PRACTICE TRANSFER AND SOCIAL VALUE CREATION: EXISTING DILEMMAS IN CROSS-SECTOR COLLABORATIONS

Fernando D. Domingos[1]
fernandodd1@al.insper.edu.br

Sandro Cabral[1]
sandroc2@insper.edu.br

Sergio G. Lazzarini[1]
sergiogl1@insper.edu.br

Ricardo Paes-de-Barros[1]
ricardopb1@insper.edu.br

Bertrand V. Quelin[2]
quelin@hec.fr

This version: October 2020

Abstract

Practice transfer in cross-sector collaborations (CSCs) can help governments and external partners (private and nonprofit) to create and capture value. Yet, these collaborations are plagued with intellectual rights and performance measurement dilemmas. As CSCs are expected to generate positive externalities, the transferring-side partner's contributions may be replicated to units beyond the original contract and lead to a severe underestimation of the CSC performance. We examine these dilemmas in a CSC in education, involving the transfer of management practices from an external partner whose performance is assessed via randomized control trials. We found that public managers acted as boundary spanners, with replication efforts leading to an underestimation of the external partner performance by 36%. We also observe that the replicable nature of practices escalates these dilemmas.

Keywords: cross-sector collaborations, practice transfer, performance assessment, social value creation, RCT.

Acknowledgments: We thank Olivier Chatain, Camille Doche, and Tomasz Obloj for their valuable comments and suggestions. We also thank seminar audiences at HEC Paris and SIOE. Further, we especially acknowledge the invaluable support by Unibanco Institute for granting access to relevant data. Fernando Domingos thanks all involved with the Individual Research Program at HEC Paris, who enriched his visiting period. Sandro Cabral acknowledges the support of CNPq (grant 428670/2016-4). All mistakes are our own.

[1] Insper – Institute of Education and Research, Brazil
[2] HEC Paris, France

PRACTICE TRANSFER AND SOCIAL VALUE CREATION: EXISTING DILEMMAS IN CROSS-SECTOR COLLABORATIONS

Abstract

Practice transfer in cross-sector collaborations (CSCs) can help governments and external partners (private and nonprofit) to create and capture value. Yet, these collaborations are plagued with intellectual rights and performance measurement dilemmas. As CSCs are expected to generate positive externalities, the transferring-side partner's contributions may be replicated to units beyond the original contract and lead to a severe underestimation of the CSC performance. We examine these dilemmas in a CSC in education, involving the transfer of management practices from an external partner whose performance is assessed via randomized control trials. We found that public managers acted as boundary spanners, with replication efforts leading to an underestimation of the external partner performance by 36%. We also observe that the replicable nature of practices escalates these dilemmas.

Keywords: cross-sector collaborations, practice transfer, performance assessment, social value creation, RCT.

1. INTRODUCTION

Strategy research has devoted increasing attention to the emergence and performance implications of cross-sector collaborations (CSCs) involving public actors and a broad array of private and nonprofit organizations (Bruce, Figueiredo & Silverman, 2019; Quelin, Cabral, Lazzarini & Kivleniece, 2019; Rangan, Samii & Van Wassenhove, 2006). In critical areas, such as education, health, and prisons, even when governments provide core activities, public delivery is often enhanced by the presence of external partners – for-profit or nonprofit – providing new technologies, complementary services, management practices, and advice (Cabral, Lazzarini & Azevedo, 2013; Luo & Kaul, 2019; Alonso & Andrews, 2019; Chatain & Plaksenkova, 2019; Rivera-Santos, Rufin & Wassmer, 2017). In this process, CSCs are expected to create social value creation via improved outcomes valued by not only the exchange partners but also by their beneficiaries (Caldwell, Roehrich & George, 2017; Quelin, Kivleniece & Lazzarini, 2017).

Although CSCs, as other contractual agreements, are subject to the hazards of opportunism and misallocation of decision rights (Bruce et al., 2019; Kivleniece & Quelin, 2012), we present and

discuss additional dilemmas still overlooked by the extant literature. Given the nature of public organizations, CSCs are expected to generate positive externalities – and generate social value – beyond the scope of the focal arrangement and their targeted population (McGahan, Zelner & Barney, 2013; Quelin, Kivleniece & Lazzarini, 2017). This feature has important implications in terms of contract design, as enforcing intellectual rights is problematic in the context of CSCs, especially when public actors are involved (Alonso & Andrews, 2019; Rufin & Rivera-Santos, 2012). For instance, an external partner transferring performance-enhancing practices to targeted public units may be unable to limit the dissemination of these practices to other public units beyond the original agreement. On the one hand, judicial disputes in arrangements involving governments are time-consuming and may be difficult to enforce given the public nature of the transacting party (Spiller, 2010). On the other hand, CSCs are often seen critically by stakeholders who are naturally opposed to the engagement of external (especially private) actors in public services (Dahl & Soss, 2014; Ravitch, 2013).

Complicating matters, it has been increasingly common in CSCs to use counterfactual assessment techniques that try to estimate the causal impact of interventions by comparing the performance of targeted and non-targeted units, as in the case of quasi-experimental and experimental methods (Imbens & Wooldridge, 2009; Duflo, Hanna & Ryan, 2012). Yet, allowing for practice replication beyond the targeted units in the original CSC arrangement may severely distort the assessment of causal impact given that replications increase the performance of non-targeted units that are used as control groups (that is, units that indicate what would have happened without the interventions pursued by the CSC). Thus, the actual contribution of external partners to improved performance may be severely underestimated (Banerjee et al., 2017). However, in order to positively impact as many beneficiaries as possible, the "contamination" of non-targeted groups with superior practices may be seen as socially desirable (Caldwell et al., 2017; Doh,

Tashman & Benischke, 2018). All these factors create critical dilemmas in CSCs involving the transfer of management practices from external actors.

In this paper, we examine these dilemmas by proposing a set of hypotheses that are tested in the context of transfer of practices from an external partner to a public organization. We study the transfer of management practices from the external *transferring-side* partner – in our case, a private nonprofit organization specialized in training and transfer of management practices in education – to a public bureaucracy running a set of public secondary schools (the *learning-side* partner). Our dataset covers almost 500 Brazilian public schools in the States of Espírito Santo and Piauí (whose governments signed collaboration contracts with the transferring-side partner). The main goal of the CSC was to improve the performance of the schools measured by standardized tests of learning in Math and Languages, both important indicators of performance and social value creation in education (Banerjee et al., 2017; Dobbie & Fryer, 2015). The transferring-side partner implemented two sets of management practices, related to goal setting (i.e. usage of short- and long-term goals) and performance monitoring (i.e. data collection and analyses), which have been generally found to promote the effectiveness of public schools (Bruhn, Karlan & Schoar, 2010; Sadun, Bloom & Van Reenen, 2017).

Regarding the identified dilemmas, the performance of this CSC was measured using an experimental design (randomized controlled trial, RCT) to assess the causal impact of the transferred practices. In this design, the targeted (treated) schools receiving the performance-enhancing practices were randomly selected and then compared to the non-targeted (control) schools. The collaboration scaled up managerial training to the public managers of *all* school districts, which include not only the treated units but also schools that were *not* selected to receive the intervention. To examine the effect of practice transfer beyond the treated units, we measure the extent to which the public upper-level manager of each district *internalized* the practices

transferred by the private actor. Thus, drawing from previous work on the role of *boundary spanners* in collaborative arrangements (Bryson, Crosby & Stone, 2015; Stamper & Johlke, 2003; Zhao & Anand, 2013), we investigate the role of pivotal public managers at the *interface* of the cross-sector boundary (Gatignon & Capron, 2020).

Our econometric analyses confirm that the internalization of practice allowed the learning-side partner to replicate the performance-enhancing practices beyond the scope of the primary collaboration. Our findings indicate that a one standard deviation increase in our measure of boundary spanners' practice internalization leads that an underestimation of the performance of the external partner by 36%. Further, we investigate whether the effect of successful practice replication is driven by heterogeneous traits such as human capital and physical resources as moderators at the recipients' level (Banerjee et al., 2017; Cabral et al., 2013; Cohodes, Setren & Walters, 2019; Davis, Guryan, Hallberg & Ludwig; 2017). Arguably, public managers at the cross-sector boundary may be unable to effectively replicate these practices at the non-targeted schools due to resource-based constraints (e.g., untreated schools may lack skilled teachers and principals). However, our analysis suggests that heterogeneous public resources at the learning-side level do not constraint the adoption of transferred practices. Consistent with previous work on the role of practices (Bromiley & Rau, 2014), the performance-enhancing practices under study here are apparently easily implementable and replicable, thus mitigating the importance of resource heterogeneity at the learning-side level.

Our paper provides some contributions to strategy scholarship. Prior works on CSCs recognize important boundary conditions under which public, private, and nonprofit actors in collaborative arrangements create and capture value (Cabral, Mahoney, McGahan & Potoski, 2019; Kroeger & Weber, 2014; Lazzarini, 2019; Luo & Kaul, 2019). We unveil potential dilemmas that arise when collaborative partners intend to capture the value created in CSCs and at the same time generate

broader and relevant gains captured by targeted and untargeted populations. Further, by showing that learning-side partners at the cross-sector boundary are able to successfully replicate practices to untargeted organizations beyond the contract scope, our work also connects the CSC literature with the evolving body of work examining the performance implications of practices and practice transfer (Bloom, Brynjolfsson, Foster, Jarmin, Patnaik, Saporta-Eksten, & Van Reenen, 2019; Bromiley & Rau, 2014; Sadun et al., 2017; Vermeulen, 2018). Also dialoguing with recent works focused on performance measurement of public and nonprofit organizations (Alonso & Andrews, 2019; Bruce et al., 2019; Gerrish, 2016), our study reinforces the central role of managerial practices and their replicable nature in the creation of social value.

2. THEORY AND HYPOTHESES

2.1. CSCs and Performance-Enhancing Managerial Practices

A burgeoning literature has examined the role of the adoption of management practices in organization performance. In fact, one of the most critical knowledge assets an organization possesses is the set of organizational routines it adopts (Nelson & Winter, 1982; Teece, Pisano, and Shuen 1997). More recently, Bromiley and Rau (2014) proposed a *practice-based view* with an emphasis on performance-enhancing practices that are potentially replicable. Simply put, practices refer to the set of routines that different organizations might implement. In contrast with the resource-based view (Barney, 1991; Peteraf, 1993), which emphasizes the role of unique and difficult-to-imitate resources, practice-based arguments consider the possibility of routines that might be adopted and further transferred to a variety of organizations.

However, despite their replicable nature, organizations appear to adopt distinct types of management practices, thus leading to interfirm performance heterogeneity (Bloom et al., 2019; Rahmandad & Ton, 2020). The transfer of such performance-enhancing managerial practices is also relevant in the context of CSCs (Alonso & Andrews, 2019), where partners frequently transfer

and receive relevant knowledge (Bruce et al., 2019). A myriad of collaborations combine the expertise of public managers, nonprofit, private, and multilateral organizations (Rangan et al., 2006; Quelin, Kivleniece & Lazzarini, 2017).

Indeed, examining public-private collaborations, Cabral et al. (2013) observe that practice transfer can occur in both directions, in cases where public actors can learn from external partners and vice-versa. Yet, we specifically focus on cases where the recipient – which we refer to as the learning-side partner – is a government with a collaborative contract with an external (nongovernmental) actor – which we refer to as the transferring-side partner. In this context, CSC contracts with external partners may increase the performance of government units and increase social value in several ways. External actors may encourage the internalization of novel planning and execution routines that increase performance, such as monitoring structures (Cabral et al., 2013) and targets (Alonso & Andrews, 2019). Such transfer of performance-enhancing practices to the public sector is proven efficient in settings such as the management of health care systems (e.g., Banerjee, Duflo & Glennerster, 2008), educational outcomes (e.g., Bloom et al.; 2015; Duflo et al., 2012) and prisons (Cabral et al., 2013). Therefore, this leads to our baseline hypothesis:

> ***Hypothesis 1:*** *In CSCs, the transfer of management practices from an external transferring-side partner to a public learning-side partner increases the performance of the targeted units.*

2.2. Dilemmas in CSCs: The Inherent Tensions between Socially Desirable Replication and Performance Assessment

We now examine dilemmas in the transfer of practices given that the learning-side partner is a public actor. Our key mechanism is based on the action of *boundary spanners*, namely, public managers closing interacting with external actors and responsible for the internalization and dissemination of the transferred practices. The role of boundary spanners has long been discussed

in the alliance literature: individuals enabling transfer channels (Allen, Tushman & Lee, 1979; Dollinger, 1984) by operating at the boundary of an organization and performing crucial tasks both for their internal organization as well as external partners (Leifer & Delbecq, 1978). In this context, the main functions of boundary spanners relate to filtering, translating and diffusing external information to their internal organization such that the received routines are also understandable for their peers (Allen et al., 1979).

The boundary spanner structure has been established to be an effective device for cross-sector transfer of information. In the context of CSCs, boundary spanners are crucial actors who "make things happen", and who play a strong leadership role (Bryson et al., 2015). As the cross-sector boundaries are increasingly blurred (Caldwell et al., 2017; Klein, Mahoney, McGahan & Pitelis, 2009; Luo & Kaul, 2019), mainly regarding the process of contract design and enforcement, these individuals at the boundary are pivotal in aligning initial conditions, processes, outcomes, and accountability structures that allow positive and sustained outcomes over time (Agranoff, 2007; Villani, Greco & Phillips, 2017). Hence, boundary spanners are likely to enable social value creation: benefits that accrue not only to the transacting parties but also to their beneficiaries (Kroeger & Weber, 2014; Lazzarini, 2019).

In this context, a key dilemma arises especially when the learning-side partner in CSCs is a public organization. In this case, boundary spanners are expected to generate public benefits beyond the scope of the original collaboration in order to generate the highest possible social value (McGahan et al., 2013; Quelin et al., 2017). As they internalize transferred practices, boundary spanners can enhance practice transfer not only to units involved in the CSC contracts, but also other units under their jurisdiction. Indeed, it is socially desirable that public managers at the cross-sector boundary have social concerns related to the public nature of their role and replicate the internalized practices that are known to positively impact the performance of all the subordinate

units within their jurisdictions (Caldwell et al., 2017; Doh, Tashman & Benischke, 2018; Gatignon & Capron, 2020).

Under these circumstances, external transferring-side partners might not be able to credibly implement and enforce contractual clauses avoiding such externalities. Criticism from myriad stakeholders might escalate, as the replication of practices beyond CSCs target units could be even a requirement to increase the legitimacy of cross-sector interactions. Preventing other beneficiaries from capturing these externalities may be seen as illegitimate and against the public interest, especially when external actors are perceived to act for their own (private) benefit (Dahl & Soss, 2014; Ravitch, 2013). Moreover, if the learning-side partner is a public organization, courts may not perceive any breach of the original contact preventing broad transfer as an instance of government opportunism, especially when there is pressure from governments and external stakeholders positively affected by the dissemination of valuable practices (Spiller, 2010; Valéro, 2015). Thus, we hypothesize:

> ***Hypothesis 2a:*** *In CSCs, the higher the degree of internalization of practices by the boundary spanners at the public learning-side partner, the higher the performance of the non-targeted units that were not originally defined as recipients of these practices by the external transferring-side partners.*

An additional dilemma arises when there is an increasing demand for a more rigorous performance assessment of the CSCs. The evaluation of policies usually involves an assessment of counterfactual scenarios (Imbens & Wooldridge, 2009; Duflo et al., 2012) – in our case, what would have happened to the recipient and its beneficiaries without the transferred practices. Counterfactual assessment is usually carried out using quasi-experimental or experimental designs, such as RCTs, which follow design protocols to increase internal validity and assure causality (Angrist & Pischke, 2008). Such improved analysis might not only be useful to gauge performance,

but also promote higher accountability in cases where CSCs involve public resources and funding from taxpayers (Behn, 2003; Gerrish, 2016). Even in cases where CSCs involve small-scale pilot projects (Banerjee et al., 2017), there might be an expectation that the programs could be scaled up after rigorous evidence of success (Cohodes et al., 2019; Davis et al., 2017).

However, practice replication beyond the units originally defined as targets in the original contract (as indicated by H1a) can severely undermine such counterfactual assessments. Essentially, practice transfer via boundary spanners managing both targeted (under the CSC contract) and non-targeted units may generate spillovers to the control group, thus distorting the assessment of what would have happened to the public units had they not been subject to the intervention (Banerjee et al., 2017). In other words, as boundary spanners successfully internalize and replicate practices, they may severely distort performance assessments if the transferred managerial practices also increase the performance of non-targeted units originally serving as control groups. For example, managerial practices are internalized when boundary spanners properly understand their role as a manager, monitoring and interacting with different hierarchy instances, as well as applying data-driven routines. This would undermine the perceived contribution and added value of the transferring-side external partner. Thus, we propose:

> ***Hypothesis 2b:*** *A higher internalization of practices by the boundary spanners (acting as a public learning-side partner) reduces the assessed contribution of the external (transferring-side) partners to the performance of the CSC.*

2.3. The Interplay between Practices and Resources: Securing Successful Replication

Practice-based arguments usually emphasize the replicability of practices (Bromiley & Rau, 2014) and contrast with resource-based arguments underscoring how unique resources can become impediments to imitation and replication (Barney, 1991; Peteraf, 1993). An integrative approach between these two approaches should include the resources embodied and the human actors

involved in the practice transfer (Feldman & Orlikowski, 2011; Feldman & Pentland, 2003). For instance, Dyer and Hatch (2006) show that the lack of network resources potentially limits the transfer of practices among partners. As such, effective practice transfer might not be a stand-alone process (Jarzabkowski, Kaplan, Seidl & Whittington, 2016) and may require the presence of superior resources on the recipient side.

Accordingly and drawing on our prior hypothesis, even if boundary spanners at the CSCs learning-side partners properly internalize the transferred practices, they may still be unable to effectively replicate these practices to the non-targeted units due to lack of resources at the recipient level. The interplay between resource- and practice-based traits may enable cross-sector partners to better articulate their collaboration (Quelin et al., 2017). In this vein, the presence of recipients with superior levels of *human capital resources* is likely to positively moderate the effect of internalization of practices by boundary spanners. Indeed, the leverage of human capital is known to moderate the relationship between strategy implementation and firm performance (Hitt, Bierman, Shimizu & Kochhar, 2001). In this regard, prior studies of CSCs posit that practices received through replication might not be properly adopted if recipients lack sufficient skills to implement them (Currie & Proctor; 2005; Villani et al., 2017). This is the case if unskilled employees are unable to translate received practices into their daily work. Along these lines, lower human capital resources could limit the effect of managerial practices to improve student learning (Duflo et al., 2012). In other words:

> ***Hypothesis 3a:*** *In CSCs, the performance effect of the internalization of practices by boundary spanners (as predicted by H2a) is enhanced by the presence of superior human capital resources possessed by the non-targeted units.*

The lack of *physical resources* (i.e. adequate facilities and communication infrastructure) can also undermine the efforts of boundary spanners even when they internalize the transferred

practices in CSCs. For example, there is evidence that schools with enough infrastructure to split students into smaller group sizes are known to have superior performance (Davis et al., 2017). Also, research has examined the positive effect of computer and information technology infrastructure on the ability of schools to leverage their pedagogical practices (Becker & Ravitz, 1999) and that physical resource heterogeneity across developed and developing settings affects the effective adoption of transferred practices (Bloom, Lemos, Sadun, Scur & Van Reenen, 2016). We thus propose our last hypothesis:

> ***Hypothesis 3b:*** *In CSCs, the performance effect of the internalization of practices by boundary spanners (as predicted by H2a) is enhanced by the presence of superior physical resources possessed by the non-targeted units.*

Although our previous hypotheses propose resource-based effects that moderate the link between practice internationalization and practice transfer, we note that alternative arguments lead to a distinct prediction. Namely, as emphasized by the practice-based view of organizational performance (Bromiley & Rau, 2014), practices may be relatively replicable and transferrable across units, even if these units lack critical resources. For instance, managerial practices such as setting goals, having clear performance assessment, promoting talented workers, and removing poor performance are all well known, even though not uniformly adopted by a broad range of organizations. They may be considered standard practices that do not require idiosyncratic resources for their implementation (Bloom et al., 2016). If this is the case, the effect of recipient-level resource heterogeneity may not be sufficiently relevant to influence and moderate the replication of practices.

Figure 1 summarizes the hypothesized relationships. In a nutshell, transfer of practices, in our context concerning managerial practices, should influence performance (H1). However, the effect of these performance-enhancing practices may go beyond the targeted subunits due to the

replication of practices enabled by the learning-side partner (H2a and H2b). Finally, the replication impact on the nontargeted may depend on their available resources (H3a and H3b).

[Figure 1 about here]

3. DATA AND METHOD

3.1. Context and Empirical Setting

We investigate a CSC in the Brazilian educational sector, involving a program funded by the *Unibanco Institute*, a private nonprofit organization that represents the external transferring-side partner of the collaboration. The external partner (private nonprofit organization) has developed a bundle of managerial practices with an expected positive effect on the performance of public high schools. Thus, the learning-side partners in this CSC are state-level governments, which are expected to internalize these practices.

Influenced by managerial practices commonly adopted in the private sector (such as planning and process control), the external partner has focused on two main programs: target setting (i.e. the definition of short and long term performance goals) and performance monitoring (i.e. information collection and analysis to assess the relative performance of units). These practices are well known, and their positive effect has also been documented in the context of public schools (Bloom, Lemos, Sadun, Scur & Van Reenen, 2014). Importantly, the performance of the CSC has been evaluated using an experimental design (randomized controlled trial, RCT). Namely, departing from a list of eligible units, the external partner randomly chose a set of targeted schools to directly receive training to implement these practices, with the non-targeted schools serving as a control group.

Our empirical setting is particularly well suited to test our hypotheses for two main reasons. First, although public schools were randomly selected to receive the training program following the RCT design, managerial practice training was also delivered to upper-level managers acting as boundary spanners in the CSCs. These boundary spanners were public upper-level managers acting

as heads of school districts, encompassing a group of regionally defined units. Boundary spanners were expected to enforce the adoption of the practices at the schools specified in the collaboration agreement. However, given the proposed RCT design, the set of schools under the supervision of the boundary spanners could involve *both* treated (targeted) and control (non-targeted) units. For this reason, it would be possible for boundary spanners to learn from their practices transferred by the external partner to the targeted units, and then use this knowledge to also stimulate the adoption of improved practices in the non-targeted schools.

Second, the collaboration is based on transferring managerial practices that can positively affect the performance of schools, according to our baseline hypothesis. More precisely, the intervention consists of a training program consisting of 68 contact hours with heads of school districts (the upper-level managers acting as boundary spanners for the CSCs, responsible for several schools), 48 contact hours and 120 online hours for school managers and pedagogical coordinators (lower-level managers, at the school level). Usually, all these managers had previously been elementary and high-school teachers and had received no long-term dedicated managerial training before acting as school principals or heads of districts. Furthermore, the program provides public managers with scripts, goals, protocols, and management practices that facilitate, stimulate, and promote action regarding planning and monitoring daily activities, and evaluating operational performance.

Considering such a setting, we were granted access to proprietary data on the school districts' degree of internalization of practices, which enabled us to examine whether it led to potential spillover effects to the non-targeted schools. In addition, we control for other forms of replication. Because lower-level managers (school principals) could also talk and generate direct spillovers across schools, we checked whether schools shared physical spaces and teachers, and controlled for their physical distance.

3.2. Data and Variables

Our detailed data allow us to examine features at the organizational level (i.e., schools, including aggregated features of teachers and students), boundary spanner level (i.e., school district manager), and performance outcomes (students' grades aggregated at school level). Our final sample comprises almost 400 schools in the Brazilian States of Espírito Santo and Piauí, from 2014 to 2017. These schools are divided among 9 districts in Espírito Santo and 11 districts in Piauí, each headed by one boundary spanner, thus allowing us to examine 1564 school-year observations.

3.2.1. Dependent variables

Our main dependent variables are organizational performance measures (school-level performance) reflecting students' achievement according to their average grades on standardized exams (Math and Languages). Student learning is a key performance measure in educational services and represents an outcome that is highly valued by the target populations. As learning translates into an increase in several future outcomes, such as earnings, college attendance, retirement savings (Chetty et al., 2011), and improved health outcomes (Aud & Sidhu, 2005) and crime reduction (Jacob & Lefgren, 2003), it is probably associated with social value creation (Kroeger & Weber, 2014). We observe these two indicators (Math and Languages grades) because they are both considered relevant outcomes of student learning (Dobbie & Fryer, 2011, 2015). Our data come from standardized tests performed by the state governments, which are a well-known source of data to assess school performance.

3.2.2. Independent variables

Transfer of practices from the external partner. We use a binary variable identifying targeted schools and non-targeted schools (*TargetedSchool*), as defined by the RCT procedure employed in the evaluation of the CSC (233 targeted and 161 non-targeted schools). This variable helps us to

identify the effect of the transfer of practices to the schools that were selected to be recipients and is used to test our baseline hypothesis (H1).

Practice Internalization. Concerning our hypothesized variable for internalization of practices (H2), this refers to the boundary spanners' actual adherence to the practices specified in the collaboration agreement. Our measure relies on an index developed by the external partner. Such an index is based on interviews with all boundary spanners and observational data regarding the frequency of interactions between the district and schools and the State Department of Education.[3] The higher such an index, then the higher the expectations of successful internalization of received practices from the external partner. For example, boundary spanners who properly use the proposed data-driven routines focused on students' learning have a higher score than the ones who merely use data for bureaucratic purposes (see Appendix A for detailed information). This measure (*PracticesInternalization*) is used to test H2a and H2b.

Heterogeneous resources at the recipient level. Lastly, we observe the organizational traits of nontargeted units to identify their heterogeneous resources, considering schools' and teachers' traits related to their available human capital (H3a) and physical resources (H3b) respectively. Human capital is proxied by the number of teachers per student in our main specification. Another proxy we use as robustness is the average number of teachers with a college degree at that particular school. Physical resources are proxied by the number of computers per student, meanwhile, for robustness purposes, another specification considers whether the schools have access to high-speed internet.

[3] The index composition is available on Appendix A. Originally, it is a score (from 1 to 4) consisting of the average of 5 main components related to managers' internalization and implementation of the so-called best managerial practices and routines. The Cronbach's Alpha reliability coefficient is 0.65. We use as our main independent variable a score following a factor analysis of all components, given as the principal axis factor (robustness checks consider other specifications).

Control variables. We count on several important features at the organizational level (school features and aggregated characteristics of teachers and students, as described in Table 1). We consider school resources (e.g. computers per student, access to high-speed internet) and networks (distance from other schools). We highlight the distance from the closest targeted school as an important control for any other sort of spillover among schools, so we can focus on the replication of practices from the boundary spanner. At the individual level (aggregated per school), we rely on both teacher and student features. For teachers, we control for human capital proxies (proportion of teachers with a college degree). At the student level, we also control for socioeconomic features (i.e., age, gender, ethnicity). Lastly, we consider state fixed effects.

Table 1 outlines the descriptive statistics. Table 2 shows the correlation matrix. There is no evidence of multicollinearity among the variables.

[Table 1 about here]

[Table 2 about here]

3.3. Methodology

To test our baseline hypothesis (H1), we use an experimental approach to explore how performance-enhancing managerial practices affect organizational-level performance. We highlight that the subsamples of targeted and non-targeted schools are paired according to their predicted performance evolution from 2014 to 2016, following the initial RCT design performed by the external partner. As such, schools were further randomized between targeted and non-targeted, within states. These two groups were found to be similar in terms of their observable characteristics, as revealed by Kolmogorov-Smirnov and t-tests; and the randomization procedure is expected to also promote balancing in terms of unobservable traits. Therefore, the design gives us confidence that no other observable features are biasing our results.

Next, to test our main hypothesis (H2), we investigate the boundary spanners' internalization and replication of practices. To do so, we estimate a series of difference-in-differences and triple-difference models predicting the effect of the practice transfer on the educational outcomes. All specifications reported in the paper are clustered OLS regressions at the boundary spanners' level. We thus specify our main model explaining the performance of the schools:

$$Y_{i,t} = \beta_1 TargetedSchool_i + \beta_2 PracticeInternalization_{i(b)} + \beta_3 Post + \\ + \beta_4 Post * TargetedSchool_i + \beta_5 Post * PracticeInternalization_{i(b)} + \\ + \beta_6 TargetedSchool_i * PracticeInternalization_{i(b)} + \\ + \beta_7 Post * TargetedSchool_i * PracticesInternalization_{i(b)} + \beta_8 Controls_i + \epsilon_{i,t}, \quad (1)$$

where $Y_{i,t}$ represents the performance of school *i* in period *t* (which, as discussed before, is measured as the school's average score for Math and Languages), $TargetedSchool_i$ is an indicator variable taking the value of 1 for targeted schools (intervention at the school level), $PracticesInternalization_{i(b)}$ refers to the boundary spanners (*b*) internalization of practices (index measured by the external partner) and *Post* is the post-treatment variable used according to the difference-in-differences methodology (2014 is the reference year, such that 2015, 2016 and 2017 refer to post-treatment).[4] The model further includes several school, teacher and student socioeconomic controls and state fixed effects.

As hypothesized, boundary spanners (upper-level managers, heads of school districts) could replicate practices to non-targeted schools (H2). Following H2a, internalization of practices leads to higher performance of the non-targeted units. Thus, we first narrow our analyses to the subsample of non-targeted schools. We expect a positive sign for β_5, which would indicate the

[4] We recall that the practice internalization score is the same for every school (targeted or non-targeted) under the jurisdiction of the same boundary spanner (i.e., i(b) = i'(b)).

replication efforts beyond the original contract. Further, we re-run our analyses considering the full sample (targeted and non-targeted), which enable us to properly investigate all coefficients of equation (1). The resulting interaction between post-treatment, targeted schools, and internalization of practices constitutes a triple-difference estimator, which disentangles the effect of boundary spanners' internalization of practices among the targeted and non-targeted schools after the introduction of the intervention. According to H2b, we expect a negative sign for β_7, which would understate the private efforts regarding targeted schools.

Lastly, we observe the role of recipient-level resources as potential moderators for the successful replication of practices (H3). Again, we narrow our analyses to the subsample of non-targeted schools. Then, we consider binary variables identifying the most resourceful schools both in terms of human capital (H3a) or physical resources (H3b). We code "resource-high" schools as the top 25% higher in terms of our variables measuring these two resource-based factors ($ResourceHighHumanCapital_i$ and $ResourceHighPhysicalCapacity_i$; both equal 1 if the 25% highest, 0 otherwise). We consider different thresholds for robustness. We expect the interaction between the boundary spanners' internalization of practices, resourceful schools and post-treatment to be positive; in other words, resource gaps would hamper the adoption of managerial practices.

4. ANALYSES AND RESULTS

4.1. Performance-Enhancing Managerial Practices

Table 3 outlines our results considering the average scores for Math and Languages. For both performance outcomes, specification (1) accounts only for the school's treatment (binary variable, 1 for targeted schools). Even though this variable is exogenous by design (i.e. it comes from the RCT used to define the targeted and non-targeted schools), we also add additional controls (school-, teacher-, and student-related traits) in the specification (2).

[Table 3 about here]

As expected, the baseline effect of practices can be seen in the interaction between the indicator variable for targeted units and the post-treatment variable, following the difference-in-difference methodology ($Post * TargetedSchool_i$). The results are highly significant both for Math and Languages ($p < 0.01$) in both specifications, with and without additional controls. These show that the improved learning effects result in what would have been expected in a year of additional schooling (around 10% of a standard deviation).

We emphasize that prior work has found that managerial practices have higher impacts on Math than on Languages (Dobbie & Fryer, 2011; Fryer, 2014). Research in developmental psychology has suggested that the critical period for Language development occurs early in life, whereas the critical period for developing higher cognitive functions related to Math extends into adolescence (Knudsen, Heckman, Cameron & Shonkoff, 2006). Another leading theory posits that reading scores are influenced by the language spoken when students are outside the classroom (Charity, Scarborough, & Griffin 2004). Yet, our results hold for both Math and Languages, thus emphasizing the performance-enhancing potential of the implemented managerial practices and lending support to our baseline hypothesis (H1).

4.2.Existing Dilemmas at Cross-Sector Boundaries

We next test the effect of practice internalization, as predicted by hypotheses H2. First, we narrow our scope to the subsample of non-targeted schools to investigate whether the internalization of received practices by boundary spanners indeed increases the performance of nontargeted units (H2a). To strengthen our confidence regarding our main independent variable, we rely on three different specifications. Our preferred specification consists of the principal factor (Table 4, specification 3) – which is a continuous measure (see Appendix A for further details). In sum, as previously argued, the higher the factor for internalization, the higher the expected performance.

Further, we use both a binary measure (specification 4, segregating high vs. low boundary spanners' internalization) and a three-level measure (specification 5, segregating per terciles). These help us to investigate whether there is any rupture or discontinuity in the practice internalization and performance relationship. Table 4 presents the results.

[Table 4 about here]

Overall, the results corroborate our hypothesis. The hypothesized effect is significant for both educational outcomes for every specification (β_5), except considering the continuous variable and Languages. Interestingly, only the highest tercile present significant effects for the three-level specification ($p=0.03$ for Math and $p<0.001$ for Languages), thus suggesting that successful replication relies on boundary spanners who internalize practices more vigorously. This result sheds new light on the existing literature claiming that managerial practices have little or no significant effect on Languages, although they significantly increased student performance in Math (Dobbie and Fryer, 2011; Fryer, 2014). We show that effects on Languages rely on the superior ability to internalize practices, a factor not explored yet.

We advance to our full sample to investigate whether assessments could undermine the impact of the transferring-side on the targeted units (H2b). The mechanisms behind this result would be due to pervasive consequences of the enhanced performance of non-targeted schools (as predicted by H2a). All the following specifications consider the continuous measure for internalization of practices (as a result, we benefit from heterogeneity among boundary spanners to further explore the internalization effects).

Table 5 presents the results. Specification (6) considers the internalization of practices by boundary spanners and the difference-in-difference estimator controls. Specification (7) follows our main model in equation (1), relying additionally on the triple difference estimator. We examine how the internalization of practices can lead to a reduction in the perceived contribution of the

transferring-side partner. To do so, we considering the interaction between internalization of practices, targeted schools, and the post-treatment variable ($Post * TargetedSchool_i * PracticesInternalization_{i(b)}$). This model helps us to disentangle the effects of internalization of practices between targeted and nontargeted schools, investigating whether managers replicated practices beyond the targeted schools.

[Table 5 about here]

In Table 5, considering Math, we provide support for H2b (β_7 = -2.2, p = 0.01). The negative and statistically significant coefficient provides evidence that assessments understate the impact of the performance-enhancing practices in targeted schools. Indeed, it suggests that an additional standard deviation of our measure gauging boundary spanners' internalization of practices reduces the observed impact of the primary transfer of practices to targeted schools by 36%. The result is not significant for Languages, again corroborating prior literature.[5] Moreover, the coefficients for our baseline hypothesis (β_4) are still strongly significant ($p < 0.001$) across these additional specifications, thus providing additional support to our baseline hypothesis.

4.3. Resource-based Heterogeneities

Finally, we investigate whether the effect of performance-enhancing practices is moderated by human capital (H3a) or physical resources (H3b) of non-targeted subunits. Table 6 outlines our results. Specification (8) considers the human capital moderation, meanwhile, specification (9) considers the physical capacity – we follow our prior analyses and run regressions both for Math and Languages, using the same set of control variables.

[Table 6 about here]

[5] If we use the three-level measure for practice internalization instead of the continuous measure, then the triple-difference considering the highest tercile present significant effects also for Languages (p=0.086); again suggesting replication would depend on boundary spanners who internalize practices more vigorously.

Although we notice that the hypothesized effect of internalization of practices post-treatment for nontargeted schools (H2a) is overall significant for both Math and Languages, providing additional evidence in support to our main hypothesis, we found no support for our moderator hypotheses (H3a and H3b). None of the coefficients regarding the interactions between our resource variables and internalization of practices post-treatment follow our predictions (one is even negatively significant). Other specifications considered the different thresholds to define the resourceful schools (between 5% to 25% higher resources) and the results were unchanged. We also considered different sub-samples regarding the overall level of schools' resources and conducted different regressions for each, investigating whether the coefficients for internalization of practices post-treatment differed. In this case, we expected more prominent results for the sub-sample of high-capacity schools. The results were again not significant.

Arguably, an alternative explanation for the lack of support for H3 is that our examined practices are, when available, imitable, and easily replicable (Bromiley & Rau, 2014). In other words, successful replication would not require resourceful schools with above-average human capital nor physical capacity. Indeed, Cohodes et al. (2019) have shown that schools can replicate managerial practices, such that performance improvements are not driven by student nor teacher characteristics because a standardized approach may facilitate portability. Also supporting this alternative explanation, prior studies considering outsourcing in the educational sector have observed that government can implement successful managerial practices that do not demand additional investment in skilled workers or superior infrastructure (Klees, 2018; Romero, Sandefur & Sandholtz, 2017). Thus, moderation could even go in the opposite direction if the boundary spanners focus their replication efforts on the nontargeted units with resource gaps, potentially the most vulnerable.

4.4.Additional Robustness Checks

To further strengthen confidence in our main findings, we perform several robustness checks. First, we perform placebo tests investigating whether non-targeted schools present effects of the transfer of managerial practices without actually receiving them. We drop targeted schools from our sample and estimate our main regression specification – equation (1) – randomly assigning non-targeted schools as the targeted units (identified by a binary variable $PlaceboSchool_i$ instead of $TargetedSchool_i$). As a result, we expect no effect for H1 ($Post * PlaceboSchool_i$), nor for H2b ($Post * PlaceboSchool_i * PracticesInternalization_{i(b)}$), as the effect of internalization of practices through replication should be equal between placebo or not among the non-targeted. We repeat this procedure 1000 times, and for all test scores, the true coefficients fall above the 99th percentile of the distribution of placebo coefficients. Hence, our findings are unlikely to have been obtained at random.

Second, regarding the triple difference estimator and our findings for Math, we check whether a specific boundary spanner drives the effects we find, rather than the internalization of practices itself. Therefore, we use the leave-one-out resampling method in which the model is retested by dropping out one subsample (Efron, 1982). Simply put, we estimate our main specification (following equation 1) 20 times to get a range of estimates for our model, each time excluding all the schools of one of the boundary spanners. Figure 2 shows the results of these leave-one-out tests. Our point estimates (for β_7) do not substantially change and are significant in every specification no matter which boundary spanner is dropped out of the sample (all significant at $p < 0.05$).

[Figure 2 about here]

As an additional robustness check, we observe whether the results are the same using different factor analysis extraction methodologies for our index of internalization of practices, such as the

maximum-likelihood instead of the principal axis factoring (Costello & Osborne, 2005). We re-run our analyses from Table 4 and Table 5 (see Appendix B). The findings are maintained.

5. DISCUSSION AND CONCLUSION

By showing that practice transfer from external partners (private nonprofit) to learning-side (public organizations) partners increases the performance of targeted units and that the boundary spanners who internalize transferred practices not only improve the performance of non-targeted units but also reduce the assessed contributions of external partners in CSCs, we inform the extant debates in strategy on the challenges of public-private contracting (Cabral et al., 2019; Lazzarini, 2019; Kroeger & Weber, 2014; Luo & Kaul, 2019). Moreover, we add to existing works on performance implications of practice transfer at the cross-sector boundary (Alonso & Andrews, 2019; Bruce et al., 2019; Gatignon & Capron, 2020) by highlighting the role of boundary spanners (Bryson et al., 2015; Stamper & Johlke, 2003; Zhao & Anand, 2013) on the creation of value through the replication of practices (Bloom et al., 2019; Bromiley & Rau, 2014; Sadun et al., 2017; Vermeulen, 2018).

Our study also contributes to the literature on the organizational determinants of social value creation through CSCs. While existing works have focused on the positive externalities of CSCs (Caldwell et al., 2017; Doh et al., 2018; Rangan et al., 2006) or raised concerns about potential contracting hazards associated with government opportunism (Inoue et al., 2013; Kivleniece & Quelin, 2012; Spiller, 2010; Valéro, 2015), we unveil additional dilemmas related to CSCs and the generation of broader social gains, by shedding light on the ability of boundary spanners to internalize practices and further replicate them beyond the scope of the original CSC contract. We demonstrate that the presence of boundary spanners is a double-edged sword in CSCs: they potentialize practice transfer, but at the same time can allow for practice replication beyond the scope of the original CSC contract. Although replication is highly beneficial from a public

standpoint, it may undermine the effort of the external (transferring-side) partner and lead to an underestimation of its perceived contribution.

By investigating the role of managerial practices, considering both their primary transfer and further replication through boundary spanners, we also inform the literature on performance-enhancing practice transfer (Bloom et al., 2019; Bromiley & Rau, 2014, Rahmandad & Ton, 2020; Sadun et al., 2017). Moreover, we shed light on the importance of standardized practices for successful replication (Cohodes et al., 2017) by showing that when practices are easily imitable (Bromiley & Rau, 2014), there is no apparent resource-based constraint to adoption. Thus, the nature of the practice seems to be crucial for their effective implementation in CSCs, especially when boundary spanners effectively internalize and replicate imitable performance-enhancing practices at the cross-sector boundary.

We also dialogue with recent works focused on performance measurement of public and nonprofit organizations by investigating how willingness to generate public value creates tensions in the assessment of contributions from different partners in collaborative efforts (Alonso & Andrews, 2019; Bruce et al., 2019; Gerrish, 2016; Kroeger & Weber, 2014). Our proposed dilemmas are exacerbated in the context of CSCs, given their general objective to have a broad, cross-sector impact (Cabral et al., 2013; Quelin et al., 2017, Quelin et al, 2019). However, tensions might also arise if any other types of alliances have the dual objective of promoting generalized learning and evaluating performance. For example, whenever alliances want to scale up programs, it might be challenging to propose evaluation guidelines (Cohodes et al., 2019; Davis et al., 2017). The dilemmas would intensify if scale-up routines depend on generalized learning and partners rely on counterfactual assessments to estimate the causal impact of the transferring-side partner.

Future research in the context of collaborative arrangements could investigate if our observed phenomena are persistent considering other environments and actors involved in the transfer and

replication of practices. Particularly, studies could investigate settings where replication of practices has high marginal costs and where practices are not as imitable and replicable as in educational settings (Cohodes et al., 2019; Klees, 2018; Romero et al., 2017). Moreover, scholars could observe if there are other potentially omitted variables in our empirical setting, such as heterogeneous traits of boundary spanners (Zhao & Anand, 2013), agents' career concerns and monitoring by external constituencies (Cabral et al., 2013). It is also important to emphasize that our findings are restricted to collaborations relying on performance assessments.

Finally, our framework and findings can inform managers of public, private, and nonprofit organizations including but not limited to the boundaries of CSCs. Our implications are useful to the myriad of partners interested in the assessment of their collaborative performance. Importantly, understating partners' efforts potentially harms the viability of long-term collaboration from the standpoint of actors on the transferring side. For instance, collaborations between consulting firms and their clients may also pose critical dilemmas if the latter decide to replicate best practices to organizational units that are not part of the original contract. Although we have argued that this dilemma is magnified in the context of CSCs, especially in cases where the recipient is a public organization, future research could also examine if our proposed practice transfer dilemmas are relevant in other organizational and industry contexts.

REFERENCES

Agranoff, R. (2007). *Managing within networks: Adding value to public organizations*. Georgetown University Press.

Allen, T. J., Tushman, M. L., & Lee, D. M. (1979). Technology transfer as a function of position in the spectrum from research through development to technical services. *Academy of management journal*, *22*(4), 694-708.

Alonso, J. M., & Andrews, R. (2019). Governance by targets and the performance of cross-sector partnerships: Do partner diversity and partnership capabilities matter?. *Strategic Management Journal*, *40*(4), 556-579.

Angrist, J. D., & Pischke, J. S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Auld, M. C., & Sidhu, N. (2005). Schooling, cognitive ability and health. *Health economics*, *14*(10), 1019-1034.

Banerjee, A. V., Duflo, E., & Glennerster, R. (2008). Putting a band-aid on a corpse: incentives for nurses in the Indian public health care system. *Journal of the European Economic Association*, *6*(2-3), 487-500.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., ... & Walton, M. (2017). From proof of concept to scalable policies: Challenges and solutions, with an application. *Journal of Economic Perspectives*, *31*(4), 73-102.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of management*, *17*(1), 99-120.

Becker, H. J., & Ravitz, J. (1999). The influence of computer and internet use on teachers' pedagogical practices and perceptions. *Journal of research on computing in education*, *31*(4), 356-384.

Behn, R. D. (2003). Why measure performance? Different purposes require different measures. *Public administration review*, *63*(5), 586-606.

Bloom, N., Brynjolfsson, E., Foster, L., Jarmin, R., Patnaik, M., Saporta-Eksten, I., & Van Reenen, J. (2019). What drives differences in management practices?. *American Economic Review*, *109*(5), 1648-83.

Bloom, N., Genakos, C., Sadun, R., & Van Reenen, J. (2012). Management practices across firms and countries. *Academy of Management Perspectives*, *26*(1), 12-33.

Bloom, N., Lemos, R., Sadun, R. & Van Reenen, J. (2015). Does Management Matter in Schools? *Economic Journal*, 125, 647–674.

Bloom, N., Lemos, R., Sadun, R., Scur, D., & Van Reenen, J. (2014). JEEA-FBBVA Lecture 2013: The new empirical economics of management. *Journal of the European Economic Association*, *12*(4), 835-876.

Bloom, N., Lemos, R., Sadun, R., Scur, D., & Van Reenen, J. (2016). International data on measuring management practices. *American Economic Review*, *106*(5), 152-56.

Bromiley, P., & Rau, D. (2014). Towards a practice-based view of strategy. *Strategic Management Journal*, *35*(8), 1249-1256.

Bruce, J. R., de Figueiredo, J. M., & Silverman, B. S. (2019). Public contracting for private innovation: Government capabilities, decision rights, and performance outcomes. *Strategic Management Journal*, 40(4), 533-555.

Bruhn, M., Karlan, D., & Schoar, A. (2010). What Capital is Missing in Developing Countries? *The American Economic Review: Papers & Proceedings*, *100*(2), 629–633.

Bryson, J. M., Crosby, B. C., & Stone, M. M. (2015). Designing and implementing cross-sector collaborations: Needed and challenging. *Public administration review*, *75*(5), 647-663.

Cabral, S. (2017). Reconciling conflicting policy objectives in public contracting: The enabling role of capabilities. *Journal of Management Studies*, *54*(6), 823-853.

Cabral, S., Lazzarini, S. G., & De Azevedo, P. F. (2013). Private entrepreneurs in public services: A longitudinal examination of outsourcing and statization of prisons. *Strategic Entrepreneurship Journal*, *7*(1), 6-25.

Cabral, S., Mahoney, J. T., McGahan, A. M., & Potoski, M. (2019). Value creation and value appropriation in public and nonprofit organizations. *Strategic Management Journal*, *40*(4), 465-475.

Caldwell, N. D., Roehrich, J. K., & George, G. (2017). Social value creation and relational coordination in public-private collaborations. *Journal of Management Studies*, *54*(6), 906-928.

Chatain, O., & Plaksenkova, E. (2019). NGOs and the creation of value in supply chains. *Strategic Management Journal*, *40*(4), 604-630.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *The Quarterly journal of economics*, *126*(4), 1593-1660.

Cohodes, S., Setren, E., & Walters, C. R. (2019). *Can successful schools replicate? Scaling up Boston's charter school sector* (No. w25796). National Bureau of Economic Research.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research, and evaluation*, *10*(1), 7.

Currie, G., & Procter, S. J. (2005). The antecedents of middle managers' strategic contribution: The case of a professional bureaucracy. *Journal of management studies*, *42*(7), 1325-1356.

Dahl, A., & Soss, J. (2014). Neoliberalism for the common good? Public value governance and the downsizing of democracy. *Public Administration Review*, *74*(4), 496-504.

Davis, J., Guryan, J., Hallberg, K., & Ludwig, J. (2017). *The economics of scale-up* (No. w23925). National Bureau of Economic Research.

Dobbie, W., & Fryer Jr, R. G. (2011). Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone. *American Economic Journal: Applied Economics*, *3*(3), 158-87.

Dobbie, W., & Fryer Jr, R. G. (2015). The medium-term impacts of high-achieving charter schools. *Journal of Political Economy*, 123(5), 985-1037.

Doh, J., Tashman, P., & Benischke, M. (2018). Adapting to grand environmental challenges through collective entrepreneurship. *Academy of Management Perspectives*, *33*(4), 450-468.

Dollinger, M. J. (1984). Environmental boundary spanning and information processing effects on organizational performance. *Academy of Management Journal*, *27*(2), 351-368.

Duflo, E., Hanna, R., & Ryan, S. P. (2012). Incentives work: Getting teachers to come to school. *American Economic Review*, *102*(4), 1241-78.

Dyer, J.H. and Hatch, N.W., 2006. Relation-specific capabilities and barriers to knowledge transfers: creating advantage through network relationships. *Strategic management journal*, *27*(8), pp.701-719.

Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. Society for industrial and applied mathematics.

Feldman, M. S., & Orlikowski, W. J. (2011). Theorizing practice and practicing theory. *Organization science*, *22*(5), 1240-1253.

Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative science quarterly*, *48*(1), 94-118.

Fryer Jr, R. G. (2014). Injecting charter school best practices into traditional public schools: Evidence from field experiments. *Quarterly Journal of Economics*, 129(3), 1355-1407.

Gatignon, A., & Capron, L. (2020). The firm as an architect of polycentric governance: Building open institutional infrastructure in emerging markets. *Strategic Management Journal*.

Gerrish, E. (2016). The impact of performance management on performance in public organizations: A meta-analysis. *Public Administration Review*, *76*(1), 48-66.

Hitt, M. A., Bierman, L., Shimizu, K., & Kochhar, R. (2001). Direct and moderating effects of human capital on strategy and performance in professional service firms: A resource-based perspective. *Academy of Management journal*, *44*(1), 13-28.

Imbens, G.W., & Wooldridge, J.M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*(1), 5-86.

Inoue, C. F., Lazzarini, S. G., & Musacchio, A. (2013). Leviathan as a minority shareholder: Firm-level implications of state equity purchases. *Academy of Management Journal*, *56*(6), 1775-1801.

Jacob, B. A., & Lefgren, L. (2003). Are idle hands the devil's workshop? Incapacitation, concentration, and juvenile crime. *American Economic Review*, *93*(5), 1560-1577.

Jarzabkowski, P., Kaplan, S., Seidl, D., & Whittington, R. (2016). On the risk of studying practices in isolation: Linking what, who, and how in strategy research. *Strategic Organization*, *14*(3), 248-259.

Kivleniece, I., & Quelin, B. V. (2012). Creating and capturing value in public-private ties: A private actor's perspective. *Academy of Management Review*, *37*(2), 272-299.

Klees, S. J. (2018). Liberia's experiment with privatising education: a critical analysis of the RCT study. *Compare: A Journal of Comparative and International Education*, *48*(3), 471-482.

Klein, P. G., Mahoney, J. T., McGahan, A. M., & Pitelis, C. N. (2010). Toward a theory of public entrepreneurship. *European Management Review*, *7*(1), 1-15.

Knudsen, E. I., Heckman, J. J., Cameron, J. L., & Shonkoff, J. P. (2006). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences*, *103*(27), 10155-10162.

Kroeger, A., & Weber, C. (2014). Developing a conceptual framework for comparing social value creation. *Academy of Management Review*, *39*(4), 513-540.

Leifer, R., & Delbecq, A. (1978). Organizational/environmental interchange: A model of boundary spanning activity. *Academy of Management Review*, *3*(1), 40-50.

Luo, J., & Kaul, A. (2019). Private action in public interest: The comparative governance of social issues. *Strategic Management Journal*, *40*(4), 476-502.

McGahan, A. M., Zelner, B. A., & Barney, J. B. (2013). Entrepreneurship in the public interest: Introduction to the Special Issue. *Strategic Entrepreneurship Journal*, *7*(1), 1-5.

Nelson, R. R., & Winter S. G. (1982) *An Evolutionary Theory of Economic Change*. Cambridge MA.

Peteraf, M. A. (1993). The cornerstones of competitive advantage: a resource-based view. *Strategic management journal*, *14*(3), 179-191.

Quelin, B. V., Cabral, S., Lazzarini, S., & Kivleniece, I. (2019). The Private Scope in Public–Private Collaborations: An Institutional and Capability-Based Perspective. *Organization Science*, *30*(4), 831-846.

Quelin, B. V., Kivleniece, I., & Lazzarini, S. (2017). Public-private collaboration, hybridity and social value: Towards new theoretical perspectives. *Journal of Management Studies*, *54*(6), 763-792.

Rahmandad, H., & Ton, Z. (2020). If Higher Pay Is Profitable, Why Is It So Rare? Modeling Competing Strategies in Mass Market Services. *Organization Science*.

Rangan, S., Samii, R., & Van Wassenhove, L. N. (2006). Constructive partnerships: When alliances between private firms and public actors can enable creative strategies. *Academy of Management Review*, 31(3), 738-751.

Ravitch, D. (2013). *Reign of error: The hoax of the privatization movement and the danger to America's public schools*. Vintage.

Romero, M., Sandefur, J., Sandholtz, W. (2017), 'Can Outsourcing Improve Liberia's schools?', *Centre for Global Development,* 7 September 2017.

Rivera-Santos, M., Rufin, C., & Wassmer, U. (2017). Alliances between firms and non-profits: A multiple and behavioural agency approach. *Journal of Management Studies*, *54*(6), 854-875.

Rufin, C., & Rivera-Santos, M. (2012). Between commonweal and competition: Understanding the governance of public–private partnerships. *Journal of Management*, *38*(5), 1634-1654.

Sadun, R., Bloom, N., & Van Reenen, J. (2017). Why Do We Undervalue Competent Management? *Harvard Business Review*, (October), 120–128.

Spiller, P. T. (2010). A tribute to Oliver Williamson: regulation: a transaction cost perspective. *California Management Review*, *52*(2), 147-158.

Stamper, C. L., & Johlke, M. C. (2003). The impact of perceived organizational support on the relationship between boundary spanner role stress and work outcomes. *Journal of Management*, *29*(4), 569-588.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic management journal*, *18*(7), 509-533.

Todorova, G., & Durisin, B. (2007). Absorptive capacity: Valuing a reconceptualization. *Academy of management review*, *32*(3), 774-786.

Valéro, V. (2015). Government opportunism in public-private partnerships. *Journal of Public Economic Theory*, *17*(1), 111-135.

Vermeulen, F. (2018). A basic theory of inheritance: How bad practice prevails. *Strategic Management Journal*, *39*(6), 1603-1629.

Villani, E., Greco, L., & Phillips, N. (2017). Understanding value creation in public-private partnerships: A comparative case study. *Journal of Management Studies*, *54*(6), 876-905.

Zhao, Z. J., & Anand, J. (2013). Beyond boundary spanners: The 'collective bridge'as an efficient interunit structure for transferring collective knowledge. *Strategic Management Journal*, *34*(13), 1513-1530.

FIGURES AND TABLES

FIGURE 1

Practice Transfer in CSCs: Conceptual Framework and Hypotheses

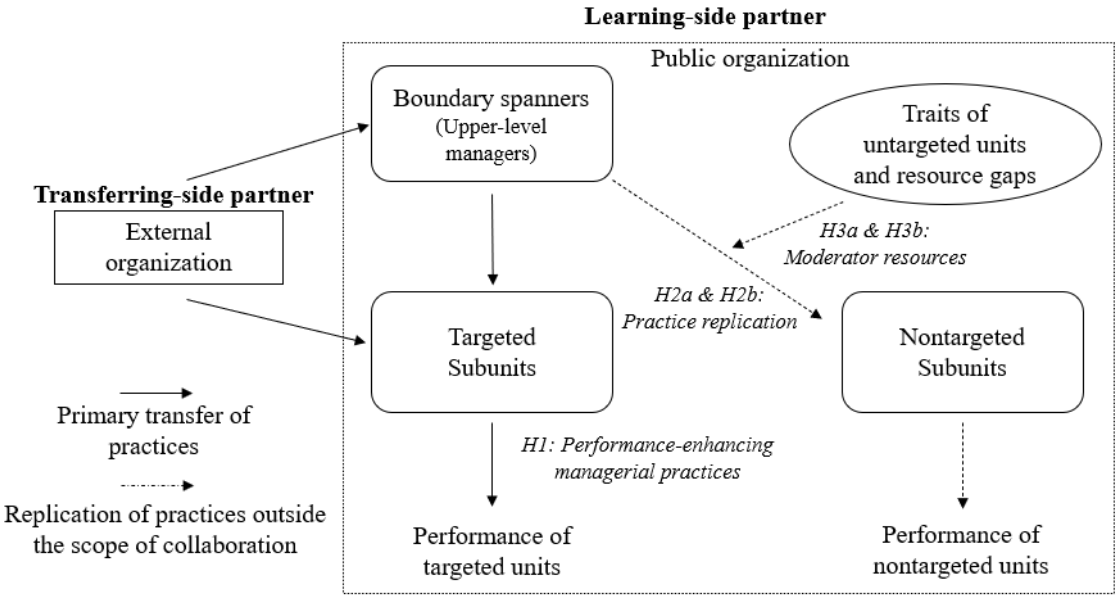
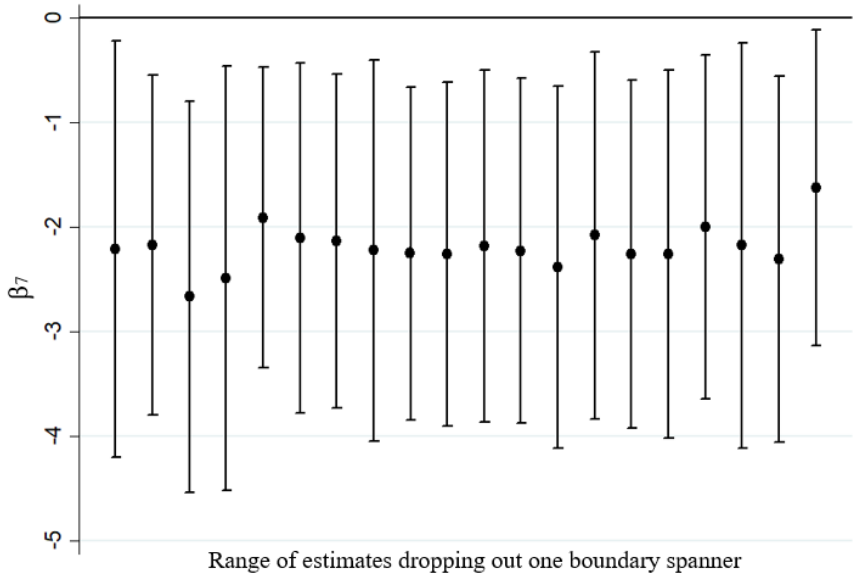


FIGURE 2

Leave-one-out Resampling Method: Dropping Boundary Spanners



Notes: The figure shows plots of our triple-difference point estimates (β7) on Math scores and corresponding 95% confidence intervals. The plots are presented from the regressions dropping the boundary spanners with the lowest internalization of practices index (the leftmost plot) to the highest (rightmost). The control variables include all school, teacher, and student features, as described in our methodology section and Table 1. Robust standard errors in parentheses (clustered at the boundary spanner level).

TABLE 1
Descriptive statistics

Variable	Description	Mean	Std. Dev.	Min	Max
Math Performance	Dependent variable (educational outcome; national standardized tests)	264.22	26.76	198.50	351.09
Language Performance	Dependent variable (educational outcome; national standardized tests)	259.68	25.40	176.00	320.90
Practice Internalization	Independent variable, main specification measured using a principal factor (per boundary spanner)	0.18	0.80	-1.62	1.29
Distance to closest targeted school	Measured according to geolocation (kilometers), per school	7.70	10.43	0.00	51.14
Distance to closest school	Measured according to geolocation (kilometers), per school	5.68	8.37	0.00	51.14
Teachers per student	Teachers/Students; proportion, per school	0.10	0.04	0.03	0.75
Computers per student	Computers/Students; proportion, per school	0.04	0.03	0.00	0.41
Teachers w/ college degree	With college degree/ All teachers; proportion, per school	0.96	0.06	0.53	1.00
Students' age	Aggregated per school	18.55	4.35	12.41	45.78
Students gender (male, %)	Male/(Male + Female); proportion, per school	0.49	0.06	0.30	0.81
Students' ethnicity (white, %)	White/All ethnicities - proportion, per school	0.21	0.17	0.00	0.84
Proportion of regular high school students	Regular high school students / all students; proportion, per school	0.58	0.27	0.00	1.00

Notes: The descriptive statistics include all 233 targeted and 161 non-targeted schools. These two groups were found to be similar in terms of their observable characteristics, as revealed by Kolmogorov-Smirnov and t-tests. Our dataset also includes binary variables for access to high-speed internet (81% have access), whether schools share rooms with any other school (6.6% share spaces) and state identification; Espírito Santo or Piauí (45% and 55% of the sample), used as fixed effects.

TABLE 2

Correlation matrix

	Variable	1	2	3	4	5	6	7	8	9	10	11	12
1	**Math Performance**	1											
2	**Language Performance**	0.914	1										
3	**Practice Internalization**	0.148	0.161	1									
4	**Distance to closest targeted school**	-0.066	-0.101	-0.074	1								
5	**Distance to closest school**	-0.019	-0.056	-0.073	0.847	1							
6	**Teachers per student**	-0.085	-0.072	-0.014	-0.090	-0.105	1						
7	**Computers per student**	0.133	0.114	-0.046	0.114	0.098	0.466	1					
8	**Teachers w/ college degree**	0.348	0.346	0.064	-0.179	-0.193	0.054	0.059	1				
9	**Students' age**	-0.327	-0.306	-0.044	0.177	0.156	-0.147	-0.125	-0.421	1			
10	**Students’ gender (male, %)**	-0.103	-0.115	0.045	-0.056	-0.042	0.042	-0.063	-0.183	0.177	1		
11	**Students' ethnicity (white, %)**	0.680	0.583	0.066	0.001	0.038	-0.064	0.165	0.269	-0.344	-0.036	1	
12	**Proportion of high school students**	-0.176	-0.173	-0.107	0.066	0.053	0.082	0.104	-0.018	-0.038	-0.290	-0.230	1

TABLE 3

Performance-enhancing managerial practices in CSCs (Baseline Hypothesis)

OLS clustered regressions	Math Performance		Language Performance	
	(1)	(2)	(1)	(2)
DV: School Performance	Only the school treatment	W/ controls	Only the school treatment	W/ controls
Hypothesized variables				
(H1) Targeted School # Post	4.23	3.51	4.45	3.62
	(0.95)	(0.89)	(1.08)	(1.03)
Targeted School	3.79	-1.45	3.74	-1.92
	(2.40)	(1.78)	(2.32)	(1.64)
Post (Post-treatment)	1.37	3.37	4.44	6.80
	(1.32)	(1.21)	(1.74)	(1.57)
Control variables				
Schools sharing spaces		-4.37		-6.36
		(3.65)		(4.09)
Distance to closest targeted school		-0.05		-0.09
		(0.12)		(0.11)
Distance to closest school		0.18		0.16
		(0.13)		(0.11)
School w/ high-speed internet		1.60		3.00
		(1.65)		(1.67)
Teachers per student		-18.80		-9.83
		(17.38)		(20.09)
Computers per student		23.04		17.82
		(25.69)		(28.22)
Teachers w/ college degree (%)		20.03		23.59
		(6.03)		(8.68)
Students' age		0.06		0.13
		(0.19)		(0.24)
Students' gender (male, %)		-27.56		-39.52
		(9.73)		(11.98)
Students' ethnicity (white, %)		59.29		34.24
		(13.16)		(10.34)
Proportion of high school students		2.31		0.77
		(3.14)		(2.88)
State FE (Espírito Santo)		24.37		26.44
		(4.65)		(4.54)
Constant	259.07	225.99	252.16	223.83
	(4.35)	(8.28)	(4.10)	(9.62)
Observations	1,564	1,564	1,564	1,564
R-squared	0.02	0.62	0.03	0.56

Notes: The table presents OLS regressions estimates explaining school performance regarding their student average Math and Languages scores. The difference-in-difference estimates refer to the average treatment effect. The control variables include all school, teacher, and student features, as described in our methodology section and Table 1.Specifications 1 and 2 differ only regarding inclusion of control variables. Analyses consider the full sample (targeted and nontargeted schools). Robust standard errors in parentheses (clustered at the boundary spanner level).

TABLE 4

Practice Internalization and Performance of Non-targeted Schools

OLS clustered regressions	Math Performance			Language Performance		
DV: Schools' Performance	(3) Continuous specification	(4) Binary specification	(5) Three-level specification	(3) Continuous specification	(4) Binary specification	(5) Three-level specification
Main specification: Continuous measure for main hypothesized variable (Practice Internalization)						
(H2a) Practice Internalization # Post	2.38			2.60		
	(1.37)			(2.09)		
Practice Internalization (continuous)	2.92			2.45		
	(1.77)			(2.10)		
Binary specification: Low/High Practice Internalization						
(H2a) Practice Internalization (High) # Post		5.57			8.81	
		(1.85)			(2.05)	
Practice Internalization (High)		3.60			2.44	
		(2.30)			(3.10)	
Three-level specification: Low/Medium/High Practice Internalization						
(H2a) Practice Internalization (Medium) # Post			1.44			-0.26
			(2.19)			(3.21)
(H2a) Practice Internalization (High) # Post			5.89			9.34
			(2.55)			(2.38)
Practice Internalization (Medium)			7.43			9.03
			(3.65)			(4.85)
Practice Internalization (High)			6.03			1.16
			(3.52)			(3.34)
Control variables						
Post (Post-treatment)	2.70	1.03	0.96	5.64	2.75	3.43
	(1.29)	(1.13)	(1.40)	(1.72)	(1.43)	(1.91)
Additional controls and FE	Y	Y	Y	Y	Y	Y
Constant	243.02	241.05	246.02	240.75	239.46	243.07
	(10.67)	(11.81)	(12.69)	(10.96)	(13.13)	(13.97)
Observations	638	638	638	638	638	638
R-squared	0.62	0.61	0.63	0.53	0.53	0.54

Notes: The table presents OLS regressions estimates explaining school performance regarding their student average Math and Languages scores. The difference-in-difference estimates refer to the average treatment effect. The control variables include all school, teacher, and student features, as described in our methodology section and Table 1. Specifications 3, 4 and 5 differ according to the measure for the main independent variable (practice internalization: either continuous, binary, or three-level variable). Analyses restricted to the subsample of nontargeted schools. Robust standard errors in parentheses (clustered at the boundary spanner level).

TABLE 5

Practice Internalization and the Assessed Performance of the External Partner

OLS clustered regressions	Math Performance		Language Performance	
	(6)	(7)	(6)	(7)
DV: Schools' Performance	Only differences in differences	W/ triple difference	Only differences in differences	W/ triple difference
Hypothesized variables				
(H2b) Practice Internalization # Targeted School # Post		-2.20		-1.68
		(0.80)		(1.18)
(H1) Targeted School # Post	3.46	3.85	3.56	3.86
	(0.86)	(0.77)	(1.00)	(1.01)
Control variables				
Targeted School	-1.24	-0.85	-1.69	-1.40
	(1.69)	(1.78)	(1.55)	(1.63)
Post (Post-treatment)	3.21	2.94	6.57	6.36
	(1.11)	(1.05)	(1.49)	(1.50)
Practice Internalization	1.54	2.70	1.54	2.39
	(1.66)	(2.07)	(1.75)	(2.36)
Practice Internalization # Post	0.99	2.38	1.41	2.47
	(1.04)	(1.35)	(1.40)	(1.94)
Targeted School # Practice Internalization		-1.82		-1.33
		(2.46)		(2.14)
Additional controls and FE	Y	Y	Y	Y
Constant	225.40	226.56	223.18	224.05
	(7.98)	(8.02)	(9.39)	(9.23)
Observations	1,564	1,564	1,564	1,564
R-squared	0.62	0.62	0.57	0.57

Notes: The table presents OLS regressions estimates explaining school performance regarding their student average Math and Languages scores. The triple-difference estimates refer to the average treatment effect. The control variables include all school, teacher, and student features, as described in our methodology section and Table 1. Specifications 6 and 7 differ only due to the additional triple difference estimator. Analyses consider the full sample (targeted and nontargeted schools). Robust standard errors in parentheses (clustered at the boundary spanner level).

TABLE 6

The Pivotal Role of Boundary Spanners in Replicating Practices

OLS clustered regressions	Math Performance		Languages Performance	
	(8)	(9)	(8)	(9)
DV: Schools' Performance	Human Capital	Physical Capacity	Human Capital	Physical Capacity
Hypothesized variables				
Main effect				
(H2a) Practice Internalization # Post	3.48	3.97	2.78	5.41
	(1.25)	(1.51)	(1.70)	(1.88)
Moderation effect				
(H3a) Practice Internalization # Human capital # Post	-2.57		0.13	
	(4.03)		(4.49)	
(H3b) Practice Internalization # Physical capacity # Post		-3.74		-5.87
		(2.67)		(0.97)
Control variables				
Practice Internalization	0.71	1.01	0.95	0.28
	(1.85)	(2.79)	(1.86)	(2.88)
Post (Post-treatment)	0.91	2.47	4.27	4.06
	(1.44)	(1.09)	(1.83)	(1.57)
Human capital (high)	-12.11		-13.13	
	(3.47)		(3.62)	
Practice Internalization # Human capital	4.80		2.35	
	(3.24)		(3.27)	
Human capital # Post	3.90		2.57	
	(3.19)		(3.74)	
Physical capacity (high)		0.19		-2.57
		(4.01)		(3.89)
Practice Internalization # Physical capacity		4.40		4.37
		(3.72)		(2.58)
Physical capacity # Post		-0.44		3.76
		(2.09)		(1.89)
Additional controls and FE	Y	Y	Y	Y
Constant	241.30	243.22	238.55	240.79
	(11.25)	(10.74)	(10.66)	(10.77)
Observations	638	638	638	638
R-squared	0.63	0.62	0.55	0.53

Notes: The table presents OLS regressions estimates explaining school performance regarding their student average Math and Languages scores. The difference-in-difference estimates refer to the average treatment effect. The control variables include all school, teachers, and student features, as described in our methodology section and Table 1. Specifications 8 and 9 differ as they investigate the human capital (H3a) or physical capacity (H3b) moderation effect. Analyses restricted to the subsample of nontargeted schools. Robust standard errors in parentheses (clustered at the boundary spanner level).

APPENDIX A

Boundary Spanners' Internalization of Practices: The External Partner Index

Index composition and score categories	Learning the Managers' Role	Internalization of Data-Driven Routines	Interactions with State Department of Education	Interactions with Supervisors	Monitoring Schools
Lowest score: 1 out of 4	Executes only documentation	No proper planning, mainly bureaucratic	None or few meetings	None or few meetings	None or few meetings
Score: 2 out of 4	Documents and presents some concerns regarding general pedagogical issues	Basic planning with no focus on students' learning	A meeting at least every 3 months	Meetings only when requested or less than monthly	A meeting at least every 3 months
Score: 3 out of 4	All previous concerns, and further mentions the importance of connections with schools	Planning focused on students' learning, but with no routines regarding data analyses	A meeting at least between 1 and 3 months	A monthly meeting at least is required	A meeting at least between 1 and 3 months
Highest score: 4 out of 4	All previous concerns, and emphasize the importance of close monitoring and feedback with schools	Planning focused on students' learning, implementing routines with data analyses	A monthly meeting at least	A meeting at least every two weeks is required	A monthly meeting at least

Notes: Index developed by the external partner (Unibanco Institute) to capture boundary spanners' internalization of received managerial practices. It consists of a score (from 1 to 4) covering 5 main components related to managers' learning, implementation and further replication of the so-called best managerial routines. The surveys were employed for all the managers of the 20 school districts from Espírito Santo and Piauí. We use as our main independent variable a score following a factor analysis of all 5 components, given as the principal factor. Robustness checks consider the maximum-likelihood instead of the principal axis factoring. Cronbach's Alpha reliability coefficient is 0.65.

APPENDIX B

Robustness to Factor Analysis Methodology – Maximum-Likelihood (Practice Internalization and Performance of Non-targeted Schools)

OLS clustered regressions	Math Performance			Language Performance		
DV: Schools' Performance	(3) Continuous specification	(4) Binary specification	(5) Three-level specification	(3) Continuous specification	(4) Binary specification	(5) Three-level specification
Main specification: Continuous measure for main hypothesized variable (Practice Internalization)						
(H2a) Practice Internalization # Post	1.76			2.00		
	(1.06)			(1.63)		
Practice Internalization (continuous)	2.58			2.39		
	(1.55)			(1.84)		
Binary specification: Low/High Practice Internalization						
(H2a) Practice Internalization (High) # Post		5.57			8.81	
		(1.85)			(2.05)	
Practice Internalization (High)		3.60			2.44	
		(2.30)			(3.10)	
Three-level specification: Low/Medium/High Practice Internalization						
(H2a) Practice Internalization (Medium) # Post			-0.68			-0.38
			(1.98)			(2.68)
(H2a) Practice Internalization (High) # Post			7.36			9.70
			(2.02)			(2.61)
Practice Internalization (Medium)			8.21			9.74
			(2.93)			(4.27)
Practice Internalization (High)			4.76			5.12
			(3.63)			(3.60)
Control variables						
Post (Post-treatment)	2.83	1.03	1.45	5.78	2.75	3.70
	(1.29)	(1.13)	(1.67)	(1.70)	(1.43)	(2.34)
Additional controls and FE	Y	Y	Y	Y	Y	Y
Constant	246.67	241.05	245.84	244.58	239.46	245.40
	(10.85)	(11.81)	(14.01)	(11.35)	(13.13)	(15.20)
Observations	638	638	638	638	638	638
R-squared	0.62	0.61	0.63	0.53	0.53	0.55

Notes: The table presents OLS regression estimates explaining school performance regarding their student average Math and Languages scores. These analyses consider the maximum-likelihood instead of principal axis factoring as robustness. The difference-in-difference estimates refer to the average treatment effect. The control variables include all school, teacher, and student features, as described in our methodology section and Table 1. Specifications 3, 4 and 5 differ according to the measure for the main independent variable (practice internalization: either continuous, binary, or three-level variable). Analyses are restricted to the subsample of nontargeted schools. Robust standard errors in parentheses (clustered at the boundary spanner level).

Robustness to Factor Analysis Methodology – Maximum-Likelihood
(Practice Internalization and the Assessed Performance of External Partner)

OLS clustered regressions	Math Performance		Language Performance	
	(6)	(7)	(6)	(7)
DV: Schools' Performance	Only differences in differences	W/ triple difference	Only differences in differences	W/ triple difference
Hypothesized variables				
(H2b) Practice Internalization # Targeted School # Post		-1.45		-1.02
		(0.66)		(0.90)
(H1) Targeted School # Post	3.47	3.65	3.58	3.70
	(0.87)	(0.81)	(1.01)	(1.02)
Control variables				
Targeted School	-1.23	-0.95	-1.67	-1.43
	(1.68)	(1.71)	(1.53)	(1.55)
Post (Post-treatment)	3.26	3.11	6.63	6.52
	(1.13)	(1.08)	(1.50)	(1.49)
Practice Internalization	1.29	2.47	1.37	2.37
	(1.33)	(1.74)	(1.42)	(2.01)
Practice Internalization # Post	0.80	1.72	1.22	1.87
	(0.80)	(1.04)	(1.15)	(1.50)
Targeted School # Practice Internalization		-1.83		-1.56
		(2.13)		(1.90)
Additional controls and FE	Y	Y	Y	Y
Constant	227.09	228.84	225.19	226.58
	(8.67)	(8.43)	(9.75)	(9.26)
Observations	1,564	1,564	1,564	1,564
R-squared	0.62	0.62	0.57	0.57

Notes: The table presents OLS regressions estimates explaining school performance regarding their student average Math and Languages scores. These analyses consider the maximum-likelihood instead of principal axis factoring as robustness. The triple-difference estimates refer to the average treatment effect. The control variables include all school, teacher, and student features, as described in our methodology section and Table 1. Specifications 6 and 7 differ only due to the additional triple difference estimator. Analyses consider the full sample (targeted and nontargeted schools). Robust standard errors in parentheses (clustered at the boundary spanner level).