link: <https://doi.org/10.1111/puar.13356>

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.

If the document is published under a Creative Commons license, this applies instead of the general rights.

Can Leadership Training Improve Organizational Effectiveness? Evidence from a Randomized Field Experiment on Transformational and Transactional Leadership

Christian Bøtcher Jacobsen

christianj@ps.au.dk

Department of Political Science, Aarhus University

Bartholins Alle 7, 8000 Aarhus C, Denmark

+45 8716 5431

Lotte Bøgh Andersen

lotte@ps.au.dk

Department of Political Science, Aarhus University

Bartholins Alle 7, 8000 Aarhus C, Denmark

+45 8716 5726

Anne Bøllingtoft

anne@mgmt.au.dk

Department of Management, Aarhus University

Fuglesangs Allé 4, 8210 Aarhus V, Denmark

+45 93508319

Tine Louise Mundbjerg Eriksen

tier@vive.dk

The Danish Center for Social Science Research

Olof Palmes Allé 22, 8200 Aarhus N, Denmark

+45 6169 7818

Author Biographies

Christian Bøtcher Jacobsen is professor at the Department of Political Science at Aarhus University, Denmark and co-director in Crown Prince Frederik Center for Public Leadership. His research focuses on leadership, motivation, and performance in the public sector with particular attention to the healthcare sector.

Lotte Bøgh Andersen is professor at the Department of Political Science at Aarhus University, Denmark. Her research focuses on leadership, management and administration in public organizations, especially motivation and performance of public employees, leadership strategies, professional norms and economic incentives. She is center director of Crown Prince Frederik Center for Public Leadership.

Anne Bøllingtoft is associate professor at Department of Management at Aarhus University, Denmark, and a member of Crown Prince Frederik Center for Public Leadership at Aarhus University. Her research interests focus on organizational behavior as well as leadership training and development in both the public and private sector. E-mail: anne@mgmt.au.dk

Tine Louise Mundbjerg Eriksen (PhD in economics) is employed as a senior researcher at The Danish Center for Social Science Research and Crown Prince Frederik Center for Public Leadership at Aarhus University. She works with empirical econometrics employing quantitative research methods using registry and survey data within the fields of personnel economics including topics such as leadership, the psychological work environment and performance as well as the economics of education focusing on school performance and well-being.

Can Leadership Training Improve Organizational Effectiveness? Evidence from a Randomized Field Experiment on Transformational and Transactional Leadership

Abstract: Based on evidence from a large-scale leadership training field experiment this article advances our knowledge about the possibilities for training leaders towards more active and effective leadership. In the field experiment, public and private leaders were randomly assigned to a control group or one of three leadership training programs: Transformational, transactional, or a combined program. Employee responses from 463 organizations show that the training can affect leadership behavior positively in very different organizations (primary and secondary schools, daycare centers, tax centers, and bank units). Furthermore, for the subsample of school principals, we find some evidence of training effects on performance in standardized tests in elementary schools and final exams in lower secondary schools. We discuss these findings in relation to training content and performance criteria.

Practitioner Points

- Training leaders in combined use of transformational and transactional leadership has the greatest positive effects on employee-perceived leadership behavior.
- The results have broad applicability as the study includes leaders from welfare and financial service sectors, representing both public and private ownership status.
- Although mostly statistically insignificant, our results on performance indicate that it may be important to pay attention to potentially different effects across different performance criteria.

The importance of leadership for directing, driving, and developing high-performing organizations is acknowledged by researchers and practitioners alike (Van Wart 2013). Public and private organizations worldwide devote considerable resources to internal and external leadership training programs with the aim of improving leadership and performance (Seidle, Fernandez, and Perry 2016). However, the value of leadership training is contested, and there is a lack of solid research-based evidence on the causal effects of leadership training on leadership behavior and performance.

Although they rarely uncover causal effects, generic management studies generally point at positive effects of leadership training on behavior and performance (Avolio et al. 2009).

Recently, leadership training has also been linked with more active leadership behavior in public organizations (Kroll and Moynihan 2015; Seidle, Fernandez, and Perry 2016). A few field experiments support that leadership training can have positive *causal* effects on performance (e.g., Dvir et al. 2002), but these studies examine specific types of leadership training in specific organizational contexts. Thus, we lack comparative studies of leadership training programs across different types of organizations to advance our understanding of the more general usefulness and causal effects of leadership training.

Based on a large field experiment (the LEAP (LEadership And Performance) project) of 673 leaders from very different private and public organizations, i.e., primary and secondary schools, private and public daycare centers, tax units, and bank branches in Denmark, this article makes three important contributions to our knowledge about the *causal* effects of leadership training on leadership behavior and organizational performance. First, existing field experimental studies mainly investigate transformational leadership (Avolio et al. 2009), which actually should be seen as complementary to transactional leadership (Bass et al. 2003). This study investigates stand-alone transformational and transactional leadership as well as combined use of the two leadership strategies, which allows a unique comparison of

the importance of leadership training content. Second, there is a profound need for studies that examine organizations with different functions and ownership status (Day 2013; Fernandez 2005; Moynihan, Pandey, and Wright 2012; Trottier, Van Wart, and Wang 2008; Wright, Moynihan, and Pandey 2012). Studies have investigated one or a few types of organizations, whereas this study includes leaders from welfare and financial service sectors in publicly and privately owned organizations, which enables us to advance our knowledge on leadership training effects in broader settings. Third, existing studies focus on immediate effects measured shortly after training, but such effects may soon wear off (Avolio et al. 2009, 780–81). This study investigates longer-term leadership training effects on leadership behavior (three months after end of training) and objective school performance (one year after end of training). Effects of training creates value through perceived leadership behavior (An et al. 2020), motivation (Bro and Jensen 2020), and value congruence (Jensen 2018), but the value of leadership training should ultimately be assessed by performance effects. These are, nonetheless, complex and multidimensional (Andersen, Heinesen, and Pedersen 2016). This study offers insights on various performance effects in schools where academic results (exam and test scores, and completion rates) are widely applied and useful because they reflect salient and comparable performance dimensions (Andersen, Heinesen, and Pedersen 2016; O’Toole and Meier 2011).

In the following sections, we formulate theoretical expectations based on existing literature and specify testable hypotheses. We discuss the field experiment we conducted, and we present expected and unexpected effects of leadership training on leadership behavior. For the subsample of school principals, we present analyses that suggest that leadership training can affect organizational performance. We conclude with a discussion of our findings and their implications for research and practice.

Transformational and Transactional Leadership

The distinction between “hard” quid pro quo leadership and “soft” leadership with focus on motivation and development is often conceptualized as transformational and transactional leadership strategies (Bass 1985). The dominant approach to measuring transformational leadership has been the multi-dimensional MLQ model (Bass and Riggio 2006), but recent studies have criticized the approach for conceptual and empirical overlap between dimensions. Furthermore, scholars increasingly underscore the visionary core of transformational leadership (Wilson 1989; Wright 2007; Wright, Moynihan, and Pandey 2012; Wright and Pandey 2010). We therefore apply a focused understanding of transformational leadership as consisting of three aspects: A transformational leader 1) develops a vision of the core goals of the organization, 2) strives to share this vision with the employees, and 3) makes an effort to sustain the shared vision in the long term (Jensen et al. 2019, 8–9). Thus, transformational leadership is conceptualized as leader behaviors that seek to develop, share, and sustain a vision. These behaviors are categorized as transformational because they are intended to stimulate employee internalization of the values and goals implied by the vision and make employees transcend their self-interest and strive toward organizational goals.

Transactional leadership, on the other hand, involves the use of contingent rewards (and/or sanctions) to align employees’ self-interest with organizational goals. When desired rewards (and undesired sanctions) are contingent on specific efforts or results, employee self-interest is expected to energize, direct, and sustain behavior toward those efforts and results.

Contingencies can be related to pecuniary/near-pecuniary rewards (e.g., bonuses), verbal rewards (e.g., praise), or sanctions (e.g., firing) (House 1998). We see these three types of contingencies as alternative ways to conduct transactional leadership, and we conceptualize transactional leadership as “the use of contingent rewards and sanctions to make individual

employees pursue their own self-interest while contributing to organizational goal attainment” (Jensen et al. 2019, 12). Verbal rewards are usually regarded as especially relevant for public managers because their use of bonuses, wage supplements, and promotions is often restrained. Furthermore, compared with pecuniary rewards, verbal rewards can be less controversial and prevent motivation crowding effects (Bellé 2015). Still, pecuniary rewards may also have positive implications in public organizations (Andersen and Pallesen 2008). Although sanctions are used in practice, they are often associated with negative effects, and the transactional training program therefore only focused on contingent pecuniary and verbal rewards.

Transformational and transactional leadership strategies are often treated as stand-alone leadership behaviors in empirical studies. However, research suggests that they are complementary with mutually augmenting effects (Bass 1990; Rainey 2009; Waldman, Bass, and Yammarino 1990) and that a combined strategy leads to higher performance than stand-alone leadership strategies (Bass et al. 2003; Hargis, Watt, and Piotrowski 2011; O’Shea et al. 2009; Rowold 2011). In order to inform practice and guide future research efforts, it is important to investigate the effects of combined leadership strategies in a robust field-experimental design.

Previous Experimental Leadership Training Studies

While the public leadership training literature is sparse (see Seidle, Fernandez, and Perry 2016 for an overview), there is a stronger tradition for studying leadership training in the generic management literature (Doh 2003). A meta-analysis identifies 62 experimental studies of leadership training and development that report predominantly positive effects (Avolio et al. 2009). Although publication bias may be an issue in this literature as in any other (Rosenthal, 1979), the existing evidence overly supports the argument that leadership

can be taught. However, many studies are lab experiments, and the results may not be valid in complex real-life organizational settings characterized by constraints, resistance, and constant changes. As expected, lab experiments are associated with stronger intervention effects than the few conducted field experiments (Avolio et al 2009, 776). Six field-experimental studies report causal effects of transformational leadership training. They all find significant increases in transformational leadership behaviors (Barling, Weber, and Kelloway 1996; Hassan, Fuwad, and Rauf 2010; Kelloway, Barling, and Helleur 2000; Parry and Sinha 2005) or significant effects on performance through improved transformational behaviors (Dvir et al. 2002; Kelloway, Barling, and Helleur 2000). Although studies have experimentally investigated the effects of incentives (e.g. Lazear 1999), we have not been able to identify field-experimental transactional leadership training studies.

Experimental evidence is especially useful if treatments can match existing experimental studies in terms of intensity and scope. In relation to intensity, the strongest treatments have been tested by Parry and Sinha (2005) in a three-month training program consisting of four days of intervention. In relation to scope, existing studies include no more than 54 leaders (32 for the training group and 22 for the control group) (Dvir et al. 2002) and no more than 50 leaders who receive training (all participating leaders, no control group) (Parry and Sinha 2005). As we will discuss later, this study includes data from 463 leaders in training programs of nine months' duration.

Leadership Training and Leadership Behavior

A classic way to depict leadership training outcomes is to explain training effects as a pathway from training through altered behavior to performance (Kirkpatrick 1979). Thus, altering behavior is vital in terms of affecting performance, and employees' perceptions of

leadership behavior is therefore a crucial intermediate outcome as leaders primarily achieve organizational performance through their employees.

According to the leadership training literature, training changes leadership behavior if the training includes feedback, classroom education, coaching, and reflections that involve work experiences (Seidle, Fernandez, and Perry 2016, 605). These features are expected to support meta-cognitive skill building, including “the ability to know how well one is performing, when one is likely to be accurate in judgement, and when one is likely to be in error” (Kruger and Dunning 1999, 1121). Our theoretical model depicts how participation in leadership training is expected to stimulate leaders to more active leadership behavior in the trained direction and thus stimulate employee effort and ultimately performance.

[FIGURE 1 HERE]

In general, leadership training offers knowledge about important leadership behaviors and can strengthen leaders’ ability to implement them in practice (Avolio et al. 2009; Seidle, Fernandez, and Perry 2016). Furthermore, effective leadership training leaves time and opportunities for individual and collective reflection about leadership practices, which can spur deeper understanding of leadership as well as more active leadership behavior (Holten, Bøllingtoft, and Wilms 2015).

As transformational leadership training focuses on improving the leaders’ knowledge of and ability to use visions to motivate and direct employees, we expect a training effect on transformational leadership behavior:

H1: Transformational leadership training has a positive effect on transformational leadership behavior.

In terms of transactional leadership, we distinguish between two types of contingent rewards (pecuniary and verbal). Transactional leadership training focuses on improving leaders' knowledge of and ability to use both types. We thus expect positive effects on both hypotheses:

H2: Transactional leadership training has a positive effect on use of contingent verbal rewards.

H3: Transactional leadership training has a positive effect on use of contingent pecuniary rewards.

The combined training program focuses on improving the leaders' use of visions and contingent rewards in combination. The combined leadership program may have a weaker effect on each type of leadership behavior than stand-alone programs given that time spent on training is the same in the programs. However, there may be synergies between the leadership strategies, which would counter this argument. We thus expect that the combined training program will affect both transformational and transactional leadership behaviors, but we refrain from presenting expectations about the relative size of the effects.

H4: Combined leadership training has a positive effect on use of transformational leadership.

H5: Combined leadership training has a positive effect on use of contingent verbal rewards.

H6: Combined leadership training has a positive effect on use of contingent pecuniary rewards.

Leadership and Organizational Performance

The ultimate value of leadership training lies in the potential to improve organizational performance. Performance is “the actual achievement of a unit relative to its intended achievements, such as the attainment of goals and objectives” (Jung 2011, 195). Performance in public service organizations is usually complex and multi-faceted (Andersen, Heinesen, and Pedersen 2016; Walker and Andrews 2013). In this article, we focus on organizational effectiveness, which is the achievement of formal objectives and therefore deemed a key aspect of performance (Boyne 2003).

Studies that link transformational leadership with higher performance (Judge and Piccolo 2004; Jung and Avolio 2000; Lowe, Kroeck, and Sivasubramaniam 1996) are mainly based on observational cross-sectional data, which may overestimate effects (Oberfield 2014).

Transformational leadership has been associated with positive performance effects among public employees (nurses) when they are also exposed to job design initiatives in a lab-like field experiment (Bellé 2014). Field experiments in the generic leadership training literature generally associate transformational leadership training with increased performance (Barling, Weber, and Kelloway 1996; Dvir et al. 2002; Hassan, Fuwad, and Rauf 2010; Kelloway, Barling, and Helleur 2000; Parry and Sinha 2005). We therefore expect that:

H7: Transformational leadership training has a positive effect on performance.

Attention to transactional leadership behaviors in relation to performance is much more limited. However, an experimental study of performance-contingent rewards among nurses shows that pecuniary rewards can increase performance considerably, although the effect is much smaller when pecuniary rewards are visible to others (Bellé 2015). A recent American study shows positive effects of contingent rewards to schoolteachers (Dee and Wyckoff 2015), and evidence from Israel confirms that performance-based pay in schools can affect short- and long-term student outcomes (Lavy 2009, 2020). In the generic management

literature, Lowe and colleagues' (1996) meta-analysis shows that contingent rewards are associated with higher effectiveness. We therefore expect that:

H8: Transactional leadership training has a positive effect on performance.

Older research suggests that transformational leadership augments transactional leadership in predicting performance (Bass 1985). Waldman, Bass, and Yammarino (1990) argue that leaders are more effective when they combine the leadership strategies because transformational behaviours reinforce contingent reward behaviours and increase subordinate effort and performance. Other studies that are not based on field-experimental evidence also suggest that combinations of transactional and transformational leadership increase performance more than either leadership strategy individually (Bass et al. 2003; Hargis, Watt, and Piotrowski 2011; O'Shea et al. 2009; Rowold 2011). This argument is echoed in public management research (Oberfield 2014; Wang et al. 2011). Based on this, we expect that:

H9: Combination leadership training has a positive effect on performance.

Design

We designed an experimental leadership training study, the LEAP project, with three training groups and one control group. The experiment included 673 leaders from public and private primary schools, public secondary schools, public and private daycare centers, tax units, and bank branches. This article analyses the 463 leaders who remained in the project and where we have valid post-treatment employee survey responses (see drop-out analyses).

To ensure an even representation of leaders from the different types of organizations across treatments, leaders were assigned to training groups through a stratified random sampling method with stratas based on nine subtypes of leaders (see table 1). Leaders were randomly assigned within stratas to avoid selection bias (Angrist and Pischke 2009, 15) and ensure that

participants were distributed to training and control groups independently of potential outcomes. To investigate training effects, leadership behavior was measured through employee surveys before and after training. For the participating school leaders, performance measures were available in the national registers.

Recruitment, Randomization and Dropout

The recruitment process varied across areas. Leaders in primary and secondary schools and daycare centers were invited directly by email. Leaders from two banks and Tax Denmark were invited and selected through their central HR departments, which resulted in high response rates. Pre-training surveys were mandatory for sign-up. Only leaders who had not previously participated in leadership training programs with similar content were enrolled, and they were not allowed to participate in similar courses during the experiment. The intended number of participants were successfully recruited from primary and lower secondary schools, tax units, and daycare centers, but not among upper secondary school principals and bank leaders. Many upper secondary school leaders had already completed similar training programs, and the banks' central managers preferred the banks' own training programs. Table 1 details the number of leaders who were invited and leaders who signed up.

[TABLE 1 ABOUT HERE]

Participants in each training program were assigned to seven classes (21 in total) with 15 to 25 participants based on the geographical location of their workplace. Four teachers (professors in economics, organizational behavior, and public administration) delivered the training, and the same teacher taught all lessons in a class. Each class received four eight-

hour days of leadership training. The teachers taught classes in all three programs and were randomly assigned to classes (Table A8¹). The training began in September 2014 and ended in May 2015. During training, the teachers registered deviations from the plan (see the online training description), and although some differences were noted, the leaders' evaluation of the training did not differ systematically among teachers. To ensure maximum inducement of training, leaders who were unable to participate in a session were offered participation in another class, preferably taught by the same teacher, and otherwise by another teacher. If participation was not possible, the training session could be watched on video and the teacher was available for discussion and reflection to stimulate leadership action.

Of the 673 leaders who signed up for the experiment, 506 completed the program (see table 1). Among the 166 leaders who dropped out during the training period, we have information about the reason from 69 percent. 28 percent dropped out due to work pressure, 21 percent because of job changes, 6 percent because of personal problems, and 6 percent because of illness. Only 9 percent left for training-related reasons. These were slightly overrepresented in the transformational training program where a few leaders – despite the initial screening of training experiences – had completed similar training. The remaining 31 percent of the 166 leaders who left the program did not specify a reason. Among the 503 completing managers, we obtained post-training employee survey data from 463 managers.

Training Program Design

The programs were designed to overcome barriers identified in existing research, such as cognitive processes, weak transferability, and lack of organizational support (Holten, Bøllingtoft, and Wilms 2015). Following insights from leadership training research, our programs incorporate principles from especially action learning and experiential learning theory to ensure that learning is transferred from the classroom to the workplace setting.

Difficulties in transferring learning is identified in the teaching and learning literature as a main reason for the limited effects of many leadership training programs (Allio 2005; Beer, Finnström, and Schrader 2016; Burgoyne, Hirsh, and Williams 2004). A detailed account of the teaching and learning principles is provided in the online appendix.

Teaching materials were developed by the four teachers and a researcher with experience from a previous leadership intervention. Teaching material (e.g., curriculum, information material, PowerPoint presentations, and exercises), structure (e.g., timing of lectures and group works), and written communication were identical in all classes in a training program. Only the substantial content varied between programs. To ensure coordinated use of teaching material, performed teaching, and equal emphasis on central learning points, the four teachers calibrated the training before the programs began.

A field experiment involves multiple ethical dilemmas, which we have handled based on explicit principles. First, we informed the participants that they would learn to employ tools that could potentially improve organizational goal attainment. However, we did not tell them about our theoretical hypotheses. Second, participation was each leader's own choice, and we reported individual results to them and only general results to their superiors. Third, withdrawal from the project was possible at all times since the experiment resembled ordinary training settings, but attrition was relatively low (25 percent as discussed in relation to table 1) and equal for the different groups.

Measures

Leadership Behavior

Leadership behavior is measured via questionnaires to employees distributed before training (September 2014) and three months after training (September 2015). We use responses from

the 7,316 employees who responded after the training.² Due to criticism of traditional leadership measurement instruments (Van Knippenberg and Sitkin 2013), we apply a recently validated instrument (Jensen et al. 2019). *Transactional leadership* is measured via questions about the use of specific contingent pecuniary and non-pecuniary rewards, and *transformational leadership* is measured via questions about how leaders seek to develop, share, and sustain a vision. Table A1 in the appendix presents the specific items, and the confirmatory factor analysis (Table A2) shows that the measurement instrument performs well. Assignment to training takes place at leader level, so we aggregate perceived leadership within organizations. Included leaders have leadership ratings from at least five employees. Analyses support that ratings show sufficient agreement between employees and response consistency between employees, i.e. group mean reliability, to allow this aggregation (Jensen, Moynihan, and Salomonsen 2018) (see table A2 in the Appendix for further details of r_{wg} and ICC scores).

School Performance

Performance information from elementary and lower secondary school leaders is obtained from Statistics Denmark. From the selected organizations, we extract all students enrolled by September 1, 2014 and 2016 from the Danish student registry. We have access to data on GPA (grade point average) and completion rates from the 9th-grade exit exam (usually taken the year the student turns 16) for 141 schools (out of 163).³ The 9th-grade exit exam comprises five mandatory exams (two mandatory oral exams in respectively Danish and English plus either joint physics/chemistry or biology and geography. The two written exams are in Danish and mathematics).⁴ GPA scores are standardized within cohorts.

For 116 public school leaders, we also have access to test scores in the Danish National Tests. These standardized tests are mandatory in public schools, and students are tested in reading in

grades 2, 4, 6, and 8, in mathematics in grades 3 and 6, in English in the seventh grade, and in physics and chemistry, biology, and geography in the eighth grade.⁵ We standardize the student test scores within each subject, class, and cohort, and obtain completion rates.

Additional Covariates

In order to account for factors such as attrition and missing employee answers (in the post-training survey), we create additional controls (covariates) from the pre-treatment survey.

Employee Job Satisfaction is obtained from the question “How satisfied are you with your job?” and measured on a scale from 0 to 10. The average of employee responses is calculated for each leader. *Leader Experience* is a self-reported measure of the number of years in the current position. *Male* is a dummy variable equal to 1 if the leader is male. *Leader Education* is a dummy equal to 1 if the leader has a Master’s degree in leadership. We further construct a measure of the share of employees who responded to the questionnaire (*Mean Employee Response Rate*) to take into account that the employee-based variables are constructed from different response rates.

In the performance analyses, registers obtained from Statistics Denmark enable us to include additional information on student demography, special needs education, and parental characteristics such as work income, labor force participation, and education using the student’s social security number. All characteristics are measured when the student is 13 years old when we investigate the effects of leadership training on final exams, and six years old (when the student enters school) when we investigate the effect on the standardized test scores. Since we employ school-level models, all characteristics are measured as school averages. Leader characteristics such as gender, age, and number of students are also included (see Table A8 for a full overview).

Identification Strategy

In the analysis of leadership behavior, we exploit the randomization design to obtain causal effects of leadership training. We use Ordinary Least Squares (OLS) according to equation (1) to account for features of our research design:

$$y_{ls} = \alpha + \beta T_{ls} + \theta_s + \varepsilon_{ls}, \quad (1)$$

where y_{ls} is the average perceived leadership style of leader l in sector s . T is a vector of the three different training dummies, and θ is a strata fixed effect (we stratified the randomization on sectors). Insofar as randomization was successful and selection out of the experiment was unrelated to unobservables, β will identify the average training effect on the treated of leadership training on perceived leadership.

In a similar fashion, we investigate the effects of leadership training on performance according to equation (2),

$$y_{ls2016} = \alpha + \beta T_{ls} + \gamma X_{ls2014} + \theta_s + \varepsilon_{ls}, \quad (2)$$

where y_{ls2016} is the average performance in the school of principal l in sector s . β identifies the intention-to-treat effect of leadership training. To increase precision, we include covariates X_{ls} measured in 2014. Furthermore, the covariates serve as an additional robustness check to the randomization assumption.⁶ Finally, θ_s is the sector fixed effect referring to public versus private schools. Since the national standardized tests are only conducted in public schools, θ_s does not enter the models with standardized test scores as the dependent variable.

Multiple Hypothesis Testing

To account for multiple hypothesis testing in the LEAP project, we report sharpened False Discovery Rate (FDR) q-values following Anderson (2008) in addition to original p-values.

The FDR controls for the proportion of rejections that are false discoveries. It is particularly well suited in analyses when one is testing a large range of outcomes and treatments. The advantage of the sharpened FDR q-values is that it allows us to adjust p-values across different model specifications making it possible to incorporate all findings from this and previously published articles in the LEAP project. In the adjustment, we include all results testing the effects of the intervention. This includes both interaction and main effects. Some articles gradually expand their conditioning set. In so far that the population behind the estimations are the same, we only include the most comprehensive model. The total list of results included in the FDR adjustment can be found in Table A10 in the online appendix. The tables in this article report FDR q-values for the full family of estimates. In the appendix, we further define a family of all estimates based on employee response outcomes and a family based on register outcomes (school performance and completion rates). We also define a set of families according to type of treatment. That is, all outcomes of those individuals exposed to the transformational leadership training constitute one family, while outcomes of those individuals exposed to the transactional leadership training constitute another family and so forth.

Balance Checks

To make sure that the randomization procedure worked, we conduct a set of balancing tests. Table A3 in the appendix shows balancing tables for leadership behavior. We only observe minor differences in the employee response rates between the transactional training group (10 percent significance level), the combination training group (5 percent significance level), and the control group. We conclude that the initial randomization was successful.

However, only employees of leaders who completed the leadership training received the post-treatment survey, which makes balance checks on these leaders even more relevant (table A4). According to the table, only minor imbalances exist among the remaining

transformational leaders who initially score lower on employee-perceived transformational leadership and contingent verbal rewards compared to the control group. This reflects the LEAP teachers' observation that some leaders in the transformational training group felt that they already mastered this type of leadership and therefore dropped out. Otherwise, the groups are well balanced, as are the organizations, which obtained employee responses in both waves (table A5). In order to consider the imbalances, we also estimate lagged dependent variable (LDV) models in the next section.⁷

Tables A8 and A9 show the balancing tests for the final exam and standardized test samples in 2014 before training. For both samples, the covariates balance rather well between the different training groups and the control groups. There are only few significant differences (e.g. in the final exam sample, mothers in the combination training group are less likely to be employed) and no more differences than would be expected considering the number of comparisons. We conclude that the randomization was successful at the school level. However, to ensure coherence with the results on perceived performance and conduct and additional robustness check, we include an LDV model in the performance analysis.

Effects of Leadership Training on Leadership Behavior

The first question concerns whether leadership training affects leadership behavior¹. The short answer is “yes”, but several nuances are relevant to emphasize. All three leadership training programs significantly affected the level of at least one type of employee-perceived leadership, but not all expectations were met. Table 2 gives an overview of all effects and refers to the more detailed results. Tables 3 and 4 report these details for leadership behavior, while Tables 5-9 will later provide the details for performance.

[TABLE 2 ABOUT HERE]

[TABLE 3 ABOUT HERE]

[TABLE 4 ABOUT HERE]

First, hypothesis 1 expects transformational leadership training to increase transformational leadership behavior. This relationship is only confirmed in the analysis that employ the balanced panel (p-value = 0.045 in Table 4). If we consider the simple means in Table 3 of all employees, the average level of transformational leadership behavior is lower in the after-treatment survey in organizations with leaders, who received the transformational leadership training. This is due to the mentioned attrition bias among leaders in the transformational training group. In this group, dropout was higher among leaders with high initial level of transformational leadership. However, if we only consider the responses of employees who remain in the organizations transformational leadership behavior increases in the transformational and combination training groups, whereas it drops in the transactional training group. Thus, while the OLS and LDV analyses on the full sample do not support hypothesis 1, the balanced panel LDV model does. In other words, we find positive development in transformational leadership behavior in the balanced panel. Table 3 also show that the significant effect of transformational leadership training on leadership behavior for the balanced panel can be explained by a large drop in transformational leadership behavior is due to a drop in the control group. This drop is apparent in all three measures of leadership behavior and indicates that leadership behavior deteriorates over time if not maintained (deterioration in the control group is also evident in Dvir et al. 2002). According to the LDV models, transformational leadership training also increased the use of contingent pecuniary rewards but did not change the use of verbal rewards significantly. When we

correct the results of transformational leadership training for multiple hypothesis testing, only the effect of transformational leadership on pecuniary rewards remains significant at a ten percent level (LDV and LDV Panel). This suggests that the results concerning the effects of transformational leadership training on leadership behavior should be interpreted with caution.

Second, hypotheses 2 and 3 expect transactional leadership training to increase the use of contingent verbal and pecuniary rewards. The results only support hypothesis 3.

Transactional training is associated with a relatively strong positive effect on pecuniary rewards, which is identified across all three models also including the adjustment for multiple hypothesis testing (see details in Table 4). Although the coefficients on verbal rewards are positive, the effects are not significant, which means that hypothesis 2 finds weak support here. Surprisingly, transactional leadership training is associated with increases in transformational leadership. However, the effect only remains significant at a ten percent level in the LDV model ($q\text{-value} = 0.074$) when we employ the adjustment for multiple hypothesis testing. If such an effect exists, it could be due to the shared goal orientation in transformational and transactional leadership, which we will return to in the discussion. However, when we consider the balanced panel in table 3, transformational leadership decreases among leaders who received the transactional treatment. Relative to the control group, transformational behavior decreases less in the transactional training group. Another interesting observation is that the increase in transformational leadership behavior is driven by a higher perception of transformational leadership among new employees and a lower perception among employees who leave. This suggests that shared goal orientation has the greatest effect when employees know the leader well.

Third, the combination training was expected to have a positive effect on transformational leadership (hypothesis 4), contingent verbal rewards (hypothesis 5), and contingent pecuniary

rewards (hypothesis 6). The results support hypothesis 4 that, the combined training program increased transformational leadership also when we adjust for multiple hypothesis testing. Looking across the three training programs, the coefficients are largest for the combination training program, although the differences to the other training programs are not statistically significant. There is weak support for hypothesis 5, given that combined training is associated with positive developments in the use of contingent verbal rewards across models, which were, nonetheless, not statistically significant when we adjust p-values for multiple hypothesis testing. Finally, the combination program had a consistently significant and sizeable effect on the use of pecuniary rewards. The level of significance drops when we adjust for multiple hypothesis testing, but the effect remains significant at the 10 percent level. Again, considering the balanced panel, this effect seems to be driven by a drop in the control group. Still, this supports hypothesis 6. Overall, the selection in and out of organizations seems to be an important consideration when we evaluate leadership training programs, as perceptions may very well rely on initial reference points, training may even affect who remains in the sample, etc. This is evident in our analyses, which suggests that some of our hypotheses are only confirmed when we study employees who remain in the organization throughout the leaders' training program.

In sum, a number of positive differences between control and training groups have been identified. The significant effects on leadership behavior are generally sizeable and correspond to between one-fourth and one-ninth of a standard deviation in the perceived leadership strategy. The overall conclusion is that transactional and combination leadership training affects employee-perceived use of contingent pecuniary rewards, while the impact of transformational leadership training on perceived transformational leadership behavior seems to be less strong. The combined use of transformational and transactional leadership has the most consistent positive effects.

Effects of Leadership Training on School Performance

Having established that the leadership training programs can to some extent positively affect employee-perceived leadership behavior, next step is to investigate whether the three types of leadership training can make a difference for organizational performance. These analyses are restricted to school principals. Again, table 2 provides an overview of the results; the details can be found in tables 5-9.

We investigate different aspects of performance in schools: GPA and completion rates from final exams, and test scores and completion rates from standardized tests. If we look at GPA from final exams for all schools, there is no support for hypothesis 7, 8 or 9: OLS and LDV regressions of leadership training on student GPA and completion rates in the 9th-grade exit exam show no effect. For transactional leadership, the coefficients are even negative and sizeable although insignificant (Table 6). It is estimated that Average GPA is reduced by 0.105 points in schools where the leaders received transactional training. For completion rates in final exams, the coefficients are generally small and insignificant for the full sample, but the sizeable increase in completion rate for transactional leadership training (see also table 5) indicates that the reduction in GPA can be due to more students taking the exams. Higher completion rates are typically associated with higher shares of academically weaker students taking the exam.

The final exam results are a conservative test of training effects because they reflect skills that are acquired over the entire course of schooling for a restricted group of students (and teachers). Therefore, to investigate a broader performance measure, which may be affected more directly by the training, we turn to standardized tests. Here the effects on the average test scores and completion rates are positive (except for transactional training in the OLS model, see table 7), but the coefficients are only statistically significant at a 10 percent level

for the transactional training in the LDV model. Once we adjust for multiple hypothesis testing, none remain statistically significant. The effects of transactional training on these test scores change slightly when covariates are included, suggesting some selection in the transactional schools that remain in the sample. The results give some support hypothesis 7, 8 and 9 and suggest that transformational, transactional and combined training increased average school performance in the standardized tests by 0.02–0.08 points. Compared to a mean of zero and a standard deviation of 0.27 in the control group, the effect sizes are substantial although insignificant. However, the insignificance can be an issue of statistical power as the sample size is relatively small. We find no effects on completion rates. The average level of completion rates show a large drop in completion rates in both treatment and control groups (see table 5). This highlights the importance of a control group, because it enables us to control for changes over time that affected the entire functional sector.

Tables 8 and 9 show results for schools whose principal completed the training program. These analyses are potentially affected by selection bias if attrition is not random. However, they are relevant because they show training effects for principals who received the full training. Table 8 shows that the statistically significant negative effect of transactional training in relation to GPA only appears in the OLS model without covariates. When we control for covariates or a lagged dependent variable, the coefficient remains negative, but it becomes smaller and insignificant. The coefficients are also negative, but still insignificant, for the other two training programs. We find no effects on completion rates. Table 9 on the other hand shows very positive effects of all three training programs on standardized test scores. The coefficients are sizeable, statistically significant, and similar across training programs. Once we adjust the p-values for multiple hypotheses testing, the coefficients on the transformational and transactional leadership training remain significant at a 10-percent level.

This suggests that finalizing the training programs is important in order for the training to be effective. This shows some support of hypothesis 7, 8 and 9.

Overall, we find some evidence that the training program affected organizational performance in the investigated schools. The transactional program seems to be associated with slightly lower GPA but also with higher test scores and possibly higher completion rates in the final exams. All three training programs are associated with higher standardized test scores for principals who completed the training programs. While effect sizes are of substantial magnitude, only few results remain significant when we adjust for multiple hypothesis testing. While this may be an issue of statistical power, we cannot rule out that the leadership training was unable to improve student performance in the investigated time period.

Discussion

The study has investigated effects of leadership training on employee-perceived leadership behavior, which we see as an important measure for visible leadership behavior and for organizational performance. Below, we will discuss effects on perceived leadership and on organizational performance, as well as limitations and importance for practical leadership training.

Perceived Leadership Effects

The most conclusive evidence of this study is that leadership training affects employee-perceived leadership. This is important as research has shown that employee-perceived leadership is much more strongly correlated with organizational performance than self-rated leadership (Favero and Bullock 2014; Jacobsen and Andersen 2015) as well as with desired

outcomes such as cooperation, satisfaction, and work quality (Oberfield 2014). Our results therefore give important insights on how leadership training can be designed and implemented to improve organizations.

However, some of the results are surprising. They indicated that leadership training can affect transformational leadership behavior and use of pecuniary rewards, while other aspects of leadership, particularly use of verbal rewards, seemed to be less affected by training. Against our expectations, transformational leadership training triggered the smallest and most uncertain increase in transformational leadership behavior (only significant for the LDV balanced panel model), whereas we saw positive effects of transactional training on transformational leadership behavior. The latter may reflect the shared goal orientation of the training programs. Attention to operational goals related to transactional leadership may have drawn attention to organizational visions related with transformational leadership. The combined training program produced the largest increase in transformational leadership, which aligns with existing findings about the complementarity of transformational and transactional leadership (Bass et al. 2003; Bass and Riggio 2006; Hargis, Watt, and Piotrowski 2011; O'Shea et al. 2009; Rowold 2011).

All three training programs increased the use of pecuniary rewards, which was a surprising result for transformational training (although only significant at a ten percent level when correcting for multiple hypothesis testing). The results are clearly strongest for transactional training, but they indicate that transformational training can also trigger use of transactions. Again, we can speculate that the reason is the shared goal-oriented foundation in the training programs, which means that leaders who learn to be more explicit about goals in their organization also become more aware of following up on goals and of rewarding employees correspondingly.

Although the results for verbal rewards were positive across training programs, we only found significant effects for the balanced panel model in the combined training program. This suggests weak effects that are too small to identify statistically. We encourage future studies to investigate the potential for improving the use of contingent verbal rewards further because the literature generally shows that verbal rewards are relevant for employee motivation (Andersen, Boye, and Laursen 2018).

The tendency to affect a type of leadership behavior that has not been part of the training program challenges the traditional distinction between transactional and transformational leadership. These results are in line with research that attenuates the augmentation hypothesis that transformational and transactional leadership by no means are substitutes but rather complementary and mutually supporting (Bass and Avolio 1993; Oberfield 2014). Although recent studies have taken important steps in developing revised models of transformational and transactional leadership (e.g., Jensen et al. 2019), we need better understanding and theoretical development of how leadership strategies function in combination to produce results.

Performance Effects

In terms of performance effects, the analyses show mixed support of the expected positive effects. We find indications that leadership training has positive effects for the broadest measure of organizational effectiveness (standardized test scores) – especially for leaders who completed the training. However, the results also indicate possible negative effects of transactional training on GPA. This may be because more students take the exams, but it highlights that measuring performance effects is a complex task and that leadership training can have positive effects on some performance measures and negative effects on others.

Although the results generally show positive tendencies in terms of performance, it is important to pay attention to potentially heterogeneous effects of leadership training across performance criteria. This emphasizes the relevance of being explicit about the choice of performance criteria (as suggested by Andersen, Heinesen, and Pedersen 2016) in terms of conducting robust research and advising practice. Creating performance through leadership is a complex process in which a multitude of factors, internal and external to the organizations, can offset or attenuate developments in performance.

Study Limitations and future research

Despite the strong field-experimental design, the study has limitations. The article illustrates some of the challenges and potentials connected with large research projects. Currently six other published articles from the LEAP project report treatment effects on survey measured outcomes such as value congruence and public service motivation⁸. This necessitates control for multiple hypothesis testing in this study, which is why we also report adjusted p-values (FDR q-values, table A10). Although significance levels are smaller, most effects of transformational and combination leadership training remain statistically significant. These results are furthermore robust to the different definitions of test families (see Table A10 in the online appendix). In future projects, we encourage scholars to preregister hypotheses, to be selective about publication of results from large projects, and to adjust analyses for multiple hypothesis testing when it is relevant.

Another limitation is associated with the inability to observe leadership behavior directly. Like most other studies in the field, we have investigated leadership behavior through employee perceptions. Employee ratings are more accurate assessments of leadership behavior than leader ratings, which are known to be inflated and inaccurate (Jacobsen and

Andersen 2015), but it would be relevant for future studies to investigate leadership behavior more directly or include other related outcomes of leadership training.

Another limitation is that we were only able to obtain performance information from primary and lower secondary schools. Different organizations respond differently to leadership, and as performance is multidimensional, performance effects of leadership training are complex. We therefore urge future studies to look into other and broader varieties of organizations and performance dimensions and measures.

Like most training programs, our programs had attrition. Although 75 percent of the leaders remained in the project, differences in initial levels of leadership behavior between the remaining leaders in the transformational program and the other remaining leaders suggest asymmetrical attrition. This could have been problematic for internal validity, but we focus on changes within individual leaders, and we conduct robustness tests that include the pre-training levels in comparison with the control group. Still, we encourage future studies to pay attention to attrition and report it openly as it can be important for assessing the broader effects of leadership training programs.

Implications for Practical Leadership Training

This study provides important points for practical leadership training. First, it shows that leadership training can actually make a difference for leaders. Second, it suggests that the combined program is most effective; implying that leadership training programs should take into account that leadership requires a varied set of tools, which training programs should be designed to equip and sharpen. Third, the training program is based on an action learning design in which leaders are taught about leadership and given opportunities to reflect, exercise, and receive feedback from peers and teachers. In line with recent leadership training

research (Holten, Bøllingtoft, and Wilms 2015), our study suggests that this is an important component in the positive findings.

The evaluated training programs are among the most intense experimental leadership development programs in the leadership literature, but some leaders participate in longer and more intense real-world programs. We expect that more comprehensive leadership training could produce even stronger effects than those identified here.

Conclusion

This study has investigated the behavioral and performance effects of a field-experimental leadership training program in search of more solid answers to one of the classic discussions about leadership: Are leaders born or can they be made? A key finding of the study is that training in transformational leadership, transactional leadership, or a combination can increase employee-perceived leadership compared to a control group. The results further suggest that a combination of transformational and transactional leadership training is most effective. Furthermore, although the effects should not be exaggerated, we find indications that training is associated with increased performance. Thus, the findings support that leadership training can be worthwhile.

The investigated training programs have emphasized action learning and skill building, but we still need more research on design of effective leadership training programs and implementation of training programs into practical useful approaches in organizations.

Perhaps the most important take-away point for future research is that the effects of leadership training are not simple. We strongly urge practitioners and future studies to pay close attention to different performance criteria and to consider possible combined effects of different types of leadership behaviors.

Notes

¹ Appendices are available online.

² For a full description of the responses in waves, see the data report [available on website].

³ Table A7 shows an overview of the sample selection.

⁴ The learning goals are nationally set, and the written exams are identical in all schools. Oral exams are assessed by the student's teacher and an external censor, while written exams are only marked externally. The results are thus comparable across schools. Both high GPAs and high completion rates from the final exams are official goals in Danish schools.

⁵ The national standardized tests are IT-based, adaptive, and self-scoring. Students' performance in these tests are also part of the official goals. For more information on the tests, see Beuchert and Nandrup (2018), who confirm the predictive validity of the test scores.

⁶ We do not have performance information on all schools in the experiment. Although we have no reason to believe that the non-existence of performance data is due to selection based on treatment, including the covariates constitutes a simple test of whether selection did occur between treatment and control groups.

⁷ We chose not to present a differences-in-differences (DiD) model as we are not able to test the common trend assumption (it appeared not to hold for the schools in the performance results), and as mentioned in Angrist and Pischke (2009), the lagged dependent model will tend to underestimate positive treatment effects, resulting in a more conservative estimate. However, results of a DiD model did not alter the conclusions substantially (as expected due to the randomization). The results are available from the authors upon request.

⁸ The leadership experiment has been investigated in six published studies: An, Meier, Bøllingtoft, and Andersen 2019; An, Meier, Ladenburg, and Westergård-Nielsen 2020;

Jensen 2018; Jensen, Andersen, and Jacobsen, 2019; Jacobsen and Andersen 2019; Bro & Jensen 2020.

References

- Allio, Robert J. 2005. Leadership development: Teaching versus learning. *Management Decision* 43(7/8): 1071–77.
- An, Seung-Ho, Kenneth J. Meier, Anne Bøllingtoft, and Lotte B. Andersen. 2019. Employee Perceived Effect of Leadership Training: Comparing Public and Private Organizations. *International Public Management Journal* 22(1): 2-28.
- An, Seung-Ho, Kenneth J. Meier, Jacob Ladenburg & Niels Westergård-Nielsen. 2020. Leadership and Job Satisfaction: Addressing Endogeneity With Panel Data From a Field Experiment, *Review of Public Personnel Administration* 40(4): 589-612.
- Andersen, Lotte B., and Thomas Pallesen. 2008. “Not Just for the Money?” How Financial Incentives Affect the Number of Publications at Danish Research Institutions. *International Public Management Journal* 11(1): 28-47.
- Andersen, Lotte B., Stefan Boye, and Ronni Laursen. 2018. Building Support? The Importance of Verbal Rewards for Employee Perceptions of Governance Initiatives. *International Public Management Journal* 21(1): 1–32.
- Andersen, Lotte B., Eskil Heinesen, and Lene H. Pedersen. 2016. How Does Public Service Motivation among Teachers Affect Student Performance in Schools? *Journal of Public Administration Research and Theory* 24(3): 651–71.

- Anderson, Michael L. 2008. Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of American Statistical Society* 103(484): 1481-1495.
- Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.
- Avolio, Bruce J., Rebecca J. Reichard, Sean T. Hannah, Fred O. Walumbwa, and Adrian Chan. 2009. A Meta-analytic Review of Leadership Impact Research: Experimental and Quasi-experimental Studies. *The Leadership Quarterly* 20(5): 764–84.
- Barling, Julian, Tom Weber, and E. Kevin Kelloway. 1996. Effects of Transformational Leadership Training on Attitudinal and Financial Outcomes: A Field Experiment. *Journal of Applied Psychology* 81(6): 827–32.
- Bass, Bernard M. 1985. *Leadership and Performance beyond Expectations*. New York: The Free Press.
- . 1990. From Transactional to Transformational Leadership: Learning to Share the Vision. *Organizational Dynamics* 18(3): 19–31.
- Bass, Bernard, and Bruce J. Avolio. 1993. Transformational Leadership and Organizational Culture. *Public Administration Quarterly* 17(1): 112–21.
- Bass, Bernard M., and Ronald E. Riggio. 2006. *Transformational Leadership* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bass, Bernard M., Bruce J. Avolio, Dong I. Jung, and Yair Berson. 2003. Predicting Unit Performance by Assessing Transformational and Transactional Leadership. *Journal of Applied Psychology* 88(2): 207–18.
- Beer, Michael, Magnus Finnström, and Derek Schrader. 2016. Why Leadership Training Fails—and What To Do about It. *Harvard Business Review* 94(19): 50–7.

- Bellé, Nicola. 2014. Leading to make a Difference: A Field Experiment on the Performance Effects of Transformational Leadership, Perceived Social Impact, and Public Service Motivation. *Journal of Public Administration Research and Theory* 24(1): 109–36.
- . 2015. Performance-related Pay and the Crowding Out of Motivation in the Public Sector: A Randomized Field Experiment. *Public Administration Review* 75(2): 230–41.
- Beuchert, Louise V., and Anne B. Nandrup. 2018. The Danish National Tests at a Glance. *Danish Economic Journal* 145(1): 1-37.
- Boyne, George A. 2003. Sources of Public Service Improvement: A Critical Review and Research Agenda. *Journal of Public Administration Research and Theory* 13(3): 367–94.
- Bro, Louise, and Ulrik T. Jensen. 2020. Does Transformational Leadership Stimulate User Orientation? Evidence from a Field Experiment. *Public Administration*. 98(1): 177–193.
- Burgoyne, John, Wendy Hirsh, and Sadie Williams. 2004. *The Development of Management and Leadership Capability and Its Contribution to Performance: The Evidence, the Prospects and the Research Need*. RR560, London: Department for Education and Skills.
- Day, David V. 2013. Training and Developing Leaders: Theory and Research. In *The Oxford Handbook of Leadership*, edited by Michael G. Rumsey, 76–93. Oxford: Oxford University Press.
- Dee, Thomas S., and James Wyckoff. 2015. Incentives, Selection, and Teacher Performance: Evidence from IMPACT. *Journal of Policy Analysis and Management* 34(2): 267–97.
- Doh, Jonathan P. 2003. Can Leadership Be Taught? Perspectives from Management Educators. *Academy of Management Learning and Education* 2(1): 54–67.

- Dvir, Taly, Dov Eden, Bruce J. Avolio, and Boas Shamir. 2002. Impact of Transformational Leadership on Follower Development and Performance: A Field Experiment. *The Academy of Management Journal* 45(4): 735–44.
- Favero, Nathan, and Justin Bullock. 2014. How (Not) to Solve the Problem: An Evaluation of Scholarly Responses to Common Source Bias. *Journal of Public Administration Research and Theory* 25(1): 285-308.
- Fernandez, Sergio. 2005. Developing and Testing an Integrative Framework of Public Sector Leadership: Evidence from the Public Education Arena. *Journal of Public Administration Research and Theory* 15(2): 197–217.
- Hargis, Michael B., John D. Watt, and Chris Piotrowski. 2011. Developing Leaders: Examining the Role of Transactional and Transformational Leadership across Business Contexts. *Organization Development Journal* 29(3): 51–66.
- Hassan, Rasool A., Bashir A. Fuwad, and Azam I. Rauf. 2010. Pre-training Motivation and the Effectiveness of Transformational Leadership Training: An Experiment. *Academy of Strategic Management Journal* 9(1): 1–8.
- Holten, Ann-Louise, Anne Bøllingtoft, and Inge Wilms. 2015. Leadership in a Changing World: Developing Managers through a Teaching and Learning Programme. *Management Decision* 53(5): 1107–24.
- House, Robert J. 1998. Appendix: Measures and Assessments for the Charismatic Leadership Approach: Scales, Latent Constructs Loadings, Cronbach Alphas, Interclass Correlations. In *Leadership: The Multiple-level Approaches: Contemporary and Alternative*, edited by Fred Dansereau and Francis J. Yammarino, 23–30. London: JAI Press.

- Jacobsen, Christian B., and Lotte B. Andersen. 2015. Is Leadership in the Eye of the Beholder? A Study of Intended and Perceived Leadership Practices and Organizational Performance. *Public Administration Review* 75(6): 829–41.
- Jacobsen, Christian B., and Lotte B. Andersen. 2019. High Performance Expectations: Concept and Causes, *International Journal of Public Administration* 42(2): 108-118.
- Jensen, Ulrich T. 2018. Does Perceived Societal Impact Moderate the Effect of Transformational Leadership on Value Congruence? Evidence from a Field Experiment. *Public Administration Review* 78(1): 48–57.
- Jensen, Ulrich T., Lotte B. Andersen, and Christian B. Jacobsen. 2019. Only When We Agree! How Value Congruence Moderates the Impact of Goal-Oriented Leadership on Public Service Motivation. *Public Administration Review* 79(1): 12-24.
- Jensen, Ulrich T., Donald P. Moynihan, and Heidi H. Salomonsen. 2018. Communicating the Vision: How Face-to-Face Dialogue Facilitates Transformational Leadership. *Public Administration Review* 78(3): 350–61.
- Jensen, Ulrich T., Lotte B. Andersen, Louise L. Bro, Anne Bøllingtoft, Tine M. Eriksen, Ann-Louise Holten, Christian B. Jacobsen, Jacob Ladenburg, Poul A. Nielsen, Heidi H. Salomonsen, Niels Westergård-Nielsen, and Allan Würtz. 2019. Conceptualizing and Measuring Transformational and Transactional Leadership. *Administration & Society* 51(1): 3–33.
- Judge, Timothy A., and Ronald F. Piccolo. 2004. Transformational and Transactional Leadership: A Meta-analytic Test of Their Relative Validity. *Journal of Applied Psychology* 89(5): 755–68.
- Jung, Chan S. 2011. Why Are Goals Important in the Public Sector? Exploring the Benefits of Goal Clarity for Reducing Turnover Intention. *Journal of Public Administration Research and Theory* 24(1): 209–34.

- Jung, Dong I., and Bruce J. Avolio. 2000. Opening the Black Box: An Experimental Investigation of the Mediating Effects of Trust and Value Congruence on Transformational and Transactional Leadership. *Journal of Organizational Behavior* 21(8): 949–64.
- Kelloway, E. Kevin, Julian Barling, and Jane Helleur. 2000. Enhancing Transformational Leadership: The Roles of Training and Feedback. *Leadership & Organization Development Journal* 21(3): 145–49.
- Kirkpatrick, D. L. 1979. “Techniques for evaluating training programs.” *Journal of American Society for Training and Development* 33(6): 78–92.
- Kroll, Alexander, and Donald P. Moynihan. 2015. Does Training Matter? Evidence from Performance Management Reforms. *Public Administration Review* 75(3): 411–20.
- Kruger, Justin, and David Dunning. 1999. Unskilled and Unaware of It: How Difficulties in Recognizing One’s Own Incompetence Lead to Inflated Self-Assessments. *Journal of Personality and Social Psychology* 77(6): 1121–34.
- Lavy, Victor. 2009. Performance Pay and Teachers’ Effort, Productivity, and Grading Ethics. *The American Economic Review* 99(5): 1979–2011.
- . 2020. Teachers’ Pay for Performance in the Long-run: The Dynamic Pattern of Treatment Effects on Students’ Educational and Labor Market Outcomes in Adulthood. *The Review of Economic Studies*, 87(5) 2322–2355.
- Lazear, Edward P. 1999. Performance Pay and Productivity. *American Economic Review* 90 (5): 1346-61.
- Lowe, Kevin B., K. Galen Kroeck, and Nagaraj Sivasubramaniam. 1996. Effectiveness Correlates of Transformational and Transactional Leadership: A Meta-analytic Review of the MLQ Literature. *The Leadership Quarterly* 7(3): 385–425.

- Moynihan, Donald P., Sanjay K. Pandey, and Bradley E. Wright. 2012. Setting the Table: How Transformational Leadership Fosters Performance Information Use. *Journal of Public Administration Research and Theory* 22(1): 143–64.
- Oberfield, Zachary W. 2014. Public Management in Time: A Longitudinal Examination of the Full Range of Leadership Theory. *Journal of Public Administration Research and Theory* 24(2): 407–29.
- O’Shea, Patric G., Roseanne J. Foti, Neil M. A. Hauenstein, and Peter Bycio. 2009. Are the Best Leaders both Transformational and Transactional? A Pattern-oriented Analysis. *Leadership* 5(2): 237–59.
- O’Toole, Laurence J., and Kenneth J. Meier. 2011. *Public Management: Organizations, Governance, and Performance*. Cambridge: Cambridge University Press.
- Parry, Ken W., and Paresha N. Sinha. 2005. Researching the Trainability of Transformational Organizational Leadership. *Human Resource Development International* 8(2): 165–83.
- Rainey, Hal. 2009. *Understanding and Managing Public Organizations*. San Francisco: Jossey-Bass.
- Rosenthal, R. (1979). The "File Drawer Problem" and Tolerance for Null Results, *Psychological Bulletin*, 86(3): 638-641
- Rowold, Jens. 2011. Relationship between Leadership Behaviors and Performance. The Moderating Role of a Work Team’s Level of Age, Gender, and Cultural Heterogeneity. *Leadership & Organization Development Journal* 32(6): 628–47.
- Seidle, Brett, Sergio Fernandez, and James L. Perry. 2016. Do Leadership Training and Development Make a Difference in the Public Sector? A Panel Study. *Public Administration Review* 76(4): 603-13.

- Trottier, Tracey, Montgomery Van Wart, and XiaoHu Wang. 2008. Examining the Nature and Significance of Leadership in Government Organizations. *Public Administration Review* 68(2): 319–33.
- Van Knippenberg, Daan, and Sim B. Sitkin. 2013. A Critical Assessment of charismatic-Transformational Leadership Research: Back to the Drawing Board? *Academy of Management Perspectives* 7(1): 1–60.
- Van Wart, Montgomery. 2013. Administrative Leadership Theory: A Reassessment after 10 Years. *Public Administration* 91(3): 521–43.
- Waldman, David A., Bernard M. Bass, and Francis J. Yammarino. 1990. Adding to Contingent-Reward Behavior: The Augmenting Effect of Charismatic Leadership. *Group & Organization Management* 15(4): 381–94.
- Walker, Richard M., and Rhys Andrews. 2013. Local Government Management and Performance: A Review of Evidence. *Journal of Public Administration Research and Theory* 25(1): 101–33.
- Wang, Gang, In-Sue Oh, Stephen H. Courtright, and Amy E. Colbert. 2011. Transformational Leadership and Performance across Criteria and Levels: A Meta-analytic Review of 25 Years of Research. *Group & Organization Management* 36(2): 223–70.
- Wilson, James Q. (1989). *Bureaucracy: What Government Agencies Do and Why They Do It*. City: Basic Books.
- Wright, Bradley E. 2007. Public Service and Motivation: Does Mission Matter? *Public Administration Review* 67(1): 54–64.
- Wright, Bradley E., and Sanjay K. Pandey. 2010. Transformational Leadership in the Public Sector: Does Structure Matter? *Journal of Public Administration Research and Theory* 20(1): 75–89.

Wright, Bradley E., Donald P. Moynihan, and Sanjay K. Pandey. 2012. Pulling the Levers: Transformational Leadership, Public Service Motivation, and Mission Valence. *Public Administration Review* 72(2): 206–15.

Figure 1. Theoretical Model of Leadership Training



Table 1. Status for Invitations, Replies, Sign-ups, and Completion Across Sectors

Area	Criteria for invitation	Invited	Replies (% of invited)	Signed up (% of replies)	Completed (% of signups)
1) Upper Secondary schools	All principals*	308	186 (60.4%)	41 (22.0%)	37 (90.2%)
2) Public primary and lower sec. schools	All principals*	787	344 (43.7%)	119 (34.6%)	83 (69.7%)
3) Private primary and lower sec. schools	All principals*	278	134 (48.2%)	44 (32.8%)	31 (70.5%)
4) Daycare, type 1	Pedagogues who lead pedagogue leaders of employees	416	191 (45.9%)	84 (44.0%)	61 (72.6%)
5) Daycare, type 2	Pedagogue leaders of employees led by pedagogue leaders**	1,689	386 (22.9%)	50 (13.0%)	37 (74.0%)
6) Daycare, type 3	Pedagogue leaders of employees led by non-pedagogue leaders**	1,092	261 (23.9%)	83 (31.8%)	60 (72.3%)
7) Daycare, private	Leaders of daycare centers with private ownership**	394	154 (39.1%)	62 (40.3%)	40 (64.5%)
8) Tax	Selected by Tax	153	150 (98.0%)	144 (96.0%)	130 (90.3%)
9) Banks	Selected by two banks	51	47 (92.2%)	45 (95.7%)	27 (60.0%)
Total		5,168	1,853 (35.9%)	673 (36.3%)	506 (75.3%)

* Head of school if school is divided into independent school units.

** Leaders of daycare centers with 3- to 6-year-olds or 0- to 6-year-olds were invited. Leaders were not invited if their own leader was part of the project.

Table 2. Overview of Effects of Leadership Training on Leadership Behavior and Performance (Final Exams and Standardized Tests)

<i>Leadership training</i>	<i>Leadership behavior</i>			<i>Performance (only schools)</i>			
	Transformational leadership behavior	Contingent verbal rewards	Contingent pecuniary rewards	Final exams (student GPA)	Final exams (completion)	Standardized tests (student scores)	Standardized tests (completion rate)
Transformational	(Positive, but only for panel and less robust). H1	No consistent effect.	Positive (in panel).	No consistent effect. H7	No consistent effect.	Higher (conditional on completion). H7	No consistent effect.
Transactional	(Positive, but not robust).	No consistent effect. H2	Positive. H3	(Tendency to lower GPA, but not robust). H8	(Tendency to higher, but not robust).	Higher (conditional on completion). H8	No consistent effect.
Combination	Positive. H4	(Positive, but not robust). H5	Positive, but less robust for panel. H6	No consistent effect. H9	No consistent effect.	(Tendency to higher if completion, but not robust). H9	No consistent effect.
Detailed results in:	Tables 3 and 4			Tables 5, 6 and 8		Tables 5, 7 and 9	

Note: Results in parentheses are only statistically significant at $p < 0.1$ and/or are less consistent. If the results are not robust after the FDR correction, it is also noted in the cells (that are classified according to hypothesis H1-H9, if any of them concerns the relevant effect).

Table 3. Means of Perceived Leadership Behavior by Treatment

Leadership Training								
	All employees				Panel (only employees who remain in the organization)			
	Transformational	Transactional	Combination	Control	Transformational	Transactional	Combination	Control
<i>Perceived Leadership Behavior</i>								
<i>Transformational Leadership Behavior</i>								
Pre	68.916	70.293	69.272	69.974	68.649	71.376	72.566	73.388
Post	67.907	72.546	73.506	69.269	69.331	71.075	73.721	69.924
Difference	-1.009	2.253***	4.234***	-0.705	0.682	-0.301	1.155	-3.464***
Num. Obs.	106	113	111	128	103	109	110	124
<i>Contingent Verbal Rewards</i>								
Pre	64.462	66.956	65.662	66.456	65.662	68.687	66.933	69.298
Post	64.399	68.024	67.024	65.909	64.891	67.774	67.373	66.252
Difference	-0.063	1.068	1.362	-0.547	-0.771	-0.913	0.440	-3.046
Num. Obs.	107	113	111	128	103	109	110	124

Contingent Pecuniary Rewards

Pre	35.866	37.018	37.063	35.746	37.897	35.451	38.139	38.000
Post	37.011	39.837	38.264	35.042	36.680	39.140	36.784	34.112
Difference	1.145	2.819**	1.201	-0.704	-1.217	3.689	-1.355	-3.888
Num. Obs.	107	113	111	128	103	109	110	124

Differences are tested for significance using t-tests conditioning on sector fixed effects. “All” refers to average leadership values based on all employees who responded in either wave 0 or wave 1. “Panel” refers to average leadership values based on employees who responded in both wave 0 and wave 1. We flag missing values of the conditioning set.

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 4. Average Training Effects of Leadership Training Programs on Employee Perceived Leadership

	Employee-perceived leadership								
	Transformational leadership behavior			Contingent verbal rewards			Contingent pecuniary rewards		
	OLS	LDV	LDV	OLS	LDV	LDV	OLS	LDV	LDV
	All	All	Panel	All	All	Panel	All	All	Panel
<i>Leadership training</i>									
Transformational	-1.315	0.889	2.814	-1.724	0.949	1.511	2.144	3.120	3.348
	(0.434)	(0.535)	(0.045)	(0.390)	(0.520)	(0.285)	(0.133)	(0.018)	(0.026)
	[0.461]	[0.530]	[0.142]	[0.459]	[0.527]	[0.376]	[0.247]	[0.083]	[0.095]
Transactional	2.993	3.285	2.132	1.859	2.069	1.710	4.444	4.140	6.096
	(0.047)	(0.010)	(0.077)	(0.285)	(0.112)	(0.190)	(0.004)	(0.003)	(0.000)
	[0.144]	[0.074]	[0.193]	[0.376]	[0.229]	[0.312]	[0.050]	[0.050]	[0.011]
Combination	4.540	4.635	3.470	1.655	2.032	2.424	3.770	3.509	2.982
	(0.004)	(0.000)	(0.004)	(0.409)	(0.155)	(0.063)	(0.008)	(0.005)	(0.038)
	[0.050]	[0.012]	[0.051]	[0.459]	[0.272]	[0.172]	[0.067]	[0.051]	[0.125]

Number of observations	462	458	446	463	459	446	463	459	446
Adjusted R-squared	0.139	0.442	0.493	0.098	0.533	0.642	0.337	0.482	0.430
Sector fixed effects	X	X	X	X	X	X	X	X	X
Covariates		X	X		X	X		X	X
Lagged dependent variable		X	X		X	X		X	X

Note: OLS and LDV regressions with p-values (in parentheses) and FDR q-values [in brackets]. “All” refers to average leadership values based on all employees who responded in either wave 0 or wave 1. “Panel” refers to average leadership values based on employees who responded in both wave 0 and wave 1. We flag missing values of the conditioning set.

Table 5. Means of School Performance by Treatment

	Leadership Training			
	Transformational	Transactional	Combination	Control
Final Exam				
<i>GPA</i>				
Time 0	0.027	-0.058	-0.024	0.051
Time 1	0.034	-0.082	0.040	0.063
Difference	0.007	-0.024	0.064	0.012
Number of observations	34	34	36	34
<i>Completion rate</i>				
Time 0	0.955	0.905	0.941	0.914
Time 1	0.925	0.942	0.915	0.897
Difference	-0.030	0.037	-0.026	-0.017
Number of observations	35	34	37	35
Standardized Tests				
<i>Student Test Score</i>				
Time 0	-0.167	-0.059	0.009	0.040
Time 1	0.000	-0.077	0.018	-0.011
Difference	0.167	-0.018	0.009	-0.051

Number of observations	29	27	27	29
------------------------	----	----	----	----

Completion rate

Time 0	0.939	0.933	0.938	0.938
--------	-------	-------	-------	-------

Time 1	0.827	0.827	0.835	0.829
--------	-------	-------	-------	-------

Difference	-0.112***	-0.106***	-0.103***	-0.109***
------------	-----------	-----------	-----------	-----------

Number of observations	29	27	27	29
------------------------	----	----	----	----

Differences are tested for significance using t-tests conditioning on sector fixed effects. *** p <

0.01; ** p < 0.05; * p < 0.1.

Table 6. Average Training Effects of Leadership Training Programs on School Performance (GPA and Completion Rate), All Schools

	Final Exams					
	GPA			Completion rate		
	OLS	OLS	LDV	OLS	OLS	LDV
Leadership training						
Transformational	-0.034	-0.070	-0.043	0.029	-0.007	-0.030
	(0.660)	(0.291)	(0.503)	(0.491)	(0.858)	(0.419)
			[0.515]			[0.459]
Transactional	-0.150	-0.131	-0.105	0.045	0.040	0.037
	(0.060)	(0.036)	(0.077)	(0.162)	(0.227)	(0.110)
			[0.193]			[0.229]
Combination	-0.014	0.010	0.006	0.018	0.001	-0.007
	(0.874)	(0.893)	(0.934)	(0.661)	(0.983)	(0.871)
			[0.800]			[0.757]
Number of observations	138	138	138	141	141	141
Adjusted R-squared	0.075	0.380	0.427	-0.016	0.140	0.315
Sector fixed effects	X	X	X	X	X	X
Covariates		X	X		X	X
Lagged dependent variable			X			X

Note: The table shows results of the mandatory exams in the 9th-grade exit exam. OLS with p-values (in parentheses) and FDR q-values [in brackets]. We flag missing values of the conditioning set.

Table 7. Average Training Effects of Leadership Training Programs on School Performance (Standardized Test Scores and Completion Rates), All Schools

	Standardized Tests					
	Student test scores			Completion rate		
	OLS	OLS	LDV	OLS	OLS	LDV
Leadership training						
Transformational	0.012	0.019	0.050	-0.002	-0.001	-0.003
	(0.854)	(0.618)	(0.098)	(0.887)	(0.928)	(0.786)
			[0.216]			[0.727]
Transactional	-0.066	0.048	0.076	-0.002	0.004	0.003
	(0.393)	(0.267)	(0.098)	(0.887)	(0.740)	(0.803)
			[0.172]			[0.727]
Combination	0.029	0.001	0.017	0.005	0.012	0.012
	(0.721)	(0.984)	(0.664)	(0.722)	(0.278)	(0.319)
			[0.652]			[0.399]
Number of observations	112	112	112	112	112	112
Adjusted R-squared	-0.011	0.720	0.800	-0.024	0.251	0.282
Covariates		X	X		X	X
Lagged dep. variable			X			X

Note: The table shows results of the standardized tests conducted in grades 2 to 8 in reading, mathematics, English, physics and chemistry, biology, and geography. OLS with p-values (in parentheses) and FDR q-values [in brackets]. We flag missing values of the conditioning set.

Table 8. Average Training Effects of Leadership Training Programs on School Performance (GPA and Completion Rate), Schools Where Principal Completed Training

	Final Exam					
	GPA			Completion rate		
	OLS	OLS	LDV	OLS	OLS	LDV
Leadership training						
Transformational	-0.079	-0.121	-0.105	0.003	-0.015	-0.012
	(0.326)	(0.158)	(0.214)	(0.943)	(0.740)	(0.795)
			[0.335]			[0.727]
Transactional	-0.156	-0.056	-0.055	0.032	0.022	0.022
	(0.038)	(0.426)	(0.414)	(0.149)	(0.519)	(0.480)
			[0.459]			[0.492]
Combination	-0.077	-0.031	-0.043	-0.012	-0.018	-0.028
	(0.405)	(0.752)	(0.665)	(0.770)	(0.761)	(0.636)
			[0.652]			[0.642]
Number of observations	100	100	100	102	102	102
Adjusted R-squared	0.128	0.587	0.583	-0.028	0.542	0.568
Sector fixed effects	X	X	X	X	X	X
Covariates		X	X		X	X
Lagged dependent variable			X			X

Note: The table shows results of the mandatory exams in the 9th-grade exit exam for the leaders who completed the leadership training. OLS with p-values (in parentheses) and FDR q-values [in brackets]. We flag missing values of the conditioning set.

Table 9. Average Training Effects of Leadership Training Programs on School Performance (Standardized Test Scores and Completion Rate). Schools Where Principal Completed Training

	Standardized Tests					
	Student test scores			Completion rate		
	OLS	OLS	LDV	OLS	OLS	LDV
Leadership training						
Transformational	0.086	0.063	0.071	0.010	0.000	-0.005
	(0.203)	(0.068)	(0.017)	(0.534)	(1.000)	(0.702)
			[0.083]			[0.699]
Transactional	-0.016	0.081	0.097	0.003	0.005	0.002
	(0.842)	(0.063)	(0.013)	(0.852)	(0.755)	(0.887)
			[0.074]			[0.757]
Combination	0.084	0.080	0.073	0.012	0.016	0.013
	(0.353)	(0.073)	(0.093)	(0.455)	(0.222)	(0.356)
			[0.213]			[0.437]
Number of observations	80	80	80	80	80	80
Adjusted R-squared	-0.005	0.817	0.845	-0.027	0.352	0.392
Covariates		X	X		X	X
Lagged dependent variable			X			X

Note: The table shows results of the standardized tests conducted in grades 2 to 8 in reading, mathematics, English, physics and chemistry, biology, and geography for the leaders who

completed the training. OLS with standard errors (in parentheses). We flag missing values of the conditioning set.

Leadership training. Leadership Behavior and Organizational Effectiveness: Evidence from a Randomized Field Experiment on Transformational and Transactional Leadership

Online Appendix

Table A1. List of items in questionnaire measuring transformational and transactional leadership

Transformational leadership behavior – My leader ...		
1	Concretizes a clear vision for the [ORGANIZATION TYPES] future	Modified from Moynihan, Pandey, and Wright 2012
2	Seeks to make employees accept common goals for the [ORGANIZATION TYPE]	Modified from MacKenzie et al., 2001
3	Strives to get the [ORGANIZATION TYPE] to work together in the direction of the vision	Modified from Podsakoff et al., 1996
4	Strives to clarify for the employees how they can contribute to achieving the [ORGANIZATION TYPES] goals	Own
Transactional leadership behavior		
Contingent pecuniary rewards – My leader...		
1	Rewards the employees' performance when they live up to his/her requirements	Modified from Jacobsen and Andersen, 2015
2	Rewards the employees dependent on how well they perform their jobs	Jacobsen and Andersen, 2015
3	Lets employees' effort determine received rewards	Modified from Rainey 2009
Contingent verbal rewards – My leader...		
1	Gives individual employees positive feedback when they perform well	Modified from House, 1998
2	Actively shows his/her appreciation of employees who do their jobs better than expected	Modified from House, 1998
3	Generally does not acknowledge individual employees even though they perform as required (R)	Modified from House, 1998

Table A2. Confirmatory factor analysis of transformational and transactional leadership (pre- and post-training surveys)

Transformational leadership behavior – My leader ...		Pre: Survey before training	Post: Survey after training
1	Concretizes a clear vision for the [ORGANIZATION TYPES] future	0.791	0.796
2	Seeks to make employees accept common goals for the [ORGANIZATION TYPE]	0.874	0.878
3	Strives to get the [ORGANIZATION TYPE] to work together in the direction of the vision	0.764	0.775
4	Strives to clarify for the employees how they can contribute to achieving the [ORGANIZATION TYPES] goals	0.873	0.883
Transactional leadership behavior			
Contingent pecuniary rewards – My leader...			
1	Rewards the employees' performance when they live up to his/her requirements	0.913	0.916
2	Rewards the employees dependent on how well they perform their jobs	0.897	0.899
3	Lets employees' effort determine received rewards	0.932	0.931
Contingent verbal rewards – My leader...			
1	Gives individual employees positive feedback when they perform well	0.902	0.909
2	Actively shows his/her appreciation of employees who do their jobs better than expected	0.886	0.877
3	Generally does not acknowledge individual employees even though they perform as required (R)	0.744	0.743

Note: RMSEA: 0.038 (pre) 0.037 (post). CFI: 0.963 (pre). 0.962 (post). TLI: 0.948 (pre). 0.946 (post). SRMR: 0.031 (pre). 0.034 (post). Scores are aggregated within leaders based on satisfactory analyses of within-group agreement. r_{wg} scores in the 2014 survey range between 0.71 (verbal rewards) and 0.83 (transformational leadership). and in the 2015 survey between 0.73 (verbal rewards) and 0.83 (transformational leadership). which are satisfactory (recommended above 0.70. Kozlowski and Klein, 2000, p. 223). ICC(1) scores range between 0.20 (contingent pecuniary rewards) and 0.69 (transformational leadership). and between 0.23 (contingent

pecuniary rewards) and 0.60 (transformational leadership) in 2015. ICC(2) range between 0.67 (contingent sanctions) and 0.94 (transformational leadership), which are well beyond the suggested values for aggregation (Biemann et al., 2012). Finally, ANOVA and F-tests are significant ($p < 001$) in both survey rounds for all measures, which indicates sufficient variation between groups (Kozlowski and Klein, 2000).

Table A3. Balance table for initial leadership behavior and controls for organizations that entered the experiments

<i>Leadership Training</i>	Transformational		Transactional		Combination		Control		N
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	
Average employee job satisfaction	7.463	1.028	7.646	0.826	7.474	1.100	7.495	0.998	601
Average employee-perceived transform. leadership	68.92	12.21	70.29	11.66	69.27	13.67	69.97	12.16	601
Average employee-perceived use of verbal rewards	64.46	15.87	66.96	13.45	65.66	15.57	66.46	14.93	601
Average employee-perceived use of pecuniary rewards	35.87	11.42	37.02	10.87	37.06	12.21	35.75	12.66	601
Mean employee response rate	0.633	0.246	0.649*	0.232	0.652**	0.236	0.610	0.242	610
Leadership experience	6.512	5.865	5.633	5.267	5.433	5.654	5.907	5.748	604
Master in leadership	0.0621	0.242	0.0387	0.194	0.0649	0.247	0.0530	0.225	605
Male	0.503	0.502	0.391	0.490	0.526	0.501	0.483	0.501	610

Note: The table shows leadership characteristics by treatment status. Means are tested against the mean of the control group.

*** $p < 0.01$. ** $p < 0.05$. * $p < 0.1$.

Table A4. Balance table for initial leadership behavior and controls for organizations that completed the experiments

<i>Leadership Training</i>	Transformational		Transactional		Combination		Control		N
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	
Average employee job satisfaction	7.348	1.029	7.646	0.826	7.478	1.100	7.547	0.966	470
Average employee-perceived transform. leadership	67.91**	13.84	72.55	11.61	73.51	12.11	69.27	13.30	462
Average employee-perceived use of verbal rewards	64.40*	16.97	68.02	12.81	67.02	16.34	65.91	15.74	463
Average employee-perceived use of pecuniary rewards	37.01	13.25	39.84	13.59	38.26	12.10	35.04	14.41	463
Mean employee response rate	0.640	0.239	0.686	0.214	0.641	0.236	0.625	0.235	475
Leadership experience	6.527	5.442	5.251	4.967	5.931	5.999	6.144	5.944	471
Master in leadership	0.0463	0.211	0.0435	0.205	0.0609	0.240	0.0522	0.223	472
Male	0.518	0.502	0.371	0.485	0.487	0.502	0.463	0.500	475

Note: The table shows leadership characteristics by treatment status. Means are tested against the mean of the control group.

*** $p < 0.01$. ** $p < 0.05$. * $p < 0.1$.

Table A5. Balance table for initial leadership behavior and controls. *balanced panel* (response before *and* after training)

<i>Leadership Training</i>	Transformational training group		Transactional training group		Combination training group		Control group		N
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	
	Average employee job satisfaction	7.536	1.166	7.763	1.005	7.844	0.971	7.651	
Average employee-perceived transform. Leadership	68.65***	13.76	71.38	12.89	72.57	13.12	73.39	12.17	446
Average employee-perceived use of verbal rewards	65.66*	17.59	68.69	13.62	66.93	17.54	69.30	16.39	446
Average employee-perceived use of pecuniary rewards	37.90	13.17	35.45	14.64	38.14	13.62	38.00	14.92	446
Mean employee-response rate	6.387	5.347	5.210	4.934	5.978	6.054	6.214	6.053	442
Leadership experience	0.050	0.218	0.037	0.189	0.064	0.246	0.057	0.232	443
Master in leadership	0.505	0.502	0.391	0.490	0.495	0.502	0.476	0.501	446

Note: The table shows leadership characteristics by treatment status. Means are tested against the mean of the control.

*** $p < 0.01$. ** $p < 0.05$. * $p < 0.1$.

Table A6. Sample selection for leadership behavior analyses

	Employees		Leaders	
<i>Sample selection</i>	Wave 0	Wave 1	Wave 0	Wave 1
All leaders			673	
All employees	19.551	15.105	610(91%)	476(71%)
Responded	10.167(52%)	7.316(48%)	601(99%)	463(97%)
	Wave 0 & Wave 1		Wave 0 & Wave 1	
All employees	12.644		476	
Responded	4.325(34%)		446(94%)	

Table A7. Sample selection for leadership behavior analyses

	Schools/leaders	# Students
<i>Sample selection</i>		
<i>9th Exit exam</i>		
All schools	163	
Schools with performance information in 2014	153 (93%)	6.401
Schools with performance information in 2016	138 (90%)	5.903
<i>The Danish National Tests(standardized tests)</i>		
All schools	117	
Schools with performance information in 2014	116 (99%)	56.942
Schools with performance information in 2016	112 (97%)	48.872

Note: Three additional schools had information about completion of exit exam. meaning that 141 schools can be included in these analyses.

Table A8. Balance table. Initial leader and student-related characteristics. GPA analysis

<i>Leadership Training</i>	Transformational training group		Transactional training group		Combination training group		Control group		N
	Mean	Std. Dev.	Mean	Std.	Mean	Std.	Mean	Std. Dev.	
				Dev.		Dev.			
<i>Leader characteristics</i>									
Age	52.88	7.373	49.50	6.463	50.45	8.599	49.89	7.213	115
Male	0.316	0.471	0.351	0.484	0.179	0.389	0.333	0.478	153
No personal identification number	0.289	0.460	0.162	0.374	0.154	0.366	0.282	0.456	153
Number of Students	47.42	30.15	41.27	27.19	36.92	17.33	41.92	25.28	153
<i>Average student characteristics</i>									
Male	0.484	0.185	0.527	0.090	0.507	0.138	0.494	0.130	153
Age	16.20	0.185	16.19	0.118	16.20	0.155	16.19	0.167	153
Danish	0.889	0.147	0.875	0.223	0.817*	0.244	0.900	0.156	153
Immigrant	0.040	0.052	0.030	0.040	0.046	0.055	0.038	0.059	153
Immigrant descendent	0.070	0.138	0.095	0.197	0.137*	0.218	0.062	0.116	153
Special needs education (Physical)	0.021	0.047	0.027	0.046	0.022	0.049	0.017	0.028	153
Special needs education (Psychical)	0.031	0.061	0.035	0.055	0.033	0.052	0.036	0.066	153

Special needs education (Other)	0.058	0.162	0.071	0.082	0.068	0.089	0.054	0.090	153
No registered parents	0.007	0.022	0.009	0.020	0.006	0.016	0.010	0.026	153
Only father	0.007	0.012	0.003	0.010	0.008	0.015	0.004	0.015	153
Only mother	0.018	0.031	0.012	0.017	0.019	0.029	0.019	0.025	153
			-						
GPA 9th grade exit exam	0.012	0.340	0.072	0.323	-0.002	0.387	0.037	0.415	151
Completion rate	0.925	0.168	0.909	0.141	0.940	0.074	0.917	0.163	153
<i>Average mothers characteristics</i>									
Age	42.360	1.586	42.48	1.333	42.24	2.077	42.47	1.218	153
Single	0.332	0.150	0.306	0.117	0.343	0.112	0.317	0.109	153
Married	0.620	0.158	0.646	0.113	0.607	0.115	0.620	0.125	153
Cohabiting	0.05	0.042	0.047	0.042	0.050	0.047	0.063	0.061	153
Log-income	10.75	1.350	10.17	1.989	9.729*	1.711	10.44	1.810	153
Unemployed	0.033	0.044	0.03	0.036	0.062***	0.084	0.023	0.030	153
Employed	0.819	0.114	0.786	0.164	0.749**	0.142	0.818	0.133	153
Outside labor force	0.147	0.111	0.182	0.158	0.189	0.128	0.159	0.129	153
Low education	0.196	0.110	0.238	0.145	0.316	0.164	0.260	0.179	153
High education	0.360	0.165	0.346	0.151	0.314	0.184	0.371	0.205	153
Vocational education	0.444	0.157	0.416	0.146	0.369	0.138	0.369	0.147	153
<i>Average fathers characteristics</i>									

Age	45.31	1.595	45.58	1.529	45.21	2.275	45.25	1.457	153
Log income	10.92*	1.209	10.10	2.019	9.718	1.757	10.34	1.794	153
Unemployed	0.022	0.028	0.047	0.051	0.040	0.083	0.032	0.044	153
Employed	0.889*	0.083	0.831	0.153	0.809	0.137	0.837	0.146	153
Outside labor force	0.089*	0.070	0.123	0.124	0.150	0.118	0.131	0.138	153
Low education	0.247	0.134	0.261	0.098	0.294	0.138	0.256	0.179	153
High education	0.319	0.175	0.257	0.148	0.284	0.182	0.300	0.171	153
Vocational education	0.435	0.142	0.483	0.145	0.422	0.140	0.443	0.173	153
Number of schools (Leaders)	38		37		39		39		153

Note: The table shows leader and school level average 9th grade student characteristics by treatment status in 2014. Parental characteristics are measured at child age 13. Means are tested against the mean of the control group and t-tests are corrected for sector fixed effects.

*** $p < 0.01$. ** $p < 0.05$. * $p < 0.1$

Table A9. Balance table. Initial leader and student-related characteristics. Standardized test analysis

<i>Leadership Training</i>	Transformational		Transactional		Combination		Control		N
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	
<i>Leader characteristics</i>									
Age	53.53**	6.653	49.41	7.021	51.96	7.861	48.45	7.089	86
Male	0.367	0.490	0.357	0.488	0.143	0.356	0.300	0.466	116
No personal identification number	0.233	0.430	0.179	0.390	0.107	0.315	0.267	0.450	116
Number of Students	512.4	257.4	442.5*	189.2	471.3	151.0	528.3	198.2	116
<i>Average student characteristics</i>									
Male	0.529	0.069	0.504	0.045	0.503	0.0473	0.495	0.045	116
Age	13.20	0.499	13.41	0.683	13.23	0.293	13.25	0.429	116
Danish	0.905	0.084	0.926	0.079	0.865	0.136	0.894	0.170	116
Immigrant	0.025	0.021	0.022	0.025	0.035	0.027	0.028	0.031	116
Immigrant descendent	0.07	0.072	0.052	0.067	0.099	0.118	0.078	0.152	116
Special needs education (Physical)	0.003	0.005	0.006	0.016	0.002	0.005	0.001	0.003	116
Special needs education (Psychical)	0.060	0.184	0.023	0.040	0.016	0.025	0.022	0.029	116

Special needs education (Other)	0.031	0.043	0.027	0.043	0.016	0.031	0.022	0.028	116
No registered parents	0.003	0.004	0.001	0.003	0.004	0.007	0.003	0.005	116
Only father	0.005*	0.007	0.003	0.005	0.006**	0.008	0.002	0.004	116
Only mother	0.016	0.012	0.017	0.017	0.018	0.013	0.016	0.012	116
Avg. National Test score	-0.027	0.209	-0.055	0.253	0.022	0.308	0.039	0.314	116
Completion rate	0.934	0.050	0.932	0.050	0.935	0.049	0.932	0.058	116
<i>Average Mothers Characteristics</i>									
Age	35.86	0.910	35.73	0.904	36.07	1.515	35.85	0.912	116
Single	0.219	0.064	0.200	0.067	0.234	0.069	0.205	0.072	116
Married	0.648	0.085	0.660	0.076	0.650	0.079	0.681	0.087	116
Cohabiting	0.132	0.050	0.140**	0.037	0.116	0.037	0.114	0.036	116
Log-income	10.26	0.931	10.24	1.206	9.834	1.560	10.15	1.526	116
Unemployed	0.045	0.029	0.040	0.028	0.043	0.021	0.045	0.023	116
Employed	0.794	0.087	0.805	0.100	0.761	0.131	0.787	0.127	116
Outside labor force	0.161	0.066	0.155	0.083	0.196	0.117	0.168	0.114	116
Low education	0.264	0.083	0.271	0.077	0.296	0.135	0.264	0.115	116
High education	0.355	0.119	0.321	0.113	0.381	0.179	0.364	0.143	116
Vocational education	0.380	0.089	0.408	0.094	0.324*	0.110	0.372	0.107	116
<i>Average fathers characteristics</i>									
Age	38.48	0.862	38.33	0.876	39.02	1.463	38.54	1.010	116

Log-income	11.09	0.669	10.92	0.683	10.77	1.040	10.87	1.276	116
Unemployed	0.030	0.024	0.024	0.018	0.032	0.022	0.028	0.022	116
Employed	0.881	0.063	0.906	0.064	0.866	0.093	0.881	0.105	116
Outside labor force	0.089	0.044	0.070	0.049	0.102	0.077	0.091	0.087	116
Low education	0.255	0.065	0.277	0.071	0.286	0.101	0.253	0.105	116
High education	0.302	0.127	0.259	0.109	0.323	0.198	0.300	0.153	116
Vocational education	0.443	0.106	0.465	0.082	0.391*	0.131	0.447	0.114	116
Number of schools (Leaders)	30		28		28		30		116

Note: The table shows leader and school-level average 9th grade student characteristics by treatment status in 2014. Parental characteristics are measured at child aged six. Means are tested against the mean of the control group and t-tests are corrected for sector-fixed effects.

*** $p < 0.01$. ** $p < 0.05$. * $p < 0.1$

Table A10: FDR information

Article	Outcome	Leadership behavior	Group	Data Source	Notes	Coef.	p-value	FDR q-value ^a	FDR q-value ^b	FDR q-value ^c
Jensen, 2018	Value Congruence	TFL	Employees	Survey	Only preschool teachers	0,987	0,028	<i>0,097</i>	<i>0,078</i>	0,105
Jensen, 2018	Value Congruence	TFL × Perceived Societal Impact	Employees	Survey	Only preschool teachers	0,012	0,017	<i>0,083</i>	<i>0,066</i>	0,105
Jacobsen & Andersen, 2019	Δ High Performance Expectations	TFL	Employees	Survey		0,050	0,116	0,231	0,184	0,254
Jacobsen & Andersen, 2019	Δ High Performance Expectations	Combi	Employees	Survey		0,105	0,001	0,050	0,038	0,024
Jacobsen & Andersen, 2019	Δ High Performance Expectations	TAL	Employees	Survey		0,077	0,013	<i>0,074</i>	<i>0,057</i>	<i>0,066</i>
Jensen et al., 2019	Public Service Motivation	TFL	Employees	Survey	Short-term Effects	-4,610	0,003	0,050	0,038	0,105
Jensen et al., 2019	Public Service Motivation	COMBI	Employees	Survey	Short-term Effects	-5,500	0,002	0,050	0,038	0,024
Jensen et al., 2019	Public Service Motivation	TAL	Employees	Survey	Short-term Effects	-4,130	0,019	<i>0,083</i>	<i>0,067</i>	<i>0,074</i>
Jensen et al., 2019	Public Service Motivation	TFL × Initial Value Congruence	Employees	Survey	Short-term Effects	0,052	0,013	<i>0,074</i>	<i>0,057</i>	0,105
Jensen et al., 2019	Public Service Motivation	COMBI × Initial Value Congruence	Employees	Survey	Short-term Effects	0,066	0,006	<i>0,054</i>	0,040	0,029
Jensen et al., 2019	Public Service Motivation	TAL × Initial Value Congruence	Employees	Survey	Short-term Effects	0,044	<i>0,067</i>	0,176	0,135	0,159
Jensen et al., 2019	Public Service Motivation	TFL	Employees	Survey	Long-term Effects	-5,940	0,056	0,167	0,123	0,165
Jensen et al., 2019	Public Service Motivation	COMBI	Employees	Survey	Long-term Effects	-7,410	0,022	<i>0,086</i>	<i>0,068</i>	<i>0,072</i>

Article	Outcome	Leadership behavior	Group	Data Source	Notes	Coef.	p-value	FDR		
								FDR q-value ^a	q-value ^b	FDR q-value ^c
Jensen et al., 2019	Public Service Motivation	TAL	Employees	Survey	Long-term Effects	-3,500	0,277	0,376	0,308	0,291
Jensen et al., 2019	Public Service Motivation	TFL × Initial Value	Employees	Survey	Long-term Effects	0,065	0,113	0,229	0,184	0,254
Jensen et al., 2019	Public Service Motivation	COMBI × Initial Value	Employees	Survey	Long-term Effects	0,078	0,063	0,172	0,132	0,155
Jensen et al., 2019	Public Service Motivation	Congruence TAL × Initial Value	Employees	Survey	Long-term Effects	0,041	0,352	0,437	0,341	0,325
Bro & Jensen, 2020	User Orientation	TFL	Employees	Survey	Short-term Effects, preschool teachers	2,660	0,013	<i>0,074</i>	<i>0,057</i>	0,105
Bro & Jensen, 2020	User Orientation	TFL	Employees	Survey	Long-term Effects, preschool teachers	-0,502	0,763	0,727	0,563	0,716
Bro & Jensen, 2020	User Orientation	TFL	Employees	Survey	Short-term Effects, tax employees	0,345	0,838	0,748	0,611	0,722
Bro & Jensen, 2020	User Orientation	TFL	Employees	Survey	Long-term Effects, tax employees	-0,712	0,729	0,714	0,548	0,716
An, et al., 2020	Job Satisfaction	TFL	Employees	Survey		-0,095	0,286	0,376	0,308	0,404
An, et al., 2020	Job Satisfaction	COMBI	Employees	Survey		-0,021	0,835	0,748	0,611	0,778
An, et al., 2020	Job Satisfaction	TAL	Employees	Survey		-0,044	0,582	0,586	0,429	0,506
An et al., 2019	Transformational Leadership behavior	TFL	Employees	Survey		-3,848	0,148	0,264	0,206	0,273

Article	Outcome	Leadership behavior	Group	Data Source	Notes	Coef.	p-value	FDR		
								FDR q-value ^a	q-value ^b	FDR q-value ^c
An et al., 2019	Transformational Leadership behavior	COMBI	Employees	Survey		2,275	0,438	0,461	0,359	0,507
An et al., 2019	Transformational Leadership behavior	TFL x Public	Employees	Survey		7,937	0,012	0,074	0,057	0,105
An et al., 2019	Transformational Leadership behavior	COMBI x Public	Employees	Survey		2,668	0,419	0,459	0,359	0,507
An et al., 2019	Contingent Verbal Rewards	TFL	Employees	Survey		-6,029	0,125	0,241	0,189	0,254
An et al., 2019	Contingent Verbal Rewards	COMBI	Employees	Survey		-0,973	0,750	0,727	0,561	0,752
An et al., 2019	Contingent Verbal Rewards	TAL	Employees	Survey		-0,754	0,796	0,727	0,590	0,671
An et al., 2019	Contingent Verbal Rewards	TFL x Public	Employees	Survey		7,475	0,079	0,193	0,148	0,216
An et al., 2019	Contingent Verbal Rewards	COMBI x Public	Employees	Survey		2,150	0,539	0,530	0,409	0,596
An et al., 2019	Contingent Verbal Rewards	TAL x Public	Employees	Survey		3,040	0,359	0,437	0,342	0,325
An et al., 2019	Contingent Pecuniary Rewards	TAL	Employees	Survey		3,872	0,302	0,394	0,309	0,405
An et al., 2019	Contingent Pecuniary Rewards	COMBI	Employees	Survey		-2,758	0,435	0,461	0,359	0,507
An et al., 2019	Contingent Pecuniary Rewards	TAL x Public	Employees	Survey		0,658	0,872	0,757	0,611	0,737
An et al., 2019	Contingent Pecuniary Rewards	COMBI x Public	Employees	Survey		6,611	0,095	0,213	0,164	0,188
An et al., 2019	Δ Transformational Leadership behavior	TFL	Employees	Survey		-2,473	0,279	0,376	0,308	0,404
An et al., 2019	Δ Transformational Leadership behavior	COMBI	Employees	Survey		3,076	0,221	0,340	0,273	0,325

Article	Outcome	Leadership behavior	Group	Data Source	Notes	Coef.	p-value	FDR		
								FDR q-value ^a	q-value ^b	FDR q-value ^c
An et al., 2019	Δ Transformational Leadership behavior	TFL x Public	Employees	Survey		6,452	0,020	0,083	0,067	0,074
An et al., 2019	Δ Transformational Leadership behavior	COMBI x Public	Employees	Survey		1,862	0,524	0,527	0,409	0,516
An et al., 2019	Δ Contingent Verbal Rewards	TFL	Employees	Survey		-5,375	0,163	0,281	0,217	0,284
An et al., 2019	Δ Contingent Verbal Rewards	COMBI	Employees	Survey		-0,431	0,888	0,757	0,611	0,708
An et al., 2019	Δ Contingent Verbal Rewards	TAL	Employees	Survey		-0,445	0,875	0,757	0,611	0,778
An et al., 2019	Δ Contingent Verbal Rewards	TFL x Public	Employees	Survey		6,795	0,105	0,227	0,179	0,192
An et al., 2019	Δ Contingent Verbal Rewards	COMBI x Public	Employees	Survey		1,596	0,647	0,647	0,489	0,664
An et al., 2019	Δ Contingent Verbal Rewards	TAL x Public	Employees	Survey		2,725	0,401	0,459	0,357	0,349
An et al., 2019	Δ Contingent Pecuniary Rewards	COMBI	Employees	Survey		-3,093	0,390	0,459	0,352	0,507
An et al., 2019	Δ Contingent Pecuniary Rewards	TAL	Employees	Survey		4,062	0,316	0,399	0,314	0,312
An et al., 2019	Δ Contingent Pecuniary Rewards	COMBI x Public	Employees	Survey		6,954	0,083	0,199	0,152	0,188
An et al., 2019	Δ Contingent Pecuniary Rewards	TAL x Public	Employees	Survey		0,429	0,922	0,795	0,626	0,713
This paper	Transformational Leadership behavior	TFL	Employees	Survey	OLS	-1,315	0,434	0,461	0,359	0,464
This paper	Transformational Leadership behavior	TAL	Employees	Survey	OLS	2,993	0,047	0,144	0,107	0,138
This paper	Transformational Leadership behavior	COMBI	Employees	Survey	OLS	4,540	0,004	0,050	0,038	0,028

Article	Outcome	Leadership behavior	Group	Data Source	Notes	Coef.	p-value	FDR		
								FDR q-value ^a	q-value ^b	FDR q-value ^c
This paper	Transformational Leadership behavior	TFL	Employees	Survey	LDV	0,889	0,537	0,530	0,409	0,516
This paper	Transformational Leadership behavior	TAL	Employees	Survey	LDV	3,285	0,010	<i>0,074</i>	<i>0,054</i>	<i>0,066</i>
This paper	Transformational Leadership behavior	COMBI	Employees	Survey	LDV	4,635	0,000	0,012	0,009	0,008
This paper	Transformational Leadership behavior	TFL	Employees	Survey	LDV, Panel	2,814	0,045	0,142	0,106	0,145
This paper	Transformational Leadership behavior	TAL	Employees	Survey	LDV, Panel	2,132	<i>0,077</i>	0,193	0,148	0,159
This paper	Transformational Leadership behavior	COMBI	Employees	Survey	LDV, Panel	3,470	0,004	<i>0,051</i>	<i>0,038</i>	0,028
This paper	Contingent Verbal Rewards	TFL	Employees	Survey	OLS	-1,724	0,390	0,459	0,352	0,448
This paper	Contingent Verbal Rewards	TAL	Employees	Survey	OLS	1,859	0,285	0,376	0,308	0,291
This paper	Contingent Verbal Rewards	COMBI	Employees	Survey	OLS	1,655	0,409	0,459	0,359	0,507
This paper	Contingent Verbal Rewards	TFL	Employees	Survey	LDV	0,949	0,520	0,527	0,409	0,516
This paper	Contingent Verbal Rewards	TAL	Employees	Survey	LDV	2,069	0,112	0,229	0,184	0,192
This paper	Contingent Verbal Rewards	COMBI	Employees	Survey	LDV	2,032	0,155	0,272	0,209	0,284
This paper	Contingent Verbal Rewards	TFL	Employees	Survey	LDV, Panel	1,511	0,285	0,376	0,308	0,404
This paper	Contingent Verbal Rewards	TAL	Employees	Survey	LDV, Panel	1,710	0,190	0,312	0,239	0,252
This paper	Contingent Verbal Rewards	COMBI	Employees	Survey	LDV, Panel	2,424	<i>0,063</i>	0,172	0,132	0,155

Article	Outcome	Leadership behavior	Group	Data Source	Notes	Coef.	p-value	FDR q-value ^a	FDR	
									q-value ^b	FDR q-value ^c
This paper	Contingent Pecuniary Rewards	TFL	Employees	Survey	OLS	2,144	0,133	0,247	0,198	0,255
This paper	Contingent Pecuniary Rewards	TAL	Employees	Survey	OLS	4,444	0,004	0,050	0,038	0,037
This paper	Contingent Pecuniary Rewards	COMBI	Employees	Survey	OLS	3,770	0,008	<i>0,067</i>	0,049	0,031
This paper	Contingent Pecuniary Rewards	TFL	Employees	Survey	LDV	3,120	0,018	<i>0,083</i>	<i>0,067</i>	0,105
This paper	Contingent Pecuniary Rewards	TAL	Employees	Survey	LDV	4,140	0,003	0,050	0,038	0,037
This paper	Contingent Pecuniary Rewards	COMBI	Employees	Survey	LDV	3,509	0,005	<i>0,051</i>	0,039	0,028
This paper	Contingent Pecuniary Rewards	TFL	Employees	Survey	LDV, Panel	3,348	0,026	<i>0,095</i>	<i>0,075</i>	0,105
This paper	Contingent Pecuniary Rewards	TAL	Employees	Survey	LDV, Panel	6,096	0,000	0,011	0,009	0,004
This paper	Contingent Pecuniary Rewards	COMBI	Employees	Survey	LDV, Panel	2,982	0,038	0,125	<i>0,096</i>	0,111
This paper	GPA, Final Exams	TFL	School performance	Registry data		-0,043	0,503	0,515	1,000	0,516
This paper	GPA, Final Exams	TAL	School performance	Registry data		-0,105	<i>0,077</i>	0,193	0,528	0,159
This paper	GPA, Final Exams	COMBI	School performance	Registry data		0,006	0,934	0,800	1,000	0,830
This paper	Completion rate, Final Exams	TFL	School performance	Registry data		-0,030	0,419	0,459	1,000	0,464
This paper	Completion rate, Final Exams	TAL	School performance	Registry data		0,037	0,110	0,229	0,528	0,192
This paper	Completion rate, Final Exams	COMBI	School performance	Registry data		-0,007	0,871	0,757	1,000	0,778

Article	Outcome	Leadership behavior	Group	Data Source	Notes	Coef.	p-value	FDR q-value ^a	FDR	
									q-value ^b	FDR q-value ^c
This paper	Student Standardized Test Scores	TFL	School performance	Registry data		0,050	0,098	0,216	0,528	0,245
This paper	Student Standardized Test Scores	TAL	School performance	Registry data		0,076	0,060	0,172	0,528	0,159
This paper	Student Standardized Test Scores	COMBI	School performance	Registry data		0,017	0,664	0,652	1,000	0,664
This paper	Completion rate, Standardized Tests	TFL	School performance	Registry data		-0,003	0,786	0,727	1,000	0,716
This paper	Completion rate, Standardized Tests	TAL	School performance	Registry data		0,003	0,803	0,727	1,000	0,671
This paper	Completion rate, Standardized Tests	COMBI	School performance	Registry data		0,012	0,319	0,399	1,000	0,507
This paper	GPA, Final Exams	TFL	School performance	Registry data	Principals completed training	-0,105	0,214	0,335	0,836	0,389
This paper	GPA, Final Exams	TAL	School performance	Registry data	Principals completed training	-0,055	0,414	0,459	1,000	0,349
This paper	GPA, Final Exams	COMBI	School performance	Registry data	Principals completed training	-0,043	0,665	0,652	1,000	0,664
This paper	Completion rate, Final Exams	TFL	School performance	Registry data	Principals completed training	-0,012	0,795	0,727	1,000	0,716
This paper	Completion rate, Final Exams	TAL	School performance	Registry data	Principals completed training	0,022	0,480	0,492	1,000	0,404
This paper	Completion rate, Final Exams	COMBI	School performance	Registry data	Principals completed training	-0,028	0,636	0,642	1,000	0,664

Article	Outcome	Leadership behavior	Group	Data Source	Notes	Coef.	p-value	FDR		
								FDR q-value ^a	q-value ^b	FDR q-value ^c
This paper	Student Standardized Test Scores	TFL	School performance	Registry data	Principals completed training	0,071	0,017	<i>0,083</i>	0,248	0,105
This paper	Student Standardized Test Scores	TAL	School performance	Registry data	Principals completed training	0,097	0,013	<i>0,074</i>	0,248	<i>0,066</i>
This paper	Student Standardized Test Scores	COMBI	School performance	Registry data	Principals completed training	0,073	<i>0,093</i>	0,213	0,528	0,188
This paper	Completion rate, Standardized Tests	TFL	School performance	Registry data	Principals completed training	-0,005	0,702	0,699	1,000	0,716
This paper	Completion rate, Standardized Tests	TAL	School performance	Registry data	Principals completed training	0,002	0,887	0,757	1,000	0,708
This paper	Completion rate, Standardized Tests	COMBI	School Performance	Registry data	Principals completed training	0,013	0,356	0,437	1,000	0,507

Note: Transformational leadership training (TFL), combination of transformational and transactional leadership training (COMBI) and transactional leadership training (TAL). Levels of significance: **Bold** significant at a 5%-level. *Italic* significant at a 10%-level.

^aFDR correction using all tests

^bFDR correction using tests of same source (survey or registry data)

^cFDR correction using tests of same treatment (TFL, TAL, COMBI)

References in the appendix:

- Bieman, Torsten, Michael S. Cole and Sven Voelpel. 2012. Within-group agreement: On the use (and misuse) of r_{WG} and $r_{WG(0)}$ in leadership research and some best practice guidelines. *The Leadership Quarterly*. 23(1): 66–80.
- House, Robert J. 1998. Appendix: measures and assessments for the charismatic leadership approach: scales. latent constructs loadings. Cronbach alphas. interclass correlations. In *Leadership: The Multiple-Level Approaches. Contemporary and Alternative*. eds. F. Dansereau and F. J. Yammarino. 23–30. London: JAI Press.
- Jacobsen, Christian B. and Lotte B. Andersen. 2015. Is leadership in the eye of the beholder? A study of intended and perceived leadership practices and organizational performance. *Public Administration Review* 75(6): 829–41.
- Kozlowski, Steve W. J. and Katherine J. Klein. 2000. A multilevel approach to theory and research in organizations: Contextual. temporal. and emergent processes. In K. J. Klein and S. W. J. Kozlowski (Eds.). *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (3–90). San Francisco. US: Jossey-Bass.
- MacKenzie, Scott B., Philip M. Podsakoff and Gregory A. Rich (2001). Transformational and transactional leadership and salesperson performance. *Journal of the Academy of Marketing Science* 29(2): 115–134.
- Moynihan, Donald P., Sanjay K. Pandey, and Bradley E. Wright. 2012. Setting the table: How transformational leadership fosters performance information use. *Journal of Public Administration Research and Theory* 22(1): 143–64.

Podsakoff, Philip M., William H. Bommer, Nathan P. Podsakoff and Scott B. MacKenzie. 2006.

Relationships between leader reward and punishment behavior and subordinate attitudes, perceptions, and behaviors: a meta-analytic review of existing and new research.

Organizational Behavior and Human Decision Processes 99(2): 113–42.

Rainey, Hal G. 2009. *Understanding and managing public organizations*. San Francisco: Jossey-

Bass.

Availability of research data and materials

Survey and register data (leadership behaviour and school performance) can only be accessed through encrypted access to Statistics Denmark. and due to Danish regulation register data cannot be shared.