The Irony of Accountability: How a Performance-Inducing Policy Reduces Motivation to Perform

Jose Eos Trinidad[1,2] (iD)

Abstract

Although sociological and organizational studies have focused on the influence of quantification on behavior, the author focuses on quantification's increasingly important consequences on well-being and motivation. Using the case of U.S. education, which has long relied on accountability policies, the author finds that attendance at schools with high-stakes accountability predicted lower student self-efficacy, that is, decreased task motivation and resilience as well as increased fear of failure—salient for low-income, urban, and public schools. These associations, however, did not spill over to social and life satisfaction dimensions of well-being. Taken together, these findings suggest the irony of accountability, where data used to induce performance may unintentionally reduce people's motivation to perform, particularly consequential in disadvantaged contexts. This article is an attempt to contribute to a broader theorization of quantification, affecting not only external behaviors and organizational structures but also internal personal dispositions. Finally, the article provides implications for the study of well-being, organizations, and education policy.

Keywords

quantification, accountability, education, well-being, organizations

Private and public organizations have often relied on data and quantitative metrics to promote efficiency and optimize performance (Marshall, Mueck, and Shockley 2015; McAfee and Brynjolfsson 2012; Schildkamp 2019). Businesses track workers' productivity, corporations measure customer satisfaction, and governments set up accountability systems and surveillance mechanisms, all of which have been growing through new technologies and greater confidence in big data and other algorithms (Brayne 2017; Ranganathan and Benson 2020; Sauder and Espeland 2009). Scholarship on organizations in general, and quantification in particular, has interrogated how behaviors and social relationships are structured, changed, and transformed by performance measures, cost-benefit analyses, interorganizational rankings, company ratings, and other forms of numerical assessment (Colyvas 2012; Martens and Niemann 2013; Mennicken and Espeland 2019). These studies highlight how different forms and processes of quantification influence individual behaviors, interpersonal relationships, organizational priorities, and interfirm competition (Kaivo-oja et al. 2015; Oswald et al. 2020). Although the intention is for objective measures to promote efficiency, critical studies have documented counterproductive effects in terms of corrupted measures, gaming strategies, destructive categorizations, and enhanced

inequities (Campbell 1979; Figlio and Getzler 2006; Sadowski 2019; Safransky 2020). In these studies, the theorization has centered on how the form, context, and scope of quantification influences behavioral and organizational consequences (Mennicken and Espeland 2019).

However, this focus on observable behavior may mask changes that take place within an individual, with aspects such as self-concept, motivation, and well-being. Although performance-inducing, quantitatively focused policies may promote behavioral compliance, it may come at the cost of people's mental health and personal motivation (Cäker and Siverbo 2018; Mitchell et al. 2018). Such is particularly important as institutions and companies are more attentive to organizational climate and personal well-being, highlighting that a positive environment can have longer term impacts on

[1]The University of Chicago, Chicago, IL, USA
[2]Ateneo de Manila University, Quezon City, Philippines

*Jose Eos Trinidad is now affiliated to University of California Berkeley, Berkeley, CA, USA

Corresponding Author:
Jose Eos Trinidad, University of California Berkeley, 2121 Berkeley Way, Berkeley, CA 94720, USA.
Email: jtrinidad@uchicago.edu

performance and productivity (Bache and Scott 2018; Warr and Nielsen 2018). Although well-being and motivation are key aspects of organizational life, quantification studies are relatively limited to qualitative research on work pressure and anxiety, and have yet to explore different dimensions of well-being and human development (Espeland and Sauder 2016; Goren 2012; Stevenson 2017). Thus, this research broadly asks how performance metrics affect well-being and intrapersonal dispositions.

Given the importance of well-being in organizations and the need to attend to how quantification regimes may potentially affect it, we investigate this with the case of a dominant quantification policy in a field that has long used this strategy: test-based accountability in U.S. education. Using students' test scores for accountability has been common in the United States, with its pinnacle being the 2001 No Child Left Behind policy. However, the form of accountability differs immensely, with some districts using these measures merely for information, others using them for school incentives, and others using them for consequential teacher evaluation and employment (Hursh 2007). In this research, we ask particularly how the practice of test-based, high-stakes teacher accountability in a school may be linked to different dimensions of student well-being, and how this link is sustained in different school contexts.[1] Although the impact of accountability on student performance is theorized to work because of how incentives are set up to encourage teachers to focus on instruction (Ingersoll and Collins 2017), this focus may inadvertently hurt the environmental climate as teachers concentrate on student performance rather than mastery, as students are pressured to perform well in narrowly set exams, and as a culture of fear is fostered for a school's potential poor performance (Conley and Glasman 2008; Schoen and Fusarelli 2008).

To preview the results, we found that accountability was linked to reduced personal efficacy but not to other dimensions of well-being, and that this finding was salient in relatively disadvantaged contexts such as low-income, public, and urban schools. With these results, we explore and discuss how quantitative mechanisms of control may be detrimental to personal motivation, especially in contexts of greater disadvantage. This finding is important in at least three ways. First, although studies have shown the supposed learning gains on students' performance with the advent of accountability policies (Figlio and Loeb 2011; Hanushek 2019), this present research suggests the potential collateral consequences for students' motivation and self-efficacy. Second, although qualitative studies have focused on teachers' resistance and fear of these teacher-focused accountability policies (Hallett 2010; Lingard 2021), this quantitative study suggests that such climate of demotivation and anxiety are also experienced by students. Third, although quantification studies have highlighted the consequences on observable behaviors and organizational structures (Espeland and Sauder 2016; Mennicken and Espeland 2019), this research motivates a new line of inquiry that attends to people's internal states.

Although the case of accountability in U.S. education may be just one particular institutional context, the study offers an important consideration for how specific forms of quantification within particular contexts are linked to internal personal dispositions, in addition to external observable actions. By opening a line of inquiry between sociology and human development, this study contributes to the literature on quantification in organizations, particularly regarding how the practice of accountability can be consequential to people's well-being and motivations. We seek to further the theorization of quantification as crucial not only for observable behaviors, interpersonal relationships, and organizational structures, but also for often invisible personal dispositions. Moreover, this research has implications for policy scholars studying the consequences of policy designs and educational researchers investigating and seeking to improve accountability systems.

Literature Review

Quantification in Organizational and Individual Performance

Organizations have long used data, quantitative information, and performance metrics to create efficiencies and to improve profit, service, ranking, or other indicators of success. For example, governments and states have been using censuses for thousands of years, corporations have used double-entry bookkeeping for hundreds of years, and profit-driven firms have been monitoring work and worker performance for decades, and all these continue with even greater sophistication and precision (Carruthers and Espeland 1991; Mennicken and Espeland 2019; Waring 2016). Studies have also kept pace in terms of investigating how the form, context, and supposed legitimacy of metrics associate with consequences for individual and organizational behaviors and relationships. For example, university rankings—widely debated yet still ever present—have consequences for university decisions and resources, staff work and pressure, student performance and inequality, donor relationship and support, and other aspects within and beyond the institution (Espeland and Sauder 2016; Hazelkorn 2015).

---

[1]The use of test-based accountability in the United States is widespread, and more frequently, standardized test scores are used to provide incentives and sanctions to schools. Although the negative impact of this practice is documented, we focus our analysis on test-based *teacher* accountability systems because of how this practice may create heightened pressures not only for the school organization but for individuals within the organization. Additionally, investigating teacher accountability provides greater analytic variation, as not all schools institute this practice, but many U.S. schools practice some form of school-based accountability.

Often, these supposedly objective metrics can provide information on organizational inefficiencies, support evidence-based decision making, evaluate program interventions, predict human behavior, institute mechanisms of accountability, and provide legitimacy to human discretion (Bovens, Goodin, and Schillemans 2014; Brayne 2017; Colyvas 2012). A sense of rationalization permeates both private and public organizations, and quantification through audits and evaluations has had an important role to play in this (Hwang and Powell 2009). Seen as critical in managing firms, using numbers has become so natural and institutionalized that organizations forego it at their peril. In general, data and information are thought as important components of organizational life and functioning, helpful in optimizing the individual and the firm's performance (Marshall et al. 2015; McAfee and Brynjolfsson 2012). Yet the use of these numbers can itself lead to detrimental consequences in terms of perverse incentives, gaming strategies, cheating, and categorical inequalities (Domina, Penner, and Penner 2017; Hibel and Penn 2020; Neal and Schanzenbach 2010). In a now classic formulation, Campbell (1979) wrote, "The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor" (p. 85). Such processes are seen in higher education, policing, health care, and government statistics, where incentives and penalties are created with quantitative indicators, and organizations fixate at the indicators rather than the improvement principles behind them (Brayne 2017; Hazelkorn 2015; Poku 2016).

However, it is not so much the mere presence of numbers or the process of quantification that affects behavioral and organizational changes, but its form, function, and context. For example, public organizations institute accountability practices that take many forms, with some measuring inputs through public spending transparency (Douglas and Meijer 2016) and others measuring outcomes such as students' test scores (Figlio and Loeb 2011). Even within similar forms of data collection, differences in function happen, as when test scores are used to provide information to the public (Diaz Rios 2020); to incentivize schools and determine failing ones (Bifulco and Schwegman 2020); or to measure, reward, and sanction teacher performance (Ingersoll and Collins 2017). More important, the context of quantification matters such that contexts that have not had such systems of accountability can experience turmoil when shifting to this practice (Hallett 2010). Taken together, the literature explores how different forms, functions, and contexts of quantification influence people and organizations' behavior, performance, and relationships. Yet the literature seems to ignore more intrapersonal processes potentially implicated in such systems.

Organizations and Subjective Well-Being

Although organizations are still concerned with people's performance, they are also expected to attend to people's subjective well-being, inclusive of meaningful work, positive climate, life satisfaction, and social belonging (Veenhoven 2008). Organizations have been instituting wellness programs, flexible policies, corporate counseling services, and other programs aimed at improving people's psychological and mental health (Kirk and Brown 2003; Ryan et al. 2021). Organizational studies have also been more interested in topics such as meaningful work, "calling" identities, and flexible work structures that affect subjective well-being (Bloom, Colbert, and Nielsen 2021; Bunderson and Thompson 2009; Gonsalves 2020).

Although the concept of well-being is not a major sociological concern, given the discipline's focus on social problems and observable behaviors, it is attracting greater scholarly attention for several reasons. First, individuals' well-being and sense of engagement are associated with positive outcomes such as lower turnover, higher task effort, and improved organizational performance (Shuck, Relo, and Rocco 2011). Second, and in contrast to the first, sociologists have interrogated the experience of stress in workplaces, particularly as it can affect burnout and interfere with an individual's nonwork life (Moen et al. 2016; Schieman, Glavin, and Milkie 2009). Finally, emotions, feelings and well-being are no longer just understood as individual psychical phenomena because of research documenting emotional contagion among groups of people (Coviello et al. 2014). These recent studies highlight the need and promise of attending to concepts of well-being and human development in organizational and sociological studies.

Studies on quantification have been attentive to some aspects of well-being. Espeland and Sauder (2016), for example, documented how university administrators, staff, and students feel pressure, and exhibit reduced or increased morale from the publication of law school rankings. Feelings of dehumanization—of being like cogs in a machine—mark workers who are under constant surveillance, such as U.S. transportation security officers monitored by a "system of cameras, testing, and observation" (Anteby and Chan 2018:8). Feelings of fear and anxiety are also routine experiences for public school teachers in disadvantaged schools, concerned about their evaluation and the metrics used to monitor their performance (Conley and Glasman 2008). Mainly qualitative in nature, these studies show how particular forms of quantification can influence reduced well-being.

These studies also provide a suggestion for how the supposed organizational improvement and efficiency may come at the expense of people's mental health—a critical intersection between the studies of organization, quantification, and social psychology. As individuals experience stress in their organizational environments, such experiences may lead to racial-ethnic and social class inequities in physical and mental health, stemming from racially and economically segregated jobs and organizational contexts (Thoits 2010). Thus, organizational processes of quantification can be considered a critical locus for investigating proximal processes that help reproduce these social inequities.

Although these studies suggest important consequences of quantification, three questions remain. First, what particular form of quantification is associated with decreased well-being? It is unlikely that every organizational use of data has implication for well-being, and so, we focus on a system that has proliferated in various public organizations: accountability (see Bovens et al. 2014). Second, are there specific aspects of well-being affected? Although most sociological studies mentioned in the previous paragraphs highlight pressures and anxieties, well-being is understood as multidimensional (Veenhoven 2008). Thus, it is important to nuance which specific aspects of well-being are linked to quantification processes. Third, how do we measure this systematically? Studies have often relied on rich qualitative cases, and our present research extends this by suggesting quantitative methods to see the link between quantification and well-being.

Case of U.S. Test-Based Accountability

Accountability policies in public organizations can span forms as varied as ethical, financial, democratic, and performance accountability—all of which require the use of data to hold account to the public (Fard and Rostamy 2007). One crucial and expansive case of accountability is that of U.S. education's use of tests to measure the performance of schools and teachers, which may include consequent incentives and sanctions in the form of pay increases, added resources, or threats of school closure or teacher dismissal (Figlio and Loeb 2011; Ingersoll and Collins 2017). Although studies have suggested positive effects of accountability in terms of increases in standardized test scores, academic performance, and postsecondary outcomes (Chiang 2009; Deming et al. 2016; Hanushek 2019), studies have also documented how different forms of accountability influence gaming strategies, school cheating, and increased inequalities (Diamond and Spillane 2004; Hibel and Penn 2020; Neal and Schanzenbach 2010). In the United States, pressures for schools to attain adequate yearly progress have provided an example for Campbell's (1979) thesis about the corruption that stems from indicators such as high-stakes tests. To attain these indicators, some schools have focused instruction on students at the threshold of passing exams or reclassified students as disabled to prevent them from taking standardized exams (Figlio and Getzler 2006; Jennings 2005)

In many of these studies, the focus has been mainly on students, teachers, or organizations' behavior and performance. Often, quantitative studies use difference-in-differences or regression discontinuity designs to compare the outcomes of students who were in schools that instituted test-based accountability and those students in schools that did not have it (Bifulco and Schwegman 2020; Chiang 2009; Dee and Jacob 2011). Multilevel and structural equation models, on the other hand, were more common in investigating teachers' behavioral responses to such systems (Hibel and Penn 2020; Ryan et al. 2017). Furthermore, qualitative studies focus on the experience of teachers and school administrators, often facing pressure and turmoil with new accountability systems (Diamond and Spillane 2004; Hallett 2010). Little, however, has been done in terms of quantitatively interrogating how school accountability predicts students' well-being, which has implications for how an organizational policy might affect not only behavior and performance but also internal personal processes.

We focus on student well-being rather than teacher well-being for two reasons. First, we want to know if the influence of accountability goes beyond those for whom it was initially directed (i.e., teachers and staff members). We hypothesize that these policies can influence school climate over and beyond the individuals monitored. It is an empirical test and extension of Golann's (2015) qualitative observation in a no-excuses charter school that the emphasis on raising test scores can undermine students' well-being as they closely monitor themselves, hold back their opinions, and defer to authority. We suggest that such sentiments can lead to reduced motivation and increased fear that can be evident in schools induced with accountability. Second, studies on accountability have focused on students' supposedly improved performance without interrogating their internal dispositions, thus necessitating the investigation into a potential collateral consequence of accountability.

A policy that uses test scores to measure student performance and incentivize teachers and schools may be potentially useful for motivating better instruction (Hong and Hong 2021; Means, Padilla, and Gallagher 2010), but it may also lead to a focus on comparability and heightened pressure to perform (Smeding et al. 2013). In other words, it may lead to a focus on quantified and quantifiable performance rather than the mastery that these forms of quantification supposedly measure. In this study, we explore the irony of accountability, where a policy that focuses on performance may unintentionally reduce organizational actors' motivation to perform. In our case, we test the following hypothesis:

> *Hypothesis 1:* Test-based teacher accountability is negatively associated with measures of student self-efficacy, inclusive of task motivation, resilience, and fear of failure.

However, we also investigate whether this form of accountability may also influence other aspects of well-being such as life satisfaction and social belonging. On the one hand, quantification may affect different domains of subjective well-being, given the increased stress and pressure felt within organizations (von der Embse et al. 2016). On the other hand, consequences may be more domain-specific as they are limited to aspects of performance rather than general well-being (Veenhoven 2008). For the case of accountability in schools, we test this latter hypothesis:

> *Hypothesis 2:* Test-based accountability is neither positively nor negatively associated with other domains of well-being such as life satisfaction and social belonging.
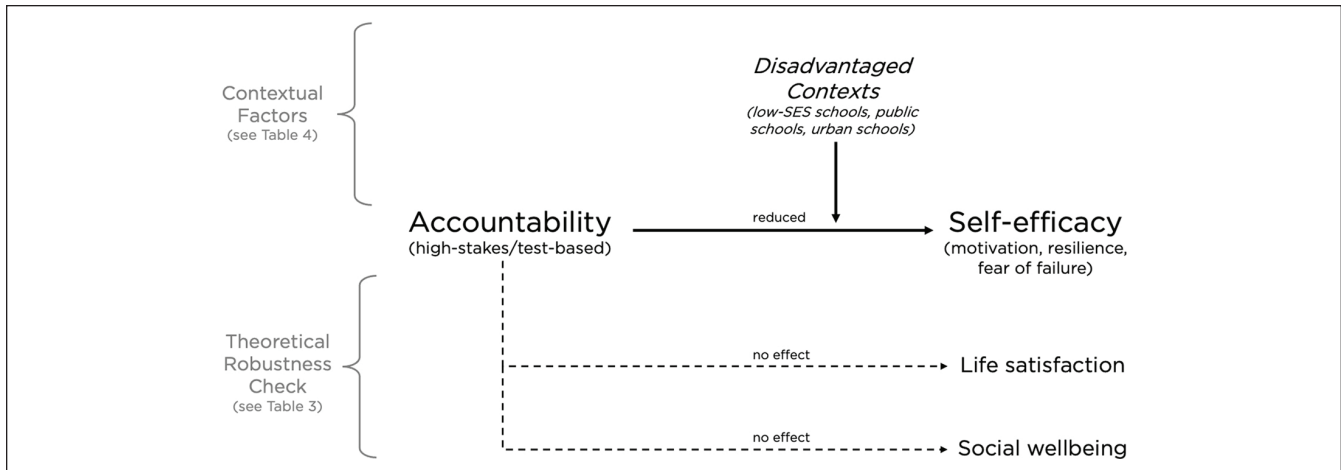
Figure 1. Conceptual model for the relationship of accountability with various dimensions of well-being.

More important, we try to investigate if this happens equally across different contexts. One interesting prospect is to see if the negative association between accountability and motivation was significant in disadvantaged contexts. For example, Diamond and Spillane (2004) showed how probationary and disadvantaged schools respond differently to accountability policies, as they felt more undue pressure from it. Thus, we hypothesize as follows:

Hypothesis 3: Test-based accountability remains negatively associated with measures of self-efficacy in disadvantaged contexts.

Figure 1 illustrates the conceptual model for this research as we investigate the influence of accountability on self-efficacy but not on life satisfaction and social belonging, and how this is salient for disadvantaged contexts.

Data and Methods

We used the U.S. data from the 2018 Programme for International Student Assessment (PISA), which had a two-stage stratified cluster sampling design, whereby a random sample of schools nationwide were first selected before selecting a random sample of 15-year-old students within the school (Rutkowski and Rutkowski 2016). Administered by the Organisation for Economic Co-operation and Development (OECD), this triennial program included data on students' standardized test scores in mathematics, reading, and science; measures of student socioeconomic status (SES) and well-being; students' beliefs and behaviors; and teachers' and principals' answers to survey questions. In 2018, PISA included more than 600,000 15-year-olds across 79 countries or administrative regions, but this research is initially limited to investigating test-based teacher accountability in the United States.

The analytic sample included 4,371 students (49 percent female) within 147 schools.[2] Missing values were imputed to preserve the cases where some variables were absent. However, 17 of the original 164 schools were not included because they did not have any school-level information, making imputation for school-level variables difficult. This is particularly important, as the main explanatory variable is school-level accountability, which we assume to be exogenous to student characteristics as logistic regressions show no association between accountability and student-level variables. Thus, although the original sample was designed to be nationally representative of 15-year-old students, we are circumspect with arguing that our research has preserved this representativeness. Nonetheless, among the 147 schools that were retained, we maximized the analytic power by imputing missing values (Royston 2004).

Measures

Test-based teacher accountability was a dichotomous variable coded 1 if a school used standardized student tests for evaluating and making judgement about teachers' effectiveness. This definition of accountability emphasizes the role of performance standards (i.e., tests), teachers' evaluation using these standards, and concomitant rewards and sanctions for passing or failing (Ingersoll and Collins 2017). Although this form of accountability can itself show up in different forms and with different gravity (e.g., some schools use it to fire teachers, whereas others include it merely for evaluation),

---

[2]In the results section, we provide descriptive statistical comparisons between the full U.S. sample and the analytic subsample and show that no statistically significant differences were present between the two in terms of the key variables being investigated. See Table 2.

Table 1. Dimensions of Student Well-Being.

	Description	Example Items
Personal efficacy dimension		
Task motivation	Measures how much the student finds satisfaction in working as hard as they can and in improving their performance.	"I find satisfaction in working as hard as I can"; "Part of the enjoyment I get from doing things is when I improve on my past performance"
Resilience	Reflects the student's perception of efficacy and ability to accomplish things, handle difficulties, and get through hard times.	"I usually manage one way or another"; "When I'm in a difficult situation, I can usually find my way out of it"
Fear of failure	Indicates the student's insecurity, doubt, or lack of confidence in their ability and talent.	"When I am failing, I worry about what others think of me"; "When I am failing, I am afraid that I might not have enough talent"
Satisfaction dimension		
Positive affect	Reflects the student's general emotional disposition and frequency of feeling positive emotions.	Frequency of feeling: happy, joyful, and cheerful (never, rarely, sometimes, always)
Life satisfaction	Measures the student's overall evaluation of their life.	"On a scale of 1–10, how satisfied are you with life these days?" (subsequently standardized)
Social dimension		
Belonging	Indexes the student's sense of belonging, integration, and connectedness with people in school.	"I make friends easily at school"; "I feel like I belong at school"
Perception of competition	Indicates the student's belief of the level and environment of competitiveness in a school.	"Students seem to value competition"; "It seems that students are competing with each other"

Note: All indices are standardized and composite measures, composed of Likert-type scale questions with item examples on the last column. Only life satisfaction was not a composite variable, as it includes only one question that has been subsequently standardized for this research.

our data cannot further disambiguate the specific form and magnitude of the school's practice of test-based teacher accountability. Nonetheless, we highlight that this indicator is the most appropriate one for our current investigation, given the data we have. This indicator also provides a more intuitive interpretation than if we had a continuous or categorical indicator of the type of accountability practiced in a school.

Student well-being was a multidimensional construct with different components such as belonging, life satisfaction, and task motivation. In PISA 2018, different indices were available to indicate students' subjective well-being (OECD 2019c). These warm likelihood estimates (WLEs) used item parameters from all students from equally weighted countries, but we standardized them so that the mean value of the US student population is zero with a standard deviation of one.[3] Each individual had a WLE in a particular well-being index, with higher scores indicating greater experience of the quality being measured (e.g., greater satisfaction). Unlike test scores that differ across contexts, these survey questions were similar across countries and the indices we

included exhibited within-country validity and cross-country comparability (OECD 2019a).

The present study used the PISA framework for the analysis of student well-being, particularly in terms of grouping the different indices into three aspects of well-being: personal efficacy, life satisfaction, and social well-being (OECD 2020). For personal efficacy, we included WLE measures of task motivation, resilience, and fear of failure. For the satisfaction dimension, we had measures of positive feelings and life satisfaction, while for the social dimension, we included students' sense of belonging and perception of competition. Table 1 presents the different dimensions of well-being, the included indices, their descriptions, and example survey items.

Control variables were included to isolate the potential relationship between teacher accountability and well-being. Student-level control variables, such as sex, SES, and standardized test scores, were added because these may independently predict subjective well-being. A dichotomous variable for sex was recoded with 1 indicating female and 0 male. The SES index was a composite variable from the 2018 PISA that included parents' highest educational attainment, parents' highest occupational status, and home possessions (OECD 2019b). This was standardized to have a mean of zero for an average OECD student. As students' level of academic achievement may be associated with subjective well-being, controls were added in terms of students' standardized mathematics, reading, and science test scores, which adhered to the general Rasch item response theory model and had been standardized for OECD countries with a mean of 500 and a standard deviation of 100 (OECD 2019b).

---

[3]Although the WLEs were computed for the OECD subsample, we opted to standardize this measure for the entire U.S. subsample to create more interpretable estimates. It must be noted, however, that our tests show little differences in estimates when using either the OECD or the U.S. WLEs. Results of the regressions with the OECD WLEs are available upon request. To learn more about the OECD's WLEs, we have included sources in the "References" section with the OECD as the author.

School-level control variables included the school type, school context, and mean SES index of the students. We had dichotomous variables for attendance in a public school (1 = public, 0 = private), defined as students not paying tuition for schools, and attendance in an urban school (1 = urban, 0 = nonurban), defined by PISA as being in a community of more than 100,000 people. The school-level SES index was created by averaging the SES indices of the students in a school.

Analytic Approach

We asked how test-based teacher accountability predicted different facets of student well-being, with a particular focus on its association with indicators of personal efficacy. In particular, we hypothesized that the presence of test-based accountability was associated with reduced motivation to perform. To test the hypothesis, we used hierarchical linear models to examine how school-level accountability practice may associate with student-level well-being indicators, using random effects on the school-level intercepts (Raudenbush and Bryk 2002). However, we recognized concerns about bias in the estimated coefficients, and so we used several analytic strategies to address these.

First, we leveraged the data and design as much as possible by adding theoretically motivated controls that may affect the relationship between test-based accountability and student well-being. Although accountability was not systematically associated with any of the control variables, some of these variables independently affected a number of well-being variables, and thus, we added these as covariates.

Second, we used theoretical and empirical robustness checks. As a theoretical robustness check, we tested the association between accountability and well-being dimensions other than personal efficacy. As accountability may induce pressures to perform, we hypothesize that this policy should only influence the personal efficacy dimension of student well-being but not necessarily be associated with the more general dimensions of well-being like life satisfaction and social belonging. As an empirical robustness check, we also disaggregated the results along different school classifications. As the strength and direction of the estimates may vary across heterogeneous populations, we provided estimates for higher and lower SES schools (bisected along the median, and divided into the highest and lowest quartiles), public and private schools, and urban and nonurban schools. While serving as robustness check, this investigation into heterogeneities also answers the question of contexts where accountability was predictive.[4]

---

[4]We were not concerned about where the association between well-being and accountability was most significant, as we were more concerned about whether the association remained significant in disadvantaged contexts. This explains our disaggregating the sample and not using interaction terms or moderation analysis.

Last, concerns about omitted variables may still be present even after these battery of controls, and theoretical and empirical robustness checks. Thus, we also quantified how much bias there would have to be due to omitted variables in order to invalidate the inferences we made (Frank et al. 2013). This method presents how many observations would have to be changed to zero effect in order to counter the inferences we made. The data are available from the PISA OECD Web site, and the clean data and code for the analysis are available at https://doi.org/10.17605/OSF.IO/9NY3Z.

Results

Descriptive statistics for the variables on student well-being, school type, demographics, and academic achievement are presented in Table 2. We compare the full sample with the analytic sample and show that despite the reduction of the sample to 147 schools, the variables in the analytic sample did not have statistically significant differences with the full sample. In the analytic sample, the student well-being measures had a mean of zero and a standard deviation of one, similar to the full sample. School composition is similar, with 93 percent identified as public schools and 40 percent situated in urban school areas. Demographic details and standardized test scores are also available.

Accountability and Personal Efficacy

Table 3 presents the association between teacher accountability and different dimensions of well-being. In the first to third columns, we investigated how teacher accountability predicted three dimensions of personal efficacy, hypothesizing that the school practice of accountability was negatively linked with these well-being dimensions.

After controlling for student background, academic achievement and school type, we found that teacher accountability was negatively associated with task motivation and resilience, and positively associated with fear of failure, all statistically significant at $p < 0.01$. Substantively, we find that being in a school with test-based teacher accountability was associated with a reduction in students' task motivation and resilience by a tenth of a standard deviation, and an increase in fear of failure by the same magnitude. To put this in perspective, the median effect size for education studies with samples larger than 2,000 students is 0.03, and effects for charter schools are often below a tenth of a standard deviation (Betts and Tang 2018; Kraft 2020). This suggests that the association is not only statistically significant but also substantively meaningful.

Taken together, these results suggest that compared with students in schools at which test scores are not used to evaluate teachers, those students in schools with such accountability practice, on average, (1) showed lower motivation to perform or improve in tasks, (2) perceived themselves to be less able to accomplish tasks or handle

Table 2. Summary Statistics.

	Full Sample			Analytic Sample		
	Mean	SD	Nonmissing Count	Mean	SD	Nonmissing Count
Student well-being measures						
Task motivation	.00	1.00	4,596	.00	.98	4,161
Resilience	.01	.99	4,623	−.01	.99	4,185
Fear of failure	.02	.99	4,626	.02	.99	4,190
Positive affect	−.01	1.00	4,594	.00	.99	4,160
Life satisfaction	−.01	1.00	4,651	−.01	1.00	4,212
Belonging	.01	1.00	4,619	.01	1.00	4,182
Perception of competition	.01	1.00	4,548	.01	1.00	4,124
School-level variables						
Accountability	.41		147	.41		147
Public school	.93		148	.93		144
Private school	.07		148	.07		144
Urban school	.41		149	.40		145
Nonurban school	.59		149	.60		145
Individual-level variables						
Female	.49		4,838	.49		4,371
SES index	.11	1.02	4,767	.14	1.01	4,307
Math test score	477.92	91.99	4,838	479.88	91.38	4,371
Reading test score	505.00	108.27	4,838	507.36	107.06	4,371
Science test score	501.84	98.04	4,838	503.62	97.28	4,371

Source: Programme for International Student Assessment 2018 (United States).
Note: The measures represent the mean, standard deviation, percentage belonging to a specific gender or school category, and number of nonmissing observations. The analytic sample is limited to schools that have information regarding their practice of accountability. From the table, we show that there were no statistically significant differences on the key variables between the full and analytic samples. SES = socioeconomic status.

Table 3. Estimates of Hierarchical Linear Models of Teacher Accountability on Student Well-Being.

	Personal Efficacy Dimension			Satisfaction Dimension		Social Dimension	
	Task Motivation	Resilience	Fear of Failure	Positive affect	Life Satisfaction	Belonging	Competition
Key variable							
School accountability	−.10** (.03)	−.11** (.04)	.10** (.04)	−.01 (.03)	.01 (.04)	.00 (.03)	.03 (.03)
Covariates at the student level							
Female	.15*** (.03)	.02 (.03)	.38*** (.03)	.02 (.03)	−.18*** (.03)	−.12*** (.03)	.03 (.03)
SES index	.08*** (.17)	.11*** (.02)	.03 (.02)	.12*** (.02)	.15*** (.02)	.12*** (.02)	.12*** (.02)
Covariates at the school level							
Public school	−.18* (.08)	.08 (.09)	.04 (.08)	−.16 (.08)	−.15 (.09)	−.32*** (.08)	.04 (.08)
Urban school	−.05 (.03)	−.02 (.04)	.00 (.04)	−.02 (.03)	−.10** (.04)	−.02 (.03)	−.19*** (.04)
School mean SES index	−.12** (.04)	−.16*** (.04)	.17*** (.04)	−.08* (.04)	−.14** (.04)	−.08* (.08)	−.01 (.04)
Intercept	−.65*** (.09)	−.14 (.09)	1.09*** (.10)	.44*** (.11)	.68*** (.12)	−.52*** (.11)	.21* (.09)
Controls for student achievement	Yes	Yes	Yes	Yes	Yes	Yes	Yes
% variance explained at level 1	4.12	2.03	6.15	2.05	2.05	2.03	2.07
% variance explained at level 2	30.56	15.97	82.27	23.44	26.53	39.51	43.75

Source: Programme for International Student Assessment 2018 (United States) (*n* = 4,371 students, 147 schools).
Note: The table presents regression coefficients from hierarchical linear models regressing student-reported well-being on the presence of test-based teacher accountability system in school, along with individual-level and school-level predictors as controls. In addition, all models control for students' achievement in standardized tests for reading, mathematics, and science. Values in parentheses are standard errors. SES = socioeconomic status.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 4. Estimates of Hierarchical Linear Models of School Accountability on Student Well-Being, Disaggregated by School Classification.

	Personal Efficacy Dimension			Satisfaction Dimension		Social Dimension	
	Task Motivation	Resilience	Fear of Failure	Positive Affect	Life Satisfaction	Belonging	Competition
A. School disadvantage median							
Higher SES schools (*n* = 73)	−.10* (.05)	−.10* (.05)	.07 (.05)	−.04 (.04)	.05 (.05)	−.00 (.04)	−.01 (.05)
Lower SES schools (*n* = 74)	−.10* (.05)	−.11* (.05)	.11* (.05)	.02 (.05)	−.05 (.05)	.02 (.04)	−.01 (.04)
B. School disadvantage quartile							
Top SES quartile (*n* = 37)	−.13 (.07)	−.14 (.08)	.07 (.06)	.04 (.06)	.01 (.07)	−.07 (.06)	.02 (.06)
Bottom SES quartile (*n* = 37)	−.13 (.07)	−.17* (.07)	.14* (.07)	−.01 (.07)	−.09 (.06)	.01 (.07)	−.09 (.07)
C. School type							
Public schools (*n* = 137)	−.11** (.03)	−.10** (.04)	.10** (.03)	−.01 (.03)	−.01 (.04)	.01 (.03)	−.01 (.03)
Private schools (*n* = 10)	.13 (.17)	.00 (.28)	−.02 (.17)	.03 (.16)	.35 (.28)	.05 (.22)	−.51* (.22)
D. Urban context							
Urban schools (*n* = 58)	−.15** (.05)	−.16** (.06)	.13* (.05)	−.05 (.06)	.03 (.06)	.03 (.04)	.03 (.06)
Nonurban schools (*n* = 89)	−.07 (.04)	−.06 (.04)	.07 (.05)	.01 (.04)	−.02 (.05)	.00 (.04)	−.05 (.04)

Source: Programme for International Student Assessment 2018 (United States) (*n* = 4,371 students, 147 schools).
Note: The table presents regression coefficients from hierarchical linear models regressing student-reported well-being on the presence of test-based teacher accountability system in school, inclusive of the same student- and school-level covariates in Table 3. Panels A to D disaggregate the sample by school disadvantage, school type, and urban context. In panel A, the schools were divided along the median school-level socioeconomic status (SES) index, while in panel B, we limited it to the top and bottom quartile schools according to school-level SES. Values in parentheses are standard errors.
*p < .05. **p < .01.

difficulties, and (3) experienced greater fear and anxiety of failing. Such associations are robust and substantively meaningful, even after considering students' academic performance and social background.

Accountability, Life Satisfaction, and Social Well-Being

Although test-based accountability may be negatively associated with their motivation to perform, could it also be associated with dimensions of satisfaction and social well-being? To answer this, Table 3 includes tests of association between accountability and the *satisfaction* and *social* dimensions of well-being. Controlling for the same covariates, the practice of accountability in schools was neither associated with having positive feelings nor was it associated with greater or less life satisfaction. The point estimates also show miniscule effect sizes of one percent of a standard deviation. Similarly, being in a school that had test-based accountability was not associated with students' sense of belonging or perception of competition. These results suggest that students' sense of life satisfaction and social well-being were unaffected by their being in a school with test-based teacher accountability practices instituted.

Heterogeneity of Associations

The previous results suggest that test-based teacher accountability was negatively associated with student's personal efficacy and motivation, but not with other dimensions like satisfaction and social well-being. The inclusion and the significant estimates of the three indices under the personal efficacy dimension provide one source of confirmation that the associations were not spurious. Moreover, the comparison with the other dimensions of well-being provides evidence that accountability only predicted personal efficacy, without spilling over to other well-being dimensions such as life satisfaction, positive affect, or social belonging. However, the strength and direction of the association between teacher accountability and self-efficacy may differ according to heterogeneous populations.

Thus, Table 4 provides estimates from models that disaggregated the sample by different school classifications. As students in more economically disadvantaged schools may show lower task motivation (Destin et al. 2019), we first bisected the sample of schools between those above and below the school SES median. Despite this division, however, accountability was still linked with lower levels of task motivation and resilience, and higher levels of fear of failure for both groups. It had similar substantive estimates of reduction or increase by around a tenth of a standard deviation. Panel A in Table 4 presents these significant coefficients, except for the coefficient for fear of failure in higher SES schools.

However, when we investigated patterns for the highest and lowest quartiles of school-level SES, we found that the coefficients for personal efficacy were only significant for students in the bottom quartile schools. Panel B in Table 4 shows

that accountability was associated with a reduction in student resilience by 17 percent of a standard deviation, and an increase in fear of failure by 14 percent of a standard deviation, both significant at $p < 0.05$. This suggests reduced efficacy dimensions for students in accountability-induced low-income schools, although we also note that the effect sizes for the top quartile are lower by just a few percentage points.

Another potential source of heterogeneity was in terms of the difference between public and private schools, with the potential of greater disadvantage in public institutions (Benveniste, Carnoy, and Rothstein 2013; Coleman, Hoffer, and Kilgore 1982). Panel C in Table 4 shows that in public schools, students in accountability-induced schools had reduced task motivation and resilience, and increased fear of failure as compared with students in schools that did not have this practice. The estimates were by a tenth of a standard deviation. These associations, however, were not statistically or substantively significant for private schools. Although some may assume that most public schools use test-based teacher accountability, our sample showed greater variability as only 39.71 percent of public schools practiced this form of accountability. Thus, the significant associations for public schools remain robust. However, we are circumspect with interpreting the lack of statistically significant association among private schools because of the small sample of schools (10 schools) and students within them (202 students).

Finally, urban and nonurban schools in the United States have critically different characteristics and outcomes, with urban schools and districts more racially and economically segregated (Frankenberg 2009; Trinidad 2020), leading us to test the estimates between these two subgroups. Panel D in Table 4 highlights that the practice of accountability in urban schools was significantly associated with lower task motivation and resilience, and higher levels of fear of failure. In particular, it was associated with decreases of 15 percent of a standard deviation in terms of task motivation and resilience, and an increase of 13 percent of a standard deviation in terms of fear of failure. However, such associations were absent for nonurban schools. Although questions may arise regarding the possible distributional difference in accountability practices between urban and nonurban settings, we find that 40 percent of both urban and nonurban schools had this form of accountability practice instituted.

Recognizing concerns about potential omitted variables invalidating our estimates, and adding to the battery of robustness checks that we have already instituted, we drew on the work of Frank et al. (2013) and quantified how much bias there would have to be to invalidate our inference. Using this analysis, we found that for the personal efficacy dimension, 34.65 percent of the estimate would have to be due to bias to conservatively invalidate our inference. Substantively, this means that more than 1,500 observations would have to be replaced with cases of zero effect to invalidate our inferences.

Discussion

Studies on quantification and data in organizations have often focused on the way these measures influence individual behaviors as well as organizational processes and priorities (Mennicken and Espeland 2019). Often, public organizations institute accountability measures to drive better performance and behavior among their staff members, hold leaders responsible for organizational efficiency, and legitimize the public enterprise (Arthur 2017; Bovens et al. 2014; Wilson 2011). However, studies have also shown how the focus on performance and the pressure to attain certain metrics have led to gaming strategies and outright cheating (Figlio and Getzler 2006; Hibel and Penn 2020; Neal and Schanzenbach 2010). Although many studies focus on the impact of organizational metrics and quantification on people's behavior, much less is known about how such a performance-inducing practice as accountability can influence well-being. Using the case of schools in the United States, we find that accountability practice was associated with reduced motivation to perform and increased fear of failure, and that the associations were salient in contexts of relative disadvantage: low-income, public, and urban schools. We suggest three key insights that have implications for the study of quantification in organizations in general, and the study of accountability practices in schools in particular.

First, policies that put a focus on and are oriented toward performance may create environments harmful to one's sense of personal efficacy. Here we highlight the irony of accountability, where a policy that was supposed to encourage better performance might unintentionally reduce people's motivation to perform. Noting the difference between actual performance and motivation to perform, we suggest that accountability policies may improve the *performance* of different tasks, such that in schools, teachers are putting more effort on instruction and students are getting higher scores in standardized examinations (Figlio and Loeb 2011; Rouse et al. 2013). But such improved performance may come at the expense of the *social climate*, which could then put a strain on certain dimensions of people's well-being (Mitchell et al. 2018). In the case of U.S. high schools, we find that the practice of using tests to bring about accountability was associated with lower levels of student task motivation and resilience, and higher levels of fear of failure, suggesting that the school's focus on quantifying performance may significantly influence individuals' feelings of self-efficacy.

Although past studies on quantification and accountability have mainly focused on observable individual and organizational performance (see Mennicken and Espeland 2019 for a review), we suggest that attention be given as well to dimensions like well-being. In organizational studies, the emphasis has been on how pressures can lead to changed behaviors such as focusing resources on measurable outcomes, improving

work efficiency, or making decisions on the basis of data (Espeland and Sauder 2016; Marshall et al. 2015; Nielsen and Riiskjær 2013). In schools in particular, studies have focused on how the presence of accountability are beneficial for short- and long-term student outcomes (Chiang 2009; Deming et al. 2016; Figlio and Loeb 2011). The present research suggests that measures of accountability are not just linked to observable performance; they are also related to internal motivation to perform. Although performance may be positively affected by a focus on accountability through increased efficiency and clearer responsibility, *well-being* and *human development* may be sacrificed by-products of this quantified logic.

Second, although research suggests that the focus on quantification may influence greater work pressure and reduced morale, our research suggests that not all facets of well-being are affected by this focus. In particular, we find that the associations were mainly in terms of the self-efficacy dimension of well-being and not necessarily on the life satisfaction and social dimensions. Although previous qualitative studies have been more domain general in their assessment of quantification and well-being, suggesting wholesale psychological harm of performance pressures (Gurova and Piattoeva 2018; Mitchell et al. 2018), our quantitative study highlights the need to be more specific on which domains of well-being are linked to these pressures. We highlight how pressures affect well-being dimensions that directly relate with performance and do not necessarily spillover to other dimensions of one's personal well-being.

This insight offers two related contributions to the study of quantification in organizations. On the one hand, it provides an impetus to examine the *specific* types of psychological outcomes related to performance pressures, quantification processes, and accountability measures. It offers a stimulus to investigate not only broad concepts of stress and fatigue but also finer grained ones of motivation, satisfaction, and social belonging. On the other hand, the study also presents what may be among the first studies to empirically analyze with a national sample how accountability influences well-being. In line with this, we suggest the potential in studying how different types of performance management measures may affect various dimensions of people's well-being across different sectors and industries. Although quantification of performance has become a widespread phenomenon (Mennicken and Espeland 2019), variations in the manner of quantification may also variably affect different dimensions of people's well-being.

Third, the influence of quantification does not operate similarly across organizational types and systems. We find that the associations were particularly salient in contexts of relative disadvantage: among schools in the lowest SES quartile, students placed in an accountability-induced school were more likely to have reduced task motivation and resilience, and increased fear of failure than students in a school that did not institute this accountability practice, net of their academic performance and sociodemographic background.

The same is true for students in public school and urban settings, suggesting that the context matters for how quantification may influence well-being. This aligns with qualitative research, which found that disadvantaged contexts can be further disadvantaged because of mechanisms of control that narrowly focus on compliance of policy demands (Diamond and Spillane 2004).

More generally, this highlights how the salience of quantification and performance metric pressures may operate differently in different contexts. In particular, the harms of accountability and quantification may be more visible in contexts of disadvantage, and in environments where the metricalization of performance were previously absent (Hallett 2010; Stevenson 2017). For example, the presence of accountability systems may highlight and exacerbate the already failing system in an organization, such that this reduces morale of its actors (von der Embse et al. 2016). Similarly, organizations that did not previously have such system of evaluation and surveillance may meet resistance from actors that have been used to the status quo (Hallett 2010). Thus, accountability and quantification per se may not immediately be associated with lower motivation, as historical and institutional contexts matter significantly for such association to be discernible. Taken together, these concepts suggest the potential harms of accountability in terms of self-efficacy and motivation, made more salient by the context of disadvantage, but not necessarily consequential for other dimensions of well-being.

Implications for Intersecting Organization Studies and Social Psychology

Although research on organizations have often focused on individual behaviors and organizational structures, this research suggests the potential for interrogating the intersection of organizations and human development. Concepts such as individual identity, well-being, motivation, risks, and vulnerabilities need to be understood within more classical theorizations of organizations because of how these internal personal factors are recognized as antecedents of organizational performance (Schieman et al. 2009; Shuck et al. 2011). Although policies and processes such as quantification and accountability are consequential for organizational behaviors, relations and structures, a more comprehensive theorization needs to also attend to the often implicit and invisible aspects of people's dispositions and motivations. Although organizational actors may act efficiently and rationally, they may also resist or experience stress when policies have counterproductive consequences for their motivation.

Our study highlights the need for research in other fields and firms to interrogate this critical intersection of organizations and human flourishing, an important nexus for sociological research. We argue that it is critical for sociologists to study this because of how marginal organizational improvements can come at the expense of individual human well-being. Such

collateral consequence may also be consequential for social and racial inequities when quantified logics are more often directed to, and more negatively affect, jobs and organizations for racial minorities and working-class individuals (Thoits 2010). Thus, our research reemphasizes the need for studies that look at social psychological effects in organizations, and how such can contribute to social reproduction and inequality.

Implications for Education Policy and Inequality

Historically, accountability policies sprang from efforts to deal with racial inequality and the poor performance in urban schools. In a way, it was originally intended to address social reproduction and stratification. However, decades of social scientific research have also shown the unintended consequences of well-intentioned organizational policies and the use of performance indicators. In this research, we show that facets of these high-stakes performance measures may contribute to unproductive pressure and demotivation among school staff and students. For example, the nature of these exams being used for teacher evaluation, incentives, and sanctions can create fear or resistance to testing regimes (Lingard 2021).

Given these findings, we argue for studies that investigate better forms of quantification, greater variety of performance indicators, and more supportive systems of accountability. Although it has not been the purview of this research to explore *what* these alternatives are, we show aspects that policy makers need to be attentive of. First, they must be sensitive not only to observable improvements in test score outcomes but also to people's lived experiences and well-being, as Campbell's (1979) law may manifest in improved indicators but sacrificed social climates. Second, these policies have to interrogate how inequities may be exacerbated because of schools more negatively affected by quantified logics. Third, they should provide more holistic indicators that are able to inform instruction and practice, and not simply rely on once-a-year examinations. By attending to these facets and potential pitfalls, policies that use quantitative social indicators may increase its potential for effectiveness and equity.

Limitations

Although this study has important contributions and implications for the study of organizations, quantification, and education, we acknowledge certain limitations as well. First, the study is an exploratory study into how a particular form of quantification is associated with well-being. It provides a suggestion or stimulus for investigation rather than a confirmation of a theory. In a way, it seeks to start contributing to a theory for how organizational forms and processes may influence internal personal dispositions.

Second, the study is limited to a particular form and context of quantification, i.e., test-based teacher accountability in U.S. education. Although the case is unique and other organizations have little overlap with school processes, the suggested idea of the irony of accountability may be discernible in other forms such as organizational surveillance that lower actors' morale, trust, and efficacy (Sewell and Barker 2006; Sewell, Mol, and Taskin 2019). Thus, the study encourages exploration into how the form of organizational quantification may be linked to aspects of well-being and human development.

Third, a number of methodological limitations has to be noted, including the use of cross-sectional data, elimination of observations with no school-level variables, and potential bias from unobserved confounding. As the PISA has a repeated cross-sectional design, we cannot follow students' trajectories and thus, we are limited by the associational patterns discerned in the data set. We were also limited to using a dichotomous measure for accountability, which may make our estimates more conservative, as we did not account for the extent of such practice but only if schools had that practice. Although PISA had a nationally representative U.S. sample, our analytic sample was not representative because of our exclusion of observations without any school-level variable. Although concerns may arise for this necessary exclusion, we did not find any statistically significant difference between the full and analytic samples, providing some assurance that the noninclusion of some schools had little deleterious consequences. Other factors also influence the presence of accountability and students' well-being, particularly neighborhood contexts and race. Given the importance of race in school accountability dynamics, this omission is an important one to note. However, such variables were absent in the PISA survey, and so the study only relied on controlling academic achievement and demographics.

Finally, the experience of being in environments pressured by accountability and quantified logics is best understood through qualitative studies. The present research hopes to encourage further research into the processes within organizations that affect people's motivation and morale, as well as attend to heterogeneities within organizational actors' responses to quantification. Although the present study suggests the harms (and irony) of accountability, grounded studies may find out why, how, and for whom this happens.

Conclusion

Systems and regimes of quantification permeate organizations through accountability practices, program evaluations, periodic monitoring, and other forms of impact measurement. Many studies focus on the way these processes change organizational strategies and structures, individual behaviors and actions, and social relationships and consciousness. Our study attempts to show how quantification may likewise influence internal personal processes, often invisible and sometimes in opposition to what is actually observed. More specifically, we suggest the irony of accountability, understood as how a policy that was supposed to induce

performance may inadvertently reduce organizational actors' motivation to perform. Using the case of test-based teacher accountability in the United States, we find how this practice in a school was linked with reduced task motivation and resilience, and heightened fear of failure, apparent in low-income, public, and urban schools. Although the case is specific to the educational institution, potential processes may be discernible for organizations where data have high stakes. We thus suggest the irony with how the process of quantification may lead to better observable behaviors and outcomes but may come at the expense of people's motivation and self-efficacy.

Acknowledgments

ORCID iD

Jose Eos Trinidad https://orcid.org/0000-0001-9986-8683

References

Anteby, Michel, and Curtis K. Chan. 2018. "A Self-Fulfilling Cycle of Coercive Surveillance: Workers' Invisibility Practices and Managerial Justification." *Organization Science* 29(2):247–63.

Arthur, James. 2017. *Policy Entrepreneurship in Education: Engagement, Influence and Impact*. New York: Routledge.

Bache, Ian, and Karen Scott. 2018. "Wellbeing in Politics and Policy." Pp. 1–22 in *The Politics of Wellbeing : Theory, Policy and Practice*, edited by I. Bache and K. Scott. Cham, Switzerland: Springer International.

Benveniste, Luis, Martin Carnoy, and Richard Rothstein. 2013. *All Else Equal: Are Public and Private Schools Different?* New York: Routledge.

Betts, Julian R., and Y.Emily Tang. 2018. "A Meta-analysis of the Effect of Charter Schools on Student Achievement." San Diego, CA: San Diego Education Research Alliance.

Bifulco, Robert, and David J. Schwegman. 2020. "Who Benefits from Accountability-Driven School Closure? Evidence from New York City." *Journal of Policy Analysis and Management* 39(1):96–130.

Bloom, Matt, Amy E. Colbert, and Jordan D. Nielsen. 2021. "Stories of Calling: How Called Professionals Construct Narrative Identities." *Administrative Science Quarterly* 66(2):298–338.

Bovens, Mark, Robert E. Goodin, and Thomas Schillemans. 2014. *The Oxford Handbook of Public Accountability*. Oxford, UK: Oxford University Press.

Brayne, Sarah. 2017. "Big Data Surveillance: The Case of Policing." *American Sociological Review* 82(5):977–1008.

Bunderson, J. Stuart, and Jeffery A. Thompson. 2009. "The Call of the Wild: Zookeepers, Callings, and the Double-Edged Sword of Deeply Meaningful Work." *Administrative Science Quarterly* 54(1):32–57.

Cäker, Mikael, and Sven Siverbo. 2018. "Effects of Performance Measurement System Inconsistency on Managers' Role Clarity and Well-Being." *Scandinavian Journal of Management* 34(3):256–66.

Campbell, Donald T. 1979. "Assessing the Impact of Planned Social Change." *Evaluation and Program Planning* 2(1):67–90.

Carruthers, Bruce G., and Wendy Nelson Espeland. 1991. "Accounting for Rationality: Double-Entry Bookkeeping and the Rhetoric of Economic Rationality." *American Journal of Sociology* 97(1):31–69.

Chiang, Hanley. 2009. "How Accountability Pressure on Failing Schools Affects Student Achievement." *Journal of Public Economics* 93(9):1045–57.

Coleman, James, Thomas Hoffer, and Sally Kilgore. 1982. "Cognitive Outcomes in Public and Private Schools." *Sociology of Education* 55(2):65–76.

Colyvas, Jeannette A. 2012. "Performance Metrics as Formal Structures and through the Lens of Social Mechanisms: When Do They Work and How Do They Influence?" *American Journal of Education* 118(2):167–97.

Conley, Sharon, and Naftaly S. Glasman. 2008. "Fear, the School Organization, and Teacher Evaluation." *Educational Policy* 22(1):63–85.

Coviello, Lorenzo, Yunkyu Sohn, Adam D. I. Kramer, Cameron Marlow, Massimo Franceschetti, Nicholas A. Christakis, and James H. Fowler. 2014. "Detecting Emotional Contagion in Massive Social Networks." *PLoS ONE* 9(3):e90315.

Dee, Thomas S., and Brian Jacob. 2011. "The Impact of No Child Left Behind on Student Achievement." *Journal of Policy Analysis and Management* 30(3):418–46.

Deming, David J., Sarah Cohodes, Jennifer Jennings, and Christopher Jencks. 2016. "School Accountability, Postsecondary Attainment, and Earnings." *Review of Economics and Statistics* 98(5):848–62.

Destin, Mesmin, Paul Hanselman, Jenny Buontempo, Elizabeth Tipton, and David S. Yeager. 2019. "Do Student Mindsets Differ by Socioeconomic Status and Explain Disparities in Academic Achievement in the United States?" *AERA Open* 5(3):1–12.

Diamond, John B., and James P. Spillane. 2004. "High-Stakes Accountability in Urban Elementary Schools: Challenging or Reproducing Inequality?" *Teachers College Record* 106(6):1145–76.

Diaz Rios, Claudia Milena. 2020. "The Role of Policy Legacies in the Alternative Trajectories of Test-Based Accountability." *Comparative Education Review* 64(4):619–41.

Domina, Thurston, Andrew Penner, and Emily Penner. 2017. "Categorical Inequality: Schools As Sorting Machines." *Annual Review of Sociology* 43:311–30.

Douglas, Scott, and Albert Meijer. 2016. "Transparency and Public Value—Analyzing the Transparency Practices and Value Creation of Public Utilities." *International Journal of Public Administration* 39(12):940–51.

Espeland, Wendy Nelson, and Michael Sauder. 2016. *Engines of Anxiety: Academic Rankings, Reputation, and Accountability*. New York: Russell Sage.

Fard, Hassan Danaee, and Ali Asghar Anvary Rostamy. 2007. "Promoting Public Trust in Public Organizations: Explaining the Role of Public Accountability." *Public Organization Review* 7(4):331–44.

Figlio, David, and Lawrence S. Getzler. 2006. "Accountability, Ability and Disability: Gaming the System?" Pp. 35–49 in *Improving School Accountability*, edited by T. J. Gronberg and D. W. Jansen. Bingley, UK: Emerald Group.

Figlio, David, and Susanna Loeb. 2011. "School Accountability." Pp. 383–421 in *Handbook of the Economics of Education*, Vol. 3. New York: Elsevier.

Frank, Kenneth A., Spiro J. Maroulis, Minh Q. Duong, and Benjamin M. Kelcey. 2013. "What Would It Take to Change an Inference? Using Rubin's Causal Model to Interpret the Robustness of Causal Inferences." *Educational Evaluation and Policy Analysis* 35(4):437–60.

Frankenberg, Erica. 2009. "The Demographic Context of Urban Schools and Districts." *Equity & Excellence in Education* 42(3):255–71.

Golann, Joanne W. 2015. "The Paradox of Success at a No-Excuses School." *Sociology of Education* 88(2):103–19.

Gonsalves, Leroy. 2020. "From Face Time to Flex Time: The Role of Physical Space in Worker Temporal Flexibility." *Administrative Science Quarterly* 65(4):1058–91.

Goren, Paul. 2012. "Data, Data, and More Data—What's an Educator to Do?" *American Journal of Education* 118(2):233–37.

Gurova, Galina, and Nelli Piattoeva. 2018. "A Postsocialist Perspective on Audit Culture: Changing Practices and Subjectivities of School Teachers in a Russian Region." Pp. 173–87 in *Education Policies and the Restructuring of the Educational Profession: Global and Comparative Perspectives*, edited by R. Normand, M. Liu, L. Miguel C., Dalila A. Oliveira, and L. LeVasseur. Singapore: Springer.

Hallett, Tim. 2010. "The Myth Incarnate: Recoupling Processes, Turmoil, and Inhabited Institutions in an Urban Elementary School." *American Sociological Review* 75(1):52–74.

Hanushek, Eric A. 2019. "Testing, Accountability, and the American Economy." *Annals of the American Academy of Political and Social Science* 683(1):110–28.

Hazelkorn, Ellen. 2015. *Rankings and the Reshaping of Higher Education: The Battle for World-Class Excellence*. New York: Palgrave Macmillan.

Hibel, Jacob, and Daphne M. Penn. 2020. "Bad Apples or Bad Orchards? An Organizational Analysis of Educator Cheating on Standardized Accountability Tests." *Sociology of Education* 93(4):331–52.

Hong, Yihua, and Guanglei Hong. 2021. "Schools with Test-Based Promotion: Effects on Instructional Time Allocation and Student Learning in Grade 3." *AERA Open* 7:2332858420979167.

Hursh, David. 2007. "Assessing No Child Left Behind and the Rise of Neoliberal Education Policies." *American Educational Research Journal* 44(3):493–518.

Hwang, Hokyu, and Walter W. Powell. 2009. "The Rationalization of Charity: The Influences of Professionalism in the Nonprofit Sector." *Administrative Science Quarterly* 54(2):268–98.

Ingersoll, Richard M., and Gregory J. Collins. 2017. "Accountability and Control in American Schools." *Journal of Curriculum Studies* 49(1):75–95.

Jennings, Jennifer L. 2005. "Below the Bubble: 'Educational Triage' and the Texas Accountability System." *American Educational Research Journal* 42(2):231–68.

Kaivo-oja, Jari, Petri Virtanen, Harri Jalonen, and Jari Stenvall. 2015. "The Effects of the Internet of Things and Big Data to Organizations and Their Knowledge Management Practices." Pp. 495–513 in *Knowledge Management in Organizations*, edited by L. Uden, M. Heričko, and I.-H. Ting. Cham, Switzerland: Springer International.

Kirk, Andrea K., and David F. Brown. 2003. "Employee Assistance Programs: A Review of the Management of Stress and Wellbeing through Workplace Counselling and Consulting." *Australian Psychologist* 38(2):138–43.

Kraft, Matthew A. 2020. "Interpreting Effect Sizes of Education Interventions." *Educational Researcher* 49(4):241–53.

Lingard, Bob. 2021. "Enactments and Resistances to Globalizing Testing Regimes and Performance-Based Accountability in the USA." Pp. 279–93 in *World Yearbook of Education 2021: Accountability and Datafication in the Governance of Education*, edited by S. Grek, C. Maroy, and A. Verger. London: Routledge.

Marshall, Anthony, Stefan Mueck, and Rebecca Shockley. 2015. "How Leading Organizations Use Big Data and Analytics to Innovate." *Strategy & Leadership* 43(5):32–39.

Martens, Kerstin, and Dennis Niemann. 2013. "When Do Numbers Count? The Differential Impact of the PISA Rating and Ranking on Education Policy in Germany and the US." *German Politics* 22(3):314–32.

McAfee, Andrew, and Erik Brynjolfsson. 2012. "Big Data: The Management Revolution." *Harvard Business Review* 90(10):60–68.

Means, Barbara, Christine Padilla, and Larry Gallagher. 2010. "Use of Education Data at the Local Level: From Accountability to Instructional Improvement." Jessup, MD: U.S. Department of Education. Retrieved June 3, 2023. https://eric.ed.gov/?id=ED511656.

Mennicken, Andrea, and Wendy Nelson Espeland. 2019. "What's New with Numbers? Sociological Approaches to the Study of Quantification." *Annual Review of Sociology* 45:223–45. doi:10.1146/annurev-soc-073117-041343.

Mitchell, Marie S., Rebecca L. Greenbaum, Ryan M. Vogel, Mary B. Mawritz, and David J. Keating. 2018. "Can You Handle the Pressure? The Effect of Performance Pressure on Stress Appraisals, Self-Regulation, and Behavior." *Academy of Management Journal* 62(2):531–52.

Moen, Phyllis, Erin L. Kelly, Wen Fan, Shi-Rong Lee, David Almeida, Ellen Ernst Kossek, and Orfeu M. Buxton. 2016. "Does a Flexibility/Support Organizational Initiative Improve High-Tech Employees' Well-Being? Evidence from the Work, Family, and Health Network." *American Sociological Review* 81(1):134–64.

Neal, Derek A., and Diane Whitmore Schanzenbach. 2010. "Left Behind by Design: Proficiency Counts and Test-Based Accountability." *Review of Economics and Statistics* 92(2):263–83.

Nielsen, Jørn Flohr, and Erik Riiskjær. 2013. "From Patient Surveys to Organizational Change: Rational Change Processes and Institutional Forces." *Journal of Change Management* 13(2):179–205.

OECD (Organisation for Economic Co-operation and Development). 2019a. "Construction of Indices." In *PISA 2018 Results (Volume III): What School Life Means for Students' Lives."* Paris: Organisation for Economic Co-operation and Development. Retrieved June 3, 2023. https://www.oecd-ilibrary.org/sites/0a428b07-en/index.html?itemId=/content/component/0a428b07-en.

OECD (Organisation for Economic Co-operation and Development). 2019b. *PISA 2018 Results (Volume III)*. Paris:

OECD Publishing. Retrieved June 3, 2023. https://www.oecd-ilibrary.org/content/publication/acd78851-en.

OECD (Organisation for Economic Co-operation and Development). 2019c. *PISA 2018 Well-Being Framework*. Paris: Organisation for Economic Co-operation and Development. Retrieved June 3, 2023. https://www.oecd-ilibrary.org/education/pisa-2018-assessment-and-analytical-framework_38a34353-en.

OECD (Organisation for Economic Co-operation and Development). 2020. "A Framework for the Analysis of School Climate and Student Well-Being." Pp. 35–43 in *PISA 2018 Results (Volume III)*. Paris: Organisation for Economic Co-operation and Development. Retrieved June 3, 2023. https://www.oecd-ilibrary.org/education/pisa-2018-results-volume-iii_3b42acd8-en.

Oswald, Frederick L., Tara S. Behrend, Dan J. Putka, and Evan Sinar. 2020. "Big Data in Industrial-Organizational Psychology and Human Resource Management: Forward Progress for Organizational Research and Practice." *Annual Review of Organizational Psychology and Organizational Behavior* 7:505–33.

Poku, Michael. 2016. "Campbell's Law: Implications for Health Care." *Journal of Health Services Research & Policy* 21(2):137–39.

Ranganathan, Aruna, and Alan Benson. 2020. "A Numbers Game: Quantification of Work, Auto-gamification, and Worker Productivity." *American Sociological Review* 85(4):573–609.

Raudenbush, Stephen W., and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage.

Rouse, Cecilia Elena, Jane Hannaway, Dan Goldhaber, and David N. Figlio. 2013. "Feeling the Florida Heat? How Low-Performing Schools Respond to Voucher and Accountability Pressure." *American Economic Journal: Economic Policy* 5(2):251–81.

Royston, Patrick. 2004. "Multiple Imputation of Missing Values." *Stata Journal* 4(3):227–41.

Rutkowski, Leslie, and David Rutkowski. 2016. "A Call for a More Measured Approach to Reporting and Interpreting PISA Results." *Educational Research* 45(4):252–57.

Ryan, Shannon V., Nathaniel P. von der Embse, Laura L. Pendergast, Elina Saeki, Natasha Segool, and Shelby Schwing. 2017. "Leaving the Teaching Profession: The Role of Teacher Stress and Educational Accountability Policies on Turnover Intent." *Teaching and Teacher Education* 66:1–11.

Ryan, Jillian C., G. Williams, B. W. Wiggins, A. J. Flitton, J. T. McIntosh, M. J. Carmen, and D. N. Cox. 2021. "Exploring the Active Ingredients of Workplace Physical and Psychological Wellbeing Programs: A Systematic Review." *Translational Behavioral Medicine* 11(5):1127–41.

Sadowski, Jathan. 2019. "When Data Is Capital: Datafication, Accumulation, and Extraction." *Big Data & Society* 6(1):2053951718820549.

Safransky, Sara. 2020. "Geographies of Algorithmic Violence: Redlining the Smart City." *International Journal of Urban and Regional Research* 44(2):200–18.

Sauder, Michael, and Wendy Nelson Espeland. 2009. "The Discipline of Rankings: Tight Coupling and Organizational Change." *American Sociological Review* 74(1):63–82.

Schieman, Scott, Paul Glavin, and Melissa A. Milkie. 2009. "When Work Interferes with Life: Work-Nonwork Interference and the Influence of Work-Related Demands and Resources." *American Sociological Review* 74(6):966–88.

Schildkamp, Kim. 2019. "Data-Based Decision-Making for School Improvement: Research Insights and Gaps." *Educational Research* 61(3):257–73.

Schoen, LaTefy, and Lance D. Fusarelli. 2008. "Innovation, NCLB, and the Fear Factor: The Challenge of Leading 21st-Century Schools in an Era of Accountability." *Educational Policy* 22(1):181–203.

Sewell, Graham, and James R. Barker. 2006. "Coercion versus Care: Using Irony to Make Sense of Organizational Surveillance." *Academy of Management Review* 31(4):934–61.

Sewell, Graham, Joeri M. Mol, and Laurent Taskin. 2019. "The Future of Organizational Surveillance and the Very Future of Organizations." *Academy of Management Proceedings* 2019(1):13417.

Shuck, Brad, Thomas G. Relo, and Tonette S. Rocco. 2011. "Employee Engagement: An Examination of Antecedent and Outcome Variables." *Human Resource Development International* 14(4):427–45.

Smeding, Annique, Céline Darnon, Carine Souchal, Marie-Christine Toczek-Capelle, and Fabrizio Butera. 2013. "Reducing the Socio-Economic Status Achievement Gap at University by Promoting Mastery-Oriented Assessment." *PLoS ONE* 8(8):e71678.

Stevenson, Howard. 2017. "The 'Datafication' of Teaching: Can Teachers Speak Back to the Numbers?" *Peabody Journal of Education* 92(4):537–57.

Thoits, Peggy A. 2010. "Stress and Health: Major Findings and Policy Implications." *Journal of Health and Social Behavior* 51(1):S41–53.

Trinidad, Jose Eos. 2020. "Will It Matter Who I'm in School With? Differential Influence of Collective Expectations on Urban and Rural US Schools." *International Studies in Sociology of Education* 29(4):344–65.

Veenhoven, Ruut. 2008. "Sociology Theories of Subjective Well-Being." Pp. 44–61 in *The Science of Subjective Well-Being: A Tribute to Ed Diener*, edited by M. Eid and R. J. Larsen. New York: Guilford.

von der Embse, Nathaniel P., Laura L. Pendergast, Natasha Segool, Elina Saeki, and Shannon Ryan. 2016. "The Influence of Test-Based Accountability Policies on School Climate and Teacher Stress across Four States." *Teaching and Teacher Education* 59:492–502.

Waring, Stephen P. 2016. *Taylorism Transformed: Scientific Management Theory since 1945*. Chapel Hill: University of North Carolina Press.

Warr, Peter, and Karina Nielsen. 2018. "Wellbeing and Work Performance." In *Handbook of Well-Being*, edited by E. Diener, S. Oishi, and L. Tay. Salt Lake City, UT: DEF.

Wilson, Matthew W. 2011. "Data Matter(s): Legitimacy, Coding, and Qualifications-of-Life." *Environment and Planning D: Society and Space* 29(5):857–72.

Author Biography

Jose Eos Trinidad is an assistant professor of education policy at the University of California Berkeley and received his Joint PhD in Sociology and Comparative Human Development from the University of Chicago. Bringing together the sociologies of education and organizations, his research interrogates education policies, "outside" organizations, and institutional change.