

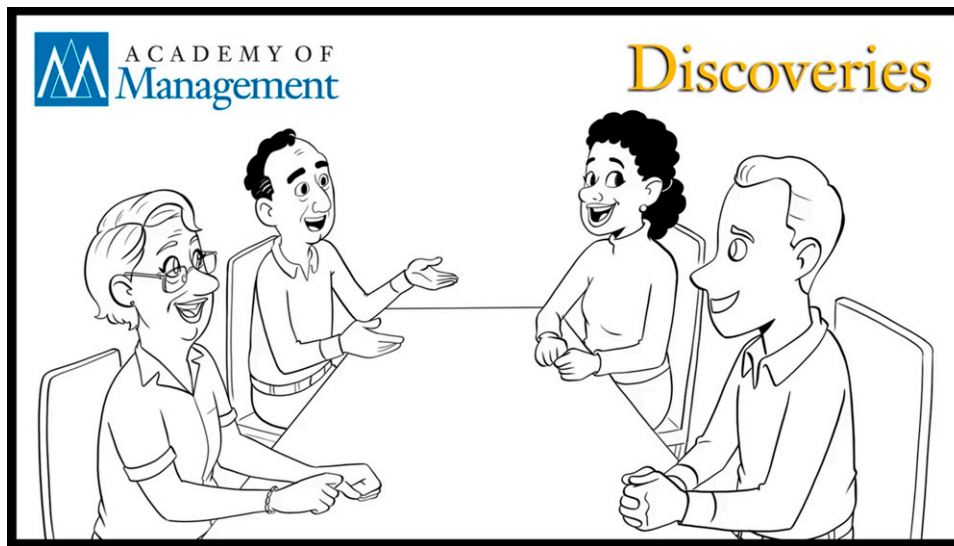
WHEN IS PSYCHOLOGICAL SAFETY HELPFUL IN ORGANIZATIONS? A LONGITUDINAL STUDY

MONICA C. HIGGINS
Harvard University

SHOSHANA R. DOBROW
London School of Economics and Political Science

JENNIE MILES WEINER
University of Connecticut

HAIYANG LIU
London School of Economics and Political Science



Research has documented many benefits associated with team-level psychological safety. However, we know little about the boundary conditions of psychological safety, such as how it operates at the organization level and if and when it is helpful over time. In this research, we explore how organization-level psychological safety, in conjunction with another aspect of workplace climate, felt accountability, impacts organizational performance over time. Our study context is the New York City public school system, a context rife with uncertainty and calls for change, including pressure on teachers to produce and improve student outcomes. Utilizing over 170,000 survey responses from teachers in 545 schools across three years, our multilevel analyses unexpectedly show that psychological safety is not on its own, nor necessarily, “helpful” with regard to organizational performance over time. Indeed, the best conditions for fostering organizational performance occur when psychological safety is relatively low and felt accountability is relatively high. Thus, these two dimensions of workplace climate appear to be interrelated in critical ways over time, albeit unexpectedly. We conclude with implications of our discoveries for future research, and propose new lines of research on the roles of interdependence, attention, and time for studying psychological safety at the organization level.

We would like to thank our editor, Sandra Robinson, for her valuable guidance and insights throughout the review process, and our two anonymous reviewers for their constructive comments. We are also grateful for the assistance and critical feedback we received from Ebony Bridwell-Mitchell, Stacey Childress, Rebecca Holcombe, Ann Ishimaru, Shauna Leung, Sola Takahashi, and Harvard’s Groups Group. *Editor’s note:* The manuscript for this article was accepted for publication during the term of AMD’s previous editor-in-chief, Peter Bamberger.

In the United States, approximately 85% of students graduate from high school; however, extraordinary variance exists such that, in some states, that number plummets to 69% (National Center for Education Statistics, 2020). Gaps in performance along racial, ethnic, and income lines are pernicious and persistent (Orfield & Frankenberg, 2014). For example, in the last 25 years, the gap in reading scores between children

coming from low- versus high-income families in the United States has increased by almost 40% (Reardon, 2011; see also Reardon, Kalogrides, & Shores, 2019). These gaps and declining performance significantly affect the capacity of children and families to move out of poverty. A critical lever for addressing these serious problems is improving the performance of schools, for the benefit of not only individual students and their families but also society overall.

Although management research on factors such as workplace climate and organizational culture could speak to such variation in school performance, studies of schools and the context of education have remained largely absent from discourse in management journals. Indeed, in *Academy of Management* journals since 2000, only three studies have focused on primary or secondary education contexts¹ in spite of the fact that some of the most path-breaking work on organizations and management has come from research done in unconventional settings (Bamberger & Pratt, 2010) such as schools (e.g., Weick, 1976). In particular, research has yet to explore how well-established constructs, such as psychological safety (Edmondson, 1999)—that is, a climate in which workers do not fear speaking up, asking for help, or admitting mistakes—might operate to improve performance in contexts like public education. This unexplored area is an opportunity for management scholars, and one that we take up in the present research.

Many settings exist in which entities such as the federal government have intervened to try to improve organization-level outcomes by placing increasing pressure on those most central to the work at hand, such as teachers in the context of public education, or, similarly, workers in the car industry after the Great Recession or health care workers with the onset of the Affordable Care Act. More recently, workers in the U.S. manufacturing industry have faced increasing pressure to improve firm performance, after the federal government's imposition of additional tariffs on foreign trade. When external accountability systems are imposed on organizations in the face of tremendous uncertainty and pressure for change, workers may respond in a variety of ways, differentially affecting performance outcomes (Patil, Vieider, & Tetlock, 2014). This variation in performance, we believe, can be explained at least in part by the climate of the workplace. Here, by surveying thousands of teachers in

¹ We examined *Academy of Management Annals*, *Academy of Management Discoveries*, *Academy of Management Learning & Education*, *Academy of Management Journal*, *Academy of Management Perspectives*, and *Academy of Management Review*, 2000–2020, using search terms including “public school,” “school reform,” “elementary school,” “high school,” and “K-12.”

Author's Voice:
What motivated you to undertake
this research?



hundreds of schools operating within the same external accountability system and high-stakes context of the New York City (NYC) public school system, we explore whether and how workplace climate—specifically, teachers' perceptions of their schools' psychological safety—impacts organizational performance.

We focus in particular on psychological safety as a key aspect of workplace climate because, for decades, it has been touted as a critical means of mitigating anxiety during times of uncertainty and change (Schein & Bennis, 1965). Building on the early work of Schein and Bennis (1965), Edmondson (1999) concretized the concept of psychological safety, including establishing important linkages with learning behaviors and outcomes in the work of teams (for a review, see Edmondson & Lei, 2014). Consistent with Schein and Bennis (1965), Edmondson and colleagues demonstrated the value of psychological safety as a lever for team performance improvement when the stakes are high and the work a team must do is complex and ambiguous, such as in hospital emergency rooms and intensive care units (e.g., Edmondson, 2003; Tucker, Nembhard, & Edmondson, 2007). In these types of contexts, it is especially important that workers feel psychologically safe. Psychological safety can yield benefits for end users, whether they be hospital patients (Edmondson, Bohmer, & Pisano, 2001) or frontline industrial workers (Morrow, McGonagle, Dove-Steinkamp, Walker, Marmet, & Barnes-Farrell, 2010). Given these studies and the parallels often drawn between contexts such as health care and education (e.g., Edmondson, Higgins, Singer, & Weiner, 2016), in the present study we explore psychological safety in the context of schools, where teachers are the focal workers.

Beyond application of the important concept of psychological safety in the novel context of public education, we embarked on this research to explore, more generally, the boundary conditions of psychological safety, as called for by recent research (Frazier, Fainshmidt, Klinger, Pezeshkan, & Vracheva, 2017). Despite the many studies reporting direct and positive relationships between team-level psychological safety and team-level learning and performance (e.g., Tucker et al., 2007), exceptions exist. Some studies have found that psychological safety has an indirect, not a direct, effect on team-level outcomes (e.g., Bunderson & Boumgarden, 2010), serves as a mediator between team structure and team

learning (e.g., Bresman & Zellmer-Bruhn, 2013), or even has negative effects in some circumstances (e.g., Deng, Leung, Lam, & Huang, 2019). Other studies have suggested the need for research on the role that psychological safety plays at the organization level (Edmondson & Lei, 2014). Together, these findings suggest the possibility of boundary conditions to psychological safety. Consistent with the idea that boundary conditions are critical factors for understanding the generalizability of a theory across contexts (Busse, Kach, & Wagner, 2016), we consider several aspects of the context of psychology safety at work. In the present research, we consider both other variables and other levels of analysis to address our overarching research question: *When is psychological safety helpful to organizational performance?*

We explore the boundary conditions of psychological safety in several ways. First, whereas the bulk of research on psychological safety has occurred at the team level, we shift the level of analysis to that of the organization, thereby addressing a call for organization-level psychological safety research (Edmondson & Lei, 2014). Unlike emergency rooms and intensive care units, where much of the empirical work on psychological safety emerged, some work contexts do not revolve as tightly around teamwork. Indeed, in some contexts, such as U.S. public education, the focal workers—teachers—do not perform the bulk of their daily work in teams; rather, teachers generally work behind classroom doors and carry out their work independently rather than interdependently (Vangrieken, Grosemans, Dochy, & Kyndt, 2017). Recent work on psychological safety has suggested, but not yet tested, the idea that work that is not done in a team setting and is not marked by interdependence might not always lead to the same positive outcomes previously found (Deng et al., 2019). Similarly, here we propose that when work occurs in contexts marked by little interdependence, analysis concerning workplace climate may best be considered at the organization (here, school) and not at the team level. Thus, we explore new territory for psychological safety research by shifting from the more common focus on team-level psychological safety to *organization*-level psychological safety.

In making this shift, we draw on research on the adjacent concept of psychological empowerment, which has expressed a similar logic. There, scholars have noted that although workplace climates characterized by psychological empowerment reflect perceptions of collectives that emerge from interactions over time (Morgeson & Hofmann, 1999), these collectives need not necessarily consist of teams (Wallace, Johnson, Mathe, & Paul, 2011). As Wallace and colleagues (2011: 2) explained, drawing from the prior

research of Hackman (1992), “climates are collective constructs... in which ambient stimuli serve as a source of information that signal the appropriateness of member behaviors.” In the present study, we make a similar suggestion: that psychological safety is a collective construct that emerges from interactions over time (see also Morgeson & Hofmann, 1999). Thus, our conceptual treatment of the construct of psychological safety is as a “shared unit-level property” (Kozlowski & Klein, 2000), meaning it reflects a characteristic that is common to, or shared by, the members of a unit, which, in this instance, is an organization as opposed to a team. Our methodology matches this conceptual treatment by asking respondents how psychologically safe workers feel “in this (organization) school” as opposed to “in this team,” as was the case in the foundational work on psychological safety (Edmondson, 1999).

Second, we explore the boundary conditions of organization-level psychological safety by investigating another dimension of workplace climate—felt accountability—that has been suggested as operating alongside psychological safety (e.g., Edmondson, 2008) but has not yet been empirically studied in this regard. Felt accountability refers to “an implicit or explicit expectation that one’s decisions or actions will be subject to evaluation by some salient audience(s) (including oneself)” (Hall, Royle, Brymer, Perrewé, Ferris, & Hochwarter, 2006: 88). Although considerable research on accountability has focused on external accountability systems and how those systems are structured (e.g., Castilla, 2015; Tetlock & Mellers, 2011), we instead take a phenomenological view (Frink & Klimoski, 1998; Hall, Frink, Ferris, Hochwarter, Kacmar, & Bowe, 2003) such that we conceive of accountability as internal or *felt* (cf. Elmore, 2004; for a review, see Frink et al., 2008). Consistent with prior management research, we view felt accountability as a “state of mind,” rather than a “state of affairs” (Frink & Klimoski, 2004: 3)—that is, as a subjective, rather than objective, construct reflecting peoples’ experiences of their work environments.

Recent research has suggested that felt accountability may help to explain the conditions under which psychological safety has a positive versus negative impact on various outcomes (Deng et al., 2019). In other words, psychological safety may be more complicated than originally conceived, in that its effects are not always positive. As Deng and colleagues (2019) recently proposed, psychological safety may operate via two different pathways, yielding positive or negative consequences depending on the underlying mechanisms at play (Deng et al., 2019). On the one hand, psychological safety may reduce group

members' fear of failure that can naturally arise when they are faced with learning tasks, and thus positively affect a group's risk-taking behaviors. On the other hand, psychological safety may reduce worker motivation such that group members exert less effort, yielding a negative effect on a group's risk-taking behaviors. Although Deng and colleagues (2019: 1139) mentioned that the arguments supporting the latter—negative influence of psychological safety—are based on “the accountability perspective,” their study did not directly examine felt accountability. Indeed, the authors suggested this area as an avenue for future research, which we explore here.

Building on Deng et al.'s (2019) study, then, we investigate felt accountability as a possible boundary condition associated with the “when” in our core research question, “When is psychological safety helpful?” This focus echoes prior work by Edmondson (2008), who suggested that psychological safety may or may not produce positive outcomes when paired with “accountability”; this work also recommended studying accountability alongside psychological safety in future research. Here, we investigate the role of felt accountability more fully, both empirically and conceptually.

Importantly, as with our organization-level focus on psychological safety, we also examine felt accountability at the organization level. Specifically, we explore workers' perceptions of how accountable people in the organization feel about *organization-level*, as opposed to individual-level, goals. Our exploration is not about individuals' views of their own goal achievement; rather, it centers on individuals' perceptions of their workplace climate at the organization level. Therefore, we define organization-level accountability as individuals' perceptions about how accountable people in the organization feel about organization-level goals. Methodologically, we thus consider felt accountability in a manner parallel to our treatment of organization-level psychological safety; we focus on participants' reports about the felt accountability “in this school,” thereby again reflecting the idea that this construct is a shared unit-level property (Kozlowski & Klein, 2000). Whereas prior felt accountability research has used individual-level measures to assess individuals' self-assessments of how accountable they feel regarding their personal goals (e.g., Mackey, Brees, McAllister, Zorn, Martinko, & Harvey, 2018), we instead assess felt accountability as a feature of workplace climate by capturing individuals' perceptions of how accountable they believe people in their organizations feel with regard to organization-level targets, standards, and goals.

In so doing, the present research addresses longstanding calls in the felt accountability literature to

extend the construct beyond the micro or individual level, which has received the “overwhelming majority of research attention” (for a review, see Hall, Frink, & Buckley, 2017: 205), to instead treat it as operating at multiple levels (e.g., Frink et al., 2008). Because regulations and professional norms can influence organizational culture and climate, scholars have suggested adopting a multilevel perspective to account for the nested structure in which individuals work (e.g., Hall et al., 2017). Contexts in which organizations are embedded in a larger external accountability system due to constraints imposed on them by political entities are commonplace in many sectors, particularly those that are highly regulated, such as health care (in which doctors, for example, work in hospitals, nested in regulated health care networks and systems [Edmondson et al., 2016]). Our study context provides a unique opportunity to focus on this multilevel perspective. By virtue of operating in the same school district in the same state with the same set of standards, all organizations in our sample were subject to the same external accountability system. By effectively controlling for this macro level of influence, we were able to focus directly on perceptions of felt accountability and psychological safety across schools.

Third, we explore the role of time as a boundary condition of organization-level psychological safety. Time is a critical yet underexplored factor in management research (e.g., George & Jones, 2000; Mitchell & James, 2001; Wright, 1997; Zaheer, Albert, & Zaheer, 1999). Scholars have regularly called for research that leverages sophisticated analytical techniques to depict complex relationships among variables over time (e.g., Song, Liu, Shi, & Wang, 2017). Thus, the fact that scholars have emphasized the need for longitudinal research on psychological safety as well is unsurprising (Frazier et al., 2017). This lack of longitudinal research limits scholars' ability to make robust claims about relationships between constructs, particularly if these constructs vary over time, as is the case for workplace climate indicators such as psychological safety. Indeed, incorporating the role of time through longitudinal research on psychological safety is needed to gain greater purchase on the causality suggested in prior studies to explore “when” psychological safety is helpful (Edmondson & Lei, 2014).

In the present research, we explore these proposed boundary conditions of organization-level psychological safety and, thus, our overarching research question—When is psychological safety helpful? Further, we do so by considering four additional questions: First, *How does psychological safety impact organizational performance over time?* Second, regarding felt accountability, *How does felt accountability impact*

organizational performance over time? Third, we explore these two dimensions of workplace climate together: *How do psychological safety and felt accountability together impact organizational performance over time?* And finally, to truly discover the power of psychological safety over time, we ask, *What is the best combination of psychological safety and felt accountability for organizational performance over time?* Is it best to have high psychological safety and high felt accountability, as Edmondson (2008) suggested, or does some other combination work “best” for organizational performance over time? Our study draws on a three-wave, three-year study of 545 schools in the NYC public school system, which allowed for a longitudinal, multilevel exploration of when psychological safety is helpful with respect to organizational performance over time, thereby offering theoretical and empirical discoveries for future research and practice.

METHOD

Sample and Procedure

We conducted our analyses on three waves of longitudinal data from the NYC public school system (for the academic years 2008–2009, 2009–2010, and 2010–2011). This school system is the largest in the United States and, indeed, is large even by corporate standards: In 2018–2019, with a total budget of \$32.3 billion, it served 1.1 million students in over 1,800 schools and employed about 75,000 teachers (New York City Department of Education, 2019).

For several reasons, this context of the NYC public school system is ideally suited for addressing our questions about when and how organization-level psychological safety and felt accountability impact organizational performance over time. First, for years, including the time period of this study, this large urban public school system has been characterized by tremendous uncertainty, complexity, and urgent calls for change, and so our core constructs should both vary and matter in this context. Second, this school system has been plagued by daunting challenges, including low performance and low graduation rates that persist. For instance, in 2015, over half of the schools that the New York State Department of Education identified as “failing” were located in the NYC school system (91 out of 178).² Traditional management responses, such as

increasing resources, have proven insufficient: New York, like other states, has increased its education budget continuously over time,³ yet student opportunity and performance gaps often widen. These issues are the kinds of pernicious problems—those characterized by high levels of uncertainty and complexity and faced by knowledge workers—that prior research has suggested psychological safety can effectively address (Edmondson & Lei, 2014).

In the context of U.S. public education, the urgent need for improved performance faces the additional obstacle of having professional norms that make this type of change very difficult. Scholars have long observed that the “egg crate structure” of schools creates a work culture in which teachers value autonomy (Lortie, 1975), independence (Little, 1990; York-Barr & Duke, 2004), and egalitarianism (Donaldson, Johnson, Kirkpatrick, Marinell, Steele, & Szczesiul, 2008; Lortie, 1975), which undermines opportunities for sharing and learning that can improve performance. At a systems level, these strong values, combined with the lack of direct oversight of teachers’ work that characterizes public schools, can create “loosely coupled systems” in which rules and regulations are often followed in name only (Meyer & Rowan, 1977). Consequently, tremendous variability in teaching practices persists across and within schools (Sass, Hannaway, Xu, Figlio, & Feng, 2012), with student performance tending to cluster among certain neighborhoods and demographics (Balfanz & Legters, 2005).

In 2001, to tighten oversight in this loosely coupled system, the U.S. government reauthorized the Elementary and Secondary Education Act as No Child Left Behind (NCLB). NCLB was an accountability system that focused on the *outcomes* that states and, by extension, schools and teachers, produce; it did not focus on *how* teachers do their work (Linn, Baker, & Betebenner, 2002). Many believe this legislation was the primary catalyst for the current and far-reaching school accountability movement in America’s schools (Hess & Petrilli, 2006; Riddle & Kober, 2011; Spillane, Camburn, Pustejovsky, Pareja, & Lewis, 2008). This impact is likely due to three aspects of NCLB. First, the law required that states develop measures to assess, and then publicly report on, student proficiency in reading or language arts and mathematics. Second, states were tasked with setting yearly performance targets to show students were making “Adequate Yearly Progress” (AYP) toward all students being deemed “proficient” by 2014. Third, when schools did not achieve their AYP targets, they were labeled as underperforming. A subset of these schools, the approximately 55% that

² Criteria for this identification included scoring in the bottom 5% in the state for English and math performance, and graduation rates below 60% (New York State Office of the Governor, 2015).

³ At a national level, U.S. spending per pupil rose from an average of \$5,773 in 1980 to \$11,762 during the 2016 fiscal year (U.S. Census Bureau, 2018).

received federal monies, were subjected to sanctions for this poor performance (U.S. Department of Education, 2006). These sanctions grew increasingly harsh when performance stalled, and could include parents obtaining the option to transfer their children out of these schools, schools losing their federal funding, and teachers and administrators being replaced, often *en masse* (Paige, 2006). Thus, not achieving AYP targets resulted in meaningful negative consequences for schools, and thus in acute pressure placed upon teachers at the time of the present study (Barrett, 2009).

Starting in 2007, the New York City Department of Education (NYCDOE) invited all students, parents, and teachers in the school system to complete its “New York City School Surveys,” with the goal of better understanding the learning environment. The magnitude of this data collection was second only to the U.S. Census (Nathanson, Cole, Kemple, Lent, McCormick, & Segeritz, 2013). These surveys focused on respondents’ perceptions of academic expectations, communication, engagement, and physical safety and respect. Given our research focus, we use data from a subset of participants for whom schools were their workplace: the teachers. The NYCDOE provided us with the survey results for teachers ($n = 52,065$, $58,940$, and $59,651$ teachers at Times 1 to 3, respectively), along with school-level demographic information about students, teachers, and principals ($n = 1,445$, $1,314$, and $1,356$ schools at Times 1 to 3, respectively).⁴ We obtained school-level performance data from the New York State Education Department website.⁵

Starting with its 2008–2009 survey, the NYCDOE agreed to include our measure of psychological safety in its teacher surveys;⁶ the survey already contained measures of felt accountability (see details of both measures below). This survey, which was Time 1 of our data collection, was conducted both electronically and on paper. However, only the electronic version contained the psychological safety questions, meaning roughly half the teachers in the district completed the electronic version ($n = 25,329$), whereas the other

⁴ Response rates for teachers in the school system were strong overall, typically in the 60–80% range across data collections.

⁵ Provided by the New York State Education Department, the New York State Report Card website provides access to a variety of school-level data to the public (<https://data.nysed.gov/>). New York State stopped using the specific performance outcome measure applied in our study (AYP, see description below) with the Time 3 survey; therefore, this survey marked the final wave of data in our study.

⁶ The addition of these items to the survey required approval from the teachers’ union, a process that took six months.

Author’s Voice:

What was the most difficult or challenging aspect of this research project and paper?



half completed the paper survey without these items ($n = 26,736$).⁷ The next annual survey (2009–2010 school year), Time 2 of our data collection, was conducted electronically only, and so included all of our focal measures, including psychological safety questions. The subsequent annual survey (2010–2011 school year), Time 3 of our data collection, was again conducted electronically and included psychological safety questions in all surveys. Table 1 displays all variables included in the analyses and the time periods during which they were measured.

Because the organization (i.e., school) is our unit of analysis, we included in our analyses only those schools that provided data at all three time points, yielding a final sample of 545 schools. To investigate selection bias in our sample, we compared school-level demographic characteristics, about students, teachers, and the schools themselves, of our final sample to those of the whole NYC school system.⁸ We found no noticeable patterns of difference, indicating that our final sample of schools was representative of the broader school system.

Measures

Organizational performance. We used AYP to assess the organizational performance of schools. AYP is a binary indicator of whether a school met a state-defined performance threshold for student achievement. During our longitudinal time frame, federal law required states to measure AYP primarily by students’ performance on state tests and to set annual benchmark targets for schools (NCLB, 2001). Additionally, AYP targets incorporated criteria regarding how students from different demographics performed.⁹ By

⁷ We conducted a variety of *t*-tests to assess whether differences existed between teachers using electronic versus written surveys, and found no evidence of differences for any of these characteristics.

⁸ We obtained data about the entire NYC school system from the NY State Department of Education’s online data archives, <https://data.nysed.gov/lists.php?type=district>

⁹ For instance, a school could have a 90% average proficiency level, but because it also had an AYP threshold of 90% proficiency required across multiple demographic subgroups (e.g., African Americans, Special Education students, English Language Learners, etc.), it would fail if less than 90% of any of those particular subgroups reached proficiency.

design, state targets were set higher each year because AYP required continual movement toward a goal of 100% student proficiency in 2014. Thus, the likelihood of achieving the AYP target became increasingly difficult as the years in our study progressed and teachers experienced heightened pressure to perform in this context, where performance was indexed by their attainment of school-level targets for performance.

New York schools tested students' proficiency in reading or language arts, mathematics, and science, as well as reporting high school graduation rates. In our analyses, we focused on NYC schools' attainment of the AYP target for reading or language arts only ("English Language Arts" or ELA) for two reasons: First, this measure applied to *all* students (from kindergarten through grade 12) in the school system, unlike science, which was assessed only for elementary and middle school students, and high school graduation rates, which applied only to high school students. Second, policy-makers and practitioners at the district and state levels expressed concerns about the substance of the AYP target for mathematics (e.g., being deemed as lacking rigor, resulting in changing standards from year to year) as well as changes made to the instrument and the scoring methods (e.g., renorming the scores differently every year [Di Carlo, 2012; Medina, 2010; Winerip, 2011]). These concerns regarding both the internal and external validity of the mathematics test results would make any form of longitudinal analysis on this measure meaningless.

We note the AYP accountability system applied to all teachers in each school, regardless of which subject they taught. That is, although the ELA AYP assessed only the subject area of reading or language arts, all teachers—not just English teachers—were held accountable for students' ELA results. Using data from the New York State Report Card website, we created a variable indicating whether each school met its AYP target in ELA at Times 1 to 3 (1 = *school met AYP target this year*, 0 = *school did not meet AYP target this year*).

Like most school performance measures, the use of AYP as a measure of school performance has sparked some controversy over the years, including the appropriateness of using it to sanction schools (Balfanz & Legters, 2005). However, despite these critiques, during our study time frame, education leaders throughout the United States used AYP to identify low-performing schools (Hochbein, Mitchell, & Pollio, 2013), and education researchers used it to quantify performance and growth (e.g., Shapiro, Solari, & Petscher, 2008). Further, it was the only performance measure that was both (a) comparable across different school types (i.e., elementary, middle, and high schools) and (b) usable for longitudinal analyses (i.e., not normalized from year to year, which would have prohibited longitudinal analyses). Perhaps most importantly, at the time of our data collection, AYP was the preeminent measure of school performance—indeed, the primary outcome measure to which schools were held

TABLE 1
Overview of Study Variables

	Source			
	Pre-Time 1	Time 1	Time 2	Time 3
Variables				
<i>Control Variables</i>				
Physical safety (Time-varying)		X	X	X
Prior performance (Pre-Time 1)	X			
Racial demographics (Average over time) ^a		X	X	X
Socioeconomic demographics (Average over time) ^a		X	X	X
Limited English proficiency (Time-varying)		X	X	X
Documented disabilities (Time-varying)		X	X	X
School type (Average over time) ^a		X	X	X
Size (Average over time) ^a		X	X	X
Teacher experience (Time-varying)		X	X	X
<i>Focal Predictor Variables</i>				
Psychological safety (Time-varying)		X	X	X
Felt accountability (Time-varying)		X	X	X
<i>Outcome Variable</i>				
AYP (Time-varying)		X	X	X
Timing				
Year	2007–2008	2008–2009	2009–2010	2010–2011

^a These variables were measured at multiple time points, but because they were relatively invariant over time, we used a single value—the mean of the Times 1, 2, and 3 values—in the analyses.

accountable—and, thus, strongly influenced the beliefs and behaviors of those working in schools (Smith & Kovacs, 2011).

Organization-level psychological safety. We assessed teachers' perceptions of psychological safety in their organization with a three-item scale included on our behalf by the NYCDOE in the Times 1, 2, and 3 teacher surveys. These items, based on Edmondson's (2003) conceptualization of psychological safety, were previously validated in another large urban school district (Higgins, Ishimaru, Holcombe, & Fowler, 2012). All items referred to perceptions of the school, using "this school" as the referent, as opposed to "team," the referent typically used in prior psychological safety research (e.g., Schulte, Cohen, & Klein, 2012). Teachers indicated their degree of agreement (1 = *strongly disagree* to 4 = *strongly agree*, reverse-coded from the original) with the following three items: "In this school, it's easy to speak up about what is on your mind"; "People in this school are eager to share information about what does and doesn't work"; and "People in this school are usually comfortable talking about problems and disagreements."

We created unstandardized composites for each teacher ($\alpha = .89, .89, \text{ and } .89$ at Times 1, 2, and 3, respectively) and then aggregated them to create a mean composite score for each school i for each year j . To assess the appropriateness of aggregating the psychological safety construct to the organization level, we examined r_{wg} , the intraclass correlation (ICC1), and the reliability of the mean (ICC2) (Bliese, 2000; Glick, 1985; LeBreton & Senter, 2008). Results supported aggregation, with all indices in the recommended range (r_{wg} (Median) = .80, .95, .77; ICC(1) = .20, .20, .20; and ICC(2) = .81, .89, .90 for Times 1 to 3, respectively).

Organization-level felt accountability. We assessed the extent to which teachers felt their organizations were accountable using a 6-item scale included in the Times 1, 2, and 3 teacher surveys. These items reflect the idea that those working in schools can expect the possibility of evaluation by a "salient audience" (Hall et al., 2017: 206) along a set of "prescribed standards" (Schlenker, Britt, Pennington, Murphy, Doherty, & Kintsch, 1994: 634). Thus, consistent with the conceptual underpinnings of the felt accountability construct (Frink et al., 2008), we assessed the extent to which teachers felt their school was accountable for helping students ultimately reach the goals set by NCLB: having expectations and setting standards and targets relating to the primary outcome of interest for schools striving to improve performance (Smith & Kovacs, 2011). As with psychological safety, we measured felt accountability as an organization-level construct in which the

organization (i.e., "my school" or "this school") was the referent.

Teachers indicated their degree of agreement (1 = *strongly disagree* to 4 = *strongly agree*, reverse-coded from the original) with the following six items: "My school has high expectations for all students," "Our school is focused on improving performance on measures of student achievement for this year," "My school has clear measures of progress for student achievement throughout the year," "Meeting targets for student progress is a priority in this school," "Helping students reach mastery for important skills and content is a priority for this school," and "This school makes it a priority to help students develop challenging learning goals." We created unstandardized composites for each teacher and then aggregated them to create a mean composite score for each school i for each year j ($\alpha = .91, .92, \text{ and } .93$ at Times 1, 2, and 3, respectively). As with psychological safety, results of r_{wg} , ICC(1) and ICC(2) provided support for aggregating the felt accountability construct from the individual teacher level to the organization level, with all indices in the recommended range (r_{wg} (Median) = .95, .95, .95; ICC1 = .22, .20, .21; ICC2 = .82, .89, .90 for Times 1 to 3, respectively).

Control variables. We controlled for several variables in our analyses, all provided to us by the NYCDOE, as follows: First, a lack of *physical safety* (e.g., gang activity, drugs, etc.) in a school has been linked to increased teacher burnout, decreased self-efficacy, and negative attitudes toward students, all of which reduce student achievement (Dinkes, Cataldi, Linkelly, & Snyder, 2009). Thus, we controlled for teachers' perceptions of physical safety in their schools. In the Times 1 through 3 surveys, teachers indicated their degree of agreement (1 = *strongly disagree* to 4 = *strongly agree*, all reverse-coded from the original) with the following four items: "Order and discipline are maintained at my school," "I can get the help I need at my school to address student behavior and discipline problems," "I am safe at my school," and "Crime and violence are a problem at my school." Two additional items were included in the Time 1 survey (both reverse-coded), which we included in the scale: "Students' use of alcohol and illegal drugs is a problem at my school" and "Gang activity is a problem in my school." We aggregated these teacher-level composites to the organization level ($\alpha = .83, .97, \text{ and } .97$ at Times 1, 2, and 3, respectively), with higher scores indicating greater feelings of safety, i , for each year, j .

Additionally, we controlled for three factors identified by prior research as influencing the likelihood of achieving AYP targets (Taylor, Stecher, O'Day, Naftel, & Le Floch, 2010): First, we controlled for

prior performance using the school's ELA AYP status for the school year preceding the start of our study (i.e., 2007–2008; 1 = *school met AYP target*, 0 = *school did not meet AYP target*), obtained from the New York State Report Card website. Of the 545 schools in our sample, 86% met their ELA AYP target in 2007–2008.¹⁰ This relatively high pass rate was consistent with national trends at that time (Usher, 2012) because fewer students were required to reach proficiency than relative to subsequent years. Indeed, by 2010–2011 (Time 3), the number of schools meeting their target dropped to 69%.

Second, we controlled for several aspects of demographic diversity, which is especially consequential for AYP because, as mentioned above, the law mandated the disaggregation of results and the attainment of specific targets for subgroups of students, known as “special populations.” Students often belonged to multiple subgroups, and their performance affected multiple AYP targets (Morrison, Morrison, & Bedford, 2007), such that schools with more subgroups tended to perform less well (Taylor et al., 2010). We controlled for the *racial demographics* of each school's student body using the percentage of White students, and we used the percentage of students receiving federally subsidized meals (i.e., free or reduced price) as an indicator of *socioeconomic demographics*. We assessed *limited English proficiency* as the percentage of students requiring English-language learning support, and *documented disabilities* as the percentage of students eligible for special education.

Third, we controlled for several aspects of the schools themselves. Because middle and high schools tended to be less likely to achieve their AYP targets compared to elementary schools (Taylor et al., 2010), we controlled for *school type* (dummy variables for *middle school*, *high school*, and *other school types*). We used enrollment figures to control for *school size*, because smaller schools are expected to be more likely to achieve AYP targets compared to larger ones (Taylor et al., 2010). In addition, because many studies have demonstrated links between teacher experience and effectiveness (e.g., Leana & Pil, 2006), with teachers with three or fewer years of experience being less effective, we controlled for *teacher experience* (in years) measured at Times 1 to 3, which we aggregated to the school level for each time period.

As displayed in Table 1, three of these control variables varied over time (limited English proficiency, documented disabilities, teacher experience), such

that we treated them as time-varying measures in our analyses. The other four control variables remained relatively constant over time (racial demographics, socioeconomic demographics, school type, and school size), and, as such, we treated them as time-invariant measures in our analyses by using the mean of each of these four variables across Times 1 through 3.

Analytic Strategy

We conducted our analyses using multilevel growth modeling techniques. This approach recognizes that our observations over time were clustered within organizations, and so accounts for organization-level attributes that may impact our outcome—namely, whether school *i* achieved its ELA AYP target for year *j* (Raudenbush & Bryk, 2002). In this two-level model, the lower level (“Level 1”) is the wave of data collection, allowing exploration of what occurs *within organizations (across years)*, and the upper level (“Level 2”) is the organization, allowing exploration of what occurs *between organizations (across years)* (Singer & Willett, 2003).¹¹

More specifically, given the binary nature of our outcome (i.e., whether schools achieved their AYP target in a given year), we fit a set of multilevel *logistic* growth models to our data (Wong & Mason, 1985)¹² using Stata's mixed-effects logistic regression command, *xtmelogit* (see Appendix A for definitions and formulae of the models). As in multiple regression models, the parameter estimates and corresponding *p*-values of the predictor variables reflect the direction, size, and significance of their relationships to the outcome variable. Here, because our models are logistic, we interpret the predictors as relating to the schools' odds of achieving their AYP targets. The time variable, “Year,” captures the passage of time over the course of the study. The intercept represents the initial likelihood of schools achieving their AYP target—that is, at Time 1, when Year is 0. The value of Year at Time 2 is 1 (i.e., one year post Time 1), and Year at Time 3 is 2 (i.e., two years

¹¹ At Level 1, we examined the relationship between time (i.e., years) and organizational performance, AYP. This generated the Level 1 parameters, an intercept and a slope, which determine the shape of each organization's “true trajectory of change” (Lenzenweger, Johnson, & Willett, 2004) because the intercept parameter represents an organization's true value of AYP at the beginning of the study and the slope parameter(s) represents an organization's true rate of change in AYP over time. The Level 2 model tests how the intercept and slope(s) from Level 1 relate to between-subjects factors (e.g., psychological safety, felt accountability, and control variables). See Appendix A for details.

¹² Our data meet the requirements for applying multilevel modeling (Singer & Willett, 2003).

¹⁰ Brand-new schools (i.e., those that opened after 2005) were not yet required to meet AYP targets in 2007–2008 and were thus excluded from our sample.

post Time 1). Because the longitudinal data set spans three years, the model estimates growth trajectories of schools' likelihood of achieving their AYP targets over time.

We estimated five models to examine the likelihood of schools' achieving AYP over time step by step (see Appendix A). We first established a baseline model ("Model 1") in which time (the variable "Year") was the primary predictor of the likelihood of schools achieving AYP, along with all control variables (three time-varying and four time-invariant) to account for interorganizational differences in the intercept of AYP. The second model ("Model 2") built on the first by adding the main effects of psychological safety and felt accountability, both time-varying variables, which indicate the relationship with AYP in any given year (research questions 1 and 2). The third model ("Model 3") extended Model 2 by adding the interaction of psychological safety and felt accountability, which indicates whether the effect of psychological safety on the likelihood of schools successfully achieving AYP targets depends on felt accountability (Research Question 3). In Model 4, we expanded Model 3 by adding the interaction of psychological safety and time, as well as felt accountability with time, which indicate the extent to which these focal constructs relate to the slope of schools' likelihood of successfully achieving AYP over time. Lastly, in our final, full model ("Model 5"), we added the three-way interaction of psychological safety, felt accountability, and time, which represents the extent to which the interaction between psychological safety and felt accountability relates to the slope of schools' likelihood successfully achieving AYP over time (also research question 3).

RESULTS

Table 2 shows means, standard deviations, and correlations for all variables. The means for both psychological safety and felt accountability were relatively high (over 3 on a 4-point scale). Because this restricted range could limit our ability to find statistically significant effects, doing so would signify a particularly robust relationship (Wang, Liao, Zhan, & Shi, 2011). We found positive correlations between our focal variables of interest, psychological safety and felt accountability (r -values = 0.70, 0.67, and 0.74 at Times 1, 2, and 3, respectively),¹³ indicating that these elements of school culture tend to cooccur yet are still distinct. We conducted a multilevel confirmatory factor

analysis, where years were nested in schools, to examine the discriminant and convergent validity of the core constructs in our measurement model. We tested a four-factor multilevel model that reflected our conceptualization of organization-level psychological safety and felt accountability as separate constructs, and where each of these focal constructs was represented by a within-level factor and a between-level factor. This model fit the data reasonably well ($n = 173,973$, $\chi^2(52) = 60,866.91$, $p < .01$, CFI = .88, RMSEA = .08, SRMR = .07 for within and .05 for between), based on the acceptable fit evidenced by the CFI being very close to the standard of being .90 or above to indicate good fit, the RMSEA meeting the standard of being .08 or below, and the SRMR meeting the standard of being .08 or below (Bentler, 1990; Browne & Cudeck, 1992; Hu & Bentler, 1999). In addition, this four-factor model is superior to a two-factor model, where psychological safety and felt accountability were combined into one within-level factor and one between-level factor ($\Delta\chi^2(2) = 64,057.36$, $p < .01$; $n = 173,973$, $\chi^2(54) = 124,924.27$, $p < .01$, CFI = .76, RMSEA = .12, SRMR = .10 for within and .11 for between). Table 3 presents the logistic multilevel growth models for AYP. Figure 1 provides a graphical representation of the results of our full model.

Across our models, the significant, negative impact of time shows that schools were, on average, slightly less likely to meet their performance target over time (e.g., in our baseline model: $\gamma = -1.52$, $p < .01$; Table 3, Model 1). Performance in an absolute sense (e.g., test scores) might appear to have been declining. However, the reality was that the performance outcome, meeting the AYP target, became increasingly difficult to attain over time because the targets set for each school became more difficult to meet as the school got closer to the ultimate target of 100% proficiency by 2014.

In terms of our seven control variables, four had the expected relationships with performance (e.g., the percentage of students receiving federally subsidized meals and the percentage of students in special education were negatively related to the likelihood of achieving AYP targets), whereas the other three (limited English proficiency, school size, and teacher experience) were not significantly related to performance.

Our first two research questions asked how psychological safety and felt accountability, respectively, impact organizational performance over time. Even though organizations characterized by higher psychological safety might be expected to have a higher likelihood of achieving their AYP target, we found the opposite; that is, schools with higher psychological safety had a *lower* likelihood of achieving their AYP target in any given year ($\gamma = -2.89$, $p < .01$; Table 3,

¹³ We conducted tests of multicollinearity, including assessing the variance inflation factor for each construct. Results show values less than 2 for psychological safety and felt accountability, indicating that multicollinearity is not a major concern for our study (Ryan, 2009).

TABLE 2
Descriptive Statistics and Correlations among Study Variables

	Mean	SD	1	2	3	4	5	6	7	8	9	10
1. AYP (Time 1)	0.93	0.26	—									
2. AYP (Time 2)	0.74	0.44	0.33**	—								
3. AYP (Time 3)	0.69	0.46	0.27**	0.70**	—							
4. Felt accountability (Time 1)	3.40	0.31	0.22**	0.28**	0.31**	—						
5. Felt accountability (Time 2)	3.40	0.26	0.21**	0.32**	0.33**	0.68**	—					
6. Felt accountability (Time 3)	3.40	0.28	0.11**	0.31**	0.35**	0.59**	0.70**	—				
7. Psychological safety (Time 1)	3.03	0.39	0.01	0.15**	0.09*	0.70**	0.47**	0.42**	—			
8. Psychological safety (Time 2)	3.03	0.36	-0.01	0.14**	0.10*	0.44**	0.67**	0.51**	0.70**	—		
9. Psychological safety (Time 3)	3.02	0.37	-0.05	0.17**	0.14**	0.39**	0.46**	0.74**	0.58**	0.72**	—	
10. Physical safety (Time 1)	3.16	0.43	0.11**	0.36**	0.35**	0.79**	0.63**	0.54**	0.65**	0.48**	0.43**	—
11. Physical safety (Time 2)	3.21	0.41	0.11*	0.38**	0.38**	0.59**	0.77**	0.59**	0.49**	0.62**	0.46**	0.75**
12. Physical safety (Time 3)	3.16	0.44	0.05	0.37**	0.38**	0.49**	0.54**	0.79**	0.41**	0.46**	0.69**	0.65**
13. Prior performance (Pre-Time 1)	0.86	0.35	0.44**	0.55**	0.45**	0.15**	0.15**	0.16**	0.04	0.04	0.08	0.12**
14. Racial demographics (Average over time) ^a	11.80	19.89	0.14**	0.29**	0.33**	0.32**	0.29**	0.32**	0.22**	0.16**	0.20**	0.37**
15. Socioeconomic demographics (Average over time) ^a	68.48	21.91	-0.08	-0.31**	-0.35**	-0.21**	-0.17**	-0.20**	-0.21**	-0.16**	-0.19**	-0.31**
16. Limited English proficiency (Time 1)	14.33	14.63	-0.07	-0.13**	-0.15**	0.00	0.02	0.02	-0.03	-0.06	-0.04	0.05
17. Limited English proficiency (Time 2)	12.88	13.95	-0.09*	-0.17**	-0.19**	-0.03	-0.01	-0.02	-0.04	-0.05	-0.04	0.02
18. Limited English proficiency (Time 3)	14.90	15.31	-0.07	-0.17**	-0.18**	-0.03	0.00	-0.01	-0.05	-0.05	-0.04	0.02
19. Documented disabilities (Time 1)	14.76	6.71	0.07	-0.18**	-0.17**	-0.08	-0.11*	-0.09*	-0.09*	-0.09*	-0.10*	-0.19**
20. Documented disabilities (Time 2)	15.73	6.86	0.05	-0.21**	-0.20**	-0.08*	-0.11*	-0.11*	-0.08	-0.08	-0.09*	-0.21**
21. Documented disabilities (Time 3)	16.21	6.89	0.05	-0.24**	-0.23**	-0.11**	-0.13**	-0.14**	-0.07	-0.06	-0.09	-0.24**
22. School type: Middle school dummy (Average over time) ^a	0.19	0.39	0.06	-0.21**	-0.24**	-0.05	-0.16**	-0.06	0.02	-0.02	0.05	-0.03
23. School type: High school dummy (Average over time) ^a	0.21	0.41	-0.32**	-.07	-0.14**	-0.24**	-0.14**	-0.17**	0.10*	0.16**	0.12*	-0.06
24. School type: Other school type (Average over time) ^a	0.13	0.34	-0.02	-.06	.00	-0.02	-0.02	-0.04	-0.04	-0.06	-0.13**	-0.08
25. Size (Average over time) ^a	559.47	380.28	0.07	.08	0.12**	-0.04	0.05	0.04	-0.10*	-0.16**	-0.12**	0.04
26. Teacher Experience (Time 1)	8.47	2.93	0.17**	0.18**	0.23**	0.19**	0.19**	0.15**	-0.10*	-0.15**	-0.11*	0.12**
27. Teacher Experience (Time 2)	9.11	2.98	0.15**	0.14**	0.20**	0.17**	0.18**	0.15**	-0.10*	-0.15**	-0.11*	0.10*
28. Teacher Experience (Time 3)	9.54	2.86	0.11*	0.09*	0.16**	0.11**	0.14**	0.13**	-0.13**	-0.15**	-0.11*	0.05

Note. Pairwise correlations resulted in a range of $n = 519$ – 545 .

^a These variables are a mean of values over time and thus are time-invariant in the analyses.

* $p < .05$

** $p < .01$; two-tailed tests

TABLE 2
(Continued)

	11	12	13	14	15	16	17	18	19	20	21	22
11. Physical safety (Time 2)	—											
12. Physical safety (Time 3)	0.73**	—										
13. Prior performance (Pre-Time 1)	0.12**	0.15**	—									
14. Racial demographics (Average over time) ^a	0.31**	0.31**	0.20**	—								
15. Socioeconomic demographics (Average over time) ^a	-0.23**	-0.28**	-0.18**	-0.75**	—							
16. Limited English proficiency (Time 1)	0.06	0.08	-0.15**	-0.20**	0.40**	—						
17. Limited English proficiency (Time 2)	0.03	0.04	-0.17**	-0.22**	0.41**	0.98**	—					
18. Limited English proficiency (Time 3)	0.04	0.05	-0.17**	-0.21**	0.40**	0.98**	0.99**	—				
19. Documented disabilities (Time 1)	-0.18**	-0.19**	-0.10*	-0.04	0.17**	-0.18**	-0.16**	-0.18**	—			
20. Documented disabilities (Time 2)	-0.19**	-0.21**	-0.12**	-0.08	0.19**	-0.20**	-0.17**	-0.19**	0.95**	—		
21. Documented disabilities (Time 3)	-0.22**	-0.24**	-0.14**	-0.09*	0.19**	-0.22**	-0.18**	-0.20**	0.89**	0.95**	—	
22. School type: Middle school dummy (Average over time) ^a	-0.09*	-0.08	-0.12**	-0.04	0.04	-0.07	-0.04	-0.02	0.12**	0.15**	0.19**	—
23. School type: High school dummy (Average over time) ^a	-0.03	-0.01	-0.20**	-0.15**	-0.06	-0.02	0.01	0.00	-0.32**	-0.30**	-0.26**	-0.25**
24. School type: Other school type (Average over time) ^a	-0.04	-0.06	-0.02	-0.02	0.04	-0.01	-0.02	-0.01	-0.01	-0.03	-0.04	-0.19**
25. Size (Average over time) ^a	0.01	0.04	0.10*	0.16**	-0.09*	0.08	0.05	0.06	-0.11*	-0.15**	-0.18**	-0.04
26. Teacher Experience (Time 1)	0.12**	0.09*	0.18**	0.23**	-0.08*	0.03	-0.01	-0.01	0.09*	0.05	-0.01	-0.20**
27. Teacher Experience (Time 2)	0.10*	0.08	0.15**	0.21**	-0.06	0.04	0.00	0.00	0.10*	0.06	-0.01	-0.19**
28. Teacher Experience (Time 3)	0.05	0.06	0.12**	0.14**	-0.02	0.04	0.00	0.00	0.10*	0.06	0.00	-0.18**

Model 2). The results for felt accountability, however, confirmed our expectation: schools with higher felt accountability had a higher likelihood of achieving their AYP target in any given year over our study's three-year time frame ($\gamma = 3.52$, $p < .01$; Table 3, Model 2).

We explored our third research question, namely, how psychological safety and felt accountability together impact organizational performance over time, from two angles.¹⁴ First, the two-way interaction between psychological safety and felt accountability (as a predictor of the intercept of schools'

likelihood of achieving their AYP target) was not significant ($\gamma = 1.00$, n.s.; Table 3, Model 3), indicating that psychological safety's impact on schools' likelihood of achieving their AYP target in any given year did not depend on felt accountability. Second, however, the three-way interaction between psychological safety, felt accountability, and time (as a predictor of the slope of schools' likelihood of achieving their AYP target) tells a different story. The significant result for this three-way interaction suggests that the impact of psychological safety on schools' likelihood of achieving their AYP target varied by the amount of felt accountability present in the school *and* over time ($\gamma = -3.47$, $p < .05$; Table 3, Model 5). Interestingly, whereas the two-way interaction (psychological safety \times felt accountability) was not a significant predictor (in Model 3), the three-way

¹⁴ Additionally, we estimated our models with various covariance structures to assess the efficiency of our final model and the quality of our estimates. Our findings remained stable across these estimations.

TABLE 2
(Continued)

	23	24	25	26	27
23. School type: High school dummy (Average over time) ^a	—				
24. School type: Other school type (Average over time) ^a	-0.20**	—			
25. Size (Average over time) ^a	-0.06	-0.02	—		
26. Teacher Experience (Time 1)	-0.34**	-0.02	0.39**	—	
27. Teacher Experience (Time 2)	-0.33**	-0.03	0.39**	0.97**	—
28. Teacher Experience (Time 3)	-0.29**	-0.01	0.37**	0.91**	0.96**

interaction (psychological safety × felt accountability × time in Model 5) was significant, indicating that as time progressed, the effect of the interaction on organizational performance became larger. That is, the effect of the combination of psychological safety and felt accountability on organizational

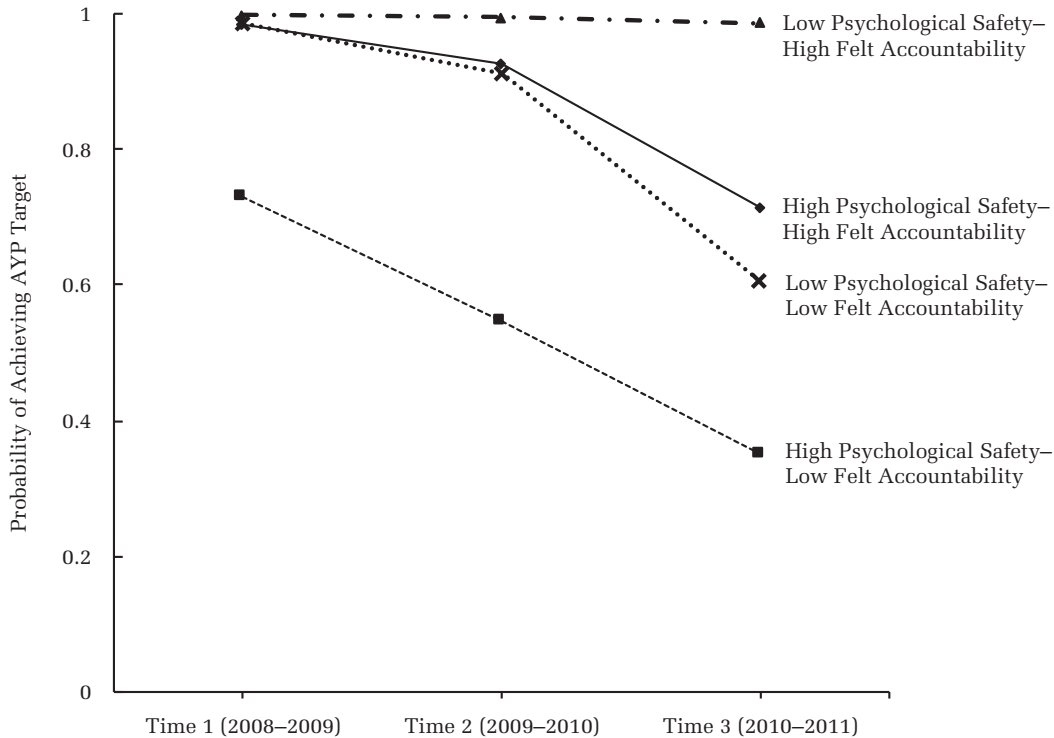
performance took hold over time. The two-way interaction did not directly include the role of time, and thus could not show this effect. Rather, we needed to explicitly include the role of time, as in the three-way interaction, to see the real effect of the combination of psychological safety and felt accountability

TABLE 3
Multilevel Logistic Growth Models: The Relationship between Psychological Safety and Felt Accountability with Organizational Performance (AYP) over Time

Parameter	Model 1		Model 2		Model 3		Model 4		Model 5	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
<i>Intercept</i>										
Constant	1.58	1.72	1.51	2.41	1.26	2.44	1.51	2.45	1.43	2.50
<i>Control Variables</i>										
Physical safety (Time-varying)	1.81**	0.32	1.90**	0.56	1.94**	0.57	1.89**	0.57	1.97**	0.58
Prior performance (Pre-Time 1)	3.08**	0.42	3.06**	0.46	3.05**	0.46	3.02**	0.46	3.07**	0.47
Racial demographics (Average over time)	0.09**	0.03	0.09*	0.04	0.09*	0.04	0.09*	0.04	0.09*	0.04
Socioeconomic demographics (Average over time)	-0.04**	0.01	-0.05**	0.02	-0.05**	0.02	-0.05**	0.02	-0.05**	0.02
Limited English proficiency (Time-varying)	-0.02	0.01	-0.02	0.01	-0.02	0.01	-0.02	0.01	-0.02	0.01
Documented disabilities (Time-varying)	-0.07**	0.02	-0.07*	0.03	-0.06*	0.03	-0.06*	0.03	-0.07*	0.03
School type: Middle school dummy (Average over time)	-2.05**	0.44	-1.68**	0.48	-1.69**	0.48	-1.65**	0.49	-1.65**	0.50
School type: High school dummy (Average over time)	-2.76**	0.52	-2.22**	0.58	-2.21**	0.59	-2.23**	0.58	-2.22**	0.59
School type: Other school type (Average over time)	-1.78**	0.49	-1.78**	0.53	-1.81**	0.54	-1.79**	0.54	-1.77**	0.54
Size (Average over time)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Teacher experience (Time-varying)	-0.01	0.06	-0.03	0.07	-0.03	0.07	-0.04	0.07	-0.04	0.07
<i>Focal Predictor Variables</i>										
Psychological safety (Time-varying)			-2.89**	0.80	-2.78**	0.81	-3.24**	1.13	-3.08*	1.15
Felt accountability (Time-varying)			3.52**	1.16	3.68**	1.18	3.30*	1.49	3.79**	1.55
Psychological safety (Time-varying) × Felt accountability (Time-varying)					1.00	1.02	0.43	1.09	2.61	1.67
<i>Slope</i>										
Year	-1.52**	0.19	-1.54**	0.23	-1.53**	0.23	-1.45**	0.25	-1.29**	0.25
Psychological safety (Time-varying) × Year							0.41	0.76	0.21	0.76
Felt accountability (Time-varying) × Year							0.60	1.11	0.19	1.12
Psychological safety (Time-varying) × Felt accountability (Time-varying) × Year									-3.47*	1.50
Log Likelihood	-448.27		-384.38		-383.87		-382.58		-379.55	
Wald χ^2	122.82**		92.40**		91.50**		90.72**		88.48**	

Notes. We report unstandardized coefficients; $n(\text{within}) = 1,541$, $n(\text{between}) = 518$.
* $p < .05$
** $p < .01$; two-tailed tests.

FIGURE 1
Fitted Probabilities of Schools Achieving AYP Targets Over Time as a Function of Low (Mean -1 SD) and High Levels (Mean $+1$ SD) of Psychological Safety and Felt Accountability



on organizational performance, thus reflecting the true longitudinal nature of these results (Dobrow & Weisman, 2021; Singer & Willett, 2003).

To help interpret our full model, especially the significant, negative effects of the three-way interaction that addresses our fourth research question about the best combination of psychological safety and felt accountability for organizational performance over time, we plotted our findings. Specifically, we plotted the probability of schools successfully achieving their AYP targets, varying both the level of psychological safety (where “high” is one standard deviation above the mean and “low” is one standard deviation below the mean) and the level of felt accountability (again, where high is one standard deviation above the mean and low is one standard deviation below the mean) over the three years of our study (Figure 1). Importantly, we note that these high and low values are relative; for instance, for psychological safety, the low value was 2.6 on a scale

of 1 to 4, meaning it was, on an absolute scale, somewhat moderate. Therefore, and again, importantly, when we use the word low for psychological safety, our reference point is the study sample, recognizing that it was not low in an absolute sense.

Interestingly, the combination of high psychological safety and low felt accountability performed the worst over time, including starting off worse (73% likelihood of achieving AYP target at Time 1) and declining more rapidly than any other combination (to 35% likelihood of achieving AYP target at Time 3). The other three combinations started out roughly the same—all close to a 100% likelihood of achieving their AYP targets at Time 1—but their trajectories over time showed notable differences. Schools with low psychological safety and high felt accountability started off well and declined only very slightly, such that their likelihood of achieving their AYP target remained close to 100% over time. By contrast, in spite of starting at the same point as the previous type of school climate, schools with high psychological safety and high felt accountability, as well as schools with low psychological safety and low felt accountability, declined much more over time, ending up with 71% and 61% likelihoods of achieving

Author’s Voice:

Was there anything that surprised you about the findings?



their AYP target, respectively. Thus, in sum, regarding our fourth research question about the best combination of psychological safety and felt accountability over time, we found the combination of relatively *low* psychological safety with relatively *high* felt accountability was the undisputed, yet unexpected, winner.¹⁵

Discovery-Oriented Supplementary Analyses

To further explore how various aspects of our findings might speak to our core research question, pertaining to *when psychological safety is helpful*, we conducted supplementary analyses in the spirit of discovery. First, we were intrigued by the possibility that the unexpected negative effect we found for psychological safety might be explained by a “too much of a good thing” (TMGT) effect (Pierce & Aguinis, 2013). This effect occurs when “antecedent variables widely accepted as leading to desirable consequences actually lead to negative outcomes” (Pierce & Aguinis, 2013: 313) due to the effect being curvilinear, rather than negative and linear. The results of our full model hint that a TMGT effect might be present (see Figure 1); however, when we reran our full model including an additional term to statistically test for the possibility of this curvilinear effect (psychological safety \times psychological safety \times time), it was not significant. This combination of results could be due to a ceiling effect or to relatively low levels of variability in our sample or, importantly, due to having three waves of data, which could be too few to reveal the

¹⁵ To examine the replicability of our results, we considered another dependent variable: the change in the percentage of students meeting the standard of performance, as indicated by New York State’s collection of English Language Arts Proficiency, measured at the third-grade level only. This measure is quite consistent across schools, but is limited by virtue of applying to only one grade level, and therefore is simply much narrower than the AYP. (We obtained these data from the NY State Department of Education’s online data archives, <https://data.nysed.gov/lists.php?type=district>) Specifically, these scores show—for students in a school who were in the third grade and took the test—the percentage who were proficient (i.e., scored at or above the state-determined threshold for proficiency). We got the same pattern of results as for our core dependent variable, AYP, thereby suggesting replication. Of primary interest, the three-way interaction of psychological safety, felt accountability, and time was again negative, but was not significant, likely because the sample size went down considerably due to focusing only on the third grade rather than on all grade levels (to $n(\text{within}) = 855$; $n(\text{between}) = 287$ vs. $n(\text{within}) = 1,365$; $n(\text{between}) = 518$ in our core analyses).

“true” pattern of psychological safety over the longer term. These ideas warrant future research.

In addition, we wondered whether exploring different compositional models of psychological safety and felt accountability (e.g., Chan, 1998) might shed light on how these aspects of workplace climate affect organizational performance. Rather than treating these focal constructs as aggregates whose measures indicate how relatively high or low these climate variables were in each school, we wanted to explore whether the *variance* in what teachers reported about their school’s climate mattered. As a first step, we note that the ICCs for both psychological safety and felt accountability were very strong in our data set (see Method section), which supports aggregating these variables from the individual teacher to the organization level and provides initial evidence of relatively low variance among teachers’ reports within a given school. Next, we calculated standard deviations for both focal constructs, which offer a direct, organization-level measure of dispersion of teachers’ responses, and added them as controls to our full model. We found that neither of the standard deviations was a significant predictor in this model, nor did including these variables alter our results. Thus, these supplementary results support our initial assumption that, in our sample, both psychological safety and felt accountability can be treated as collective, or shared, views of climate held by the teachers in a given school that can be used at the organization level both conceptually and methodologically. That said, we strongly suggest that future research more fully consider this question of consensus versus dispersion of both psychological safety and felt accountability, particularly in contexts in which greater dispersion may exist than in our sample.

DISCUSSION

The present study is the first to explore *when* psychological safety is helpful at the organization level over time. Specifically, by drawing on a three-year, three-wave multilevel study of a unique large-scale context, the NYC public school system, we aimed to build on the mixed findings in prior research about psychological safety’s effects, primarily found at the team level, by examining psychological safety’s boundary conditions at the level of the organization. With this in-depth investigation, we offer discoveries about psychological safety, along with another aspect of workplace climate, felt accountability, that we hope can inform future research and theory-building.

Our core findings were surprising in several respects, and offer insights that, along with our supplementary analyses, suggest avenues for future research.

In particular, we had expected to find a direct and positive relationship between organization-level psychological safety and organizational performance. Instead, when considered on its own, all else held constant, psychological safety had a *negative* effect on performance over time. By contrast, we found the opposite to be the case for organization-level felt accountability: it had a positive impact on schools' likelihood of reaching their performance targets over time, all else held constant. Considering psychological safety and felt accountability together in conjunction with time as a three-way interaction yielded our most interesting findings: the most promising work environment, in terms of yielding the best performance outcomes over time, was not, as might have been expected, when both elements were relatively high, or even when they were balanced (i.e., both relatively high or both relatively low). Rather, the best-performing schools in the NYC public school system were those with relatively low psychological safety and relatively high felt accountability.

These findings support our primary motivation for conducting this study, namely that compelling discoveries regarding the boundary conditions of psychological safety could be possible, particularly when considered at the organization level. Indeed, our results highlight that psychological safety's effect on performance is more nuanced than found in prior research, with higher psychological safety not simply or directly enabling better performance—and perhaps even hindering performance—when considered on its own. Yet, our findings cannot be explained fully by prior research, especially because we are treading into new territory given the level of analysis, the consideration of time as a critical and underexplored variable, and the particular context of the present study. As such, we believe our findings offer important opportunities for future research. We suggest that these opportunities for theory-building and subsequent empirical work fall into four main areas: (a) *attention* as a critical resource, (b) *task type and the role of interdependence*, (c) *outcomes beyond learning*, and (d) *the role of time in management research*.

Attention

The most robust of our findings concerned the three-way interaction effect we found between psychological safety, felt accountability, and time. Given these findings, we are persuaded that a possible area to explore next, both theoretically and empirically, is the concept of attention—here, referring to what teachers feel psychologically safe to talk about and what they feel accountable for. We are inspired by considering our findings along with those about a

construct adjacent to psychological safety—namely, psychological empowerment, or the extent to which individuals see their work as having “impact,” as being “important,” and as being “autonomous” (Wallace et al., 2011: 5). Wallace and colleagues (2011: 4) conceptualized felt accountability as enabling organizational order because it can help people remain aware of their task responsibilities and know what they should feel empowered to actually “do.” We can make a parallel conjecture here with our own study: the reason felt accountability can be so powerful alongside psychological safety is that it helps people know what they should exercise voice *about* when they feel psychologically safe.

At the center of these ideas is the notion of *attention*—namely, that *what* individual workers attend to when they feel psychologically safe to speak up matters. We suggest that attention should have an impact on whether psychological safety actually helps or hurts performance. In schools, as well as many other service-oriented contexts, particularly those that are high stakes, attention is a precious resource. That is, teachers do not simply need more time as a resource, they also need help attending to their instructional practice. Otherwise, without that attention, individuals may feel psychologically safe in a “comfort zone” (Edmondson, 2008) that does not help the organization. Such a comfortable space might morph into what we could call a “complaining zone,” rather than a “practice zone” in which people talk about the work at hand. We strongly encourage future research to explore conditions under which a complaining zone surfaces and, further, the extent to which this may interfere with the school's performance.

Indeed, some research in education has found suggestive evidence to this effect. Studies have found that when teachers are given an opportunity to engage in voice or decision-making (e.g., participating in an instructional leadership team), but do not have clear roles or a sense of what the discussion is about, their time together can devolve into discourse beyond the core work of teaching and learning (e.g., discussing the appropriate format for student presentations, scheduling, cafeteria lines, etc. [Weiner, 2014]). Further, as Deng and colleagues (2019: 1115) offered with their dual-pathway model of psychological safety, “people tend to exert less effort when they do not feel accountable.” Here, we are suggesting the possibility not of “less” effort on the part of teachers, but perhaps, instead, of *misplaced* effort, if the workplace climate is marked by high psychological safety accompanied by little felt accountability. Without that felt accountability and anchoring to specific targets and expectations for student learning, the psychologically safe environment could

backfire because it actually allows for distraction or attention being drawn *away* from the task at hand. Further, over time, these windows for complaining might accumulate, yielding a negative self-fueling cycle, which could also explain our negative coefficient for psychological safety.

We expect that high psychological safety might yield this kind of complaining zone or spiral when workers feel frustrated about their jobs, for example, as suggested by classic research on expectancy theory, when workers do not feel their effort necessarily translates into the performance expected (Vroom, 1964). Indeed, during the time period of this study, in the era of NCLB, teachers were under enormous pressure to perform and reported high levels of stress and frustration. In one longitudinal study of the effects of NCLB on teacher attitudes, scholars concluded that “much of the frustration [seemed] rooted in demands emanating from NCLB... [including] excessive paperwork, time shortages... shrinking curriculum, and prescribed lessons” (Smith & Kovacs, 2011: 218). Another study published during that same time emphasized that the teachers, under NCLB, were the ones who ultimately became the “victims of increased expectation and regimentation” (Rubin, 2011: 407). Thus, one interpretation of our negative findings for psychological safety is that teachers were complaining about issues associated with the pressures they faced under NCLB, which may have distracted them from working on problems of instructional practice.

Education research also provides insight into our findings with respect to the positive impact of felt accountability. Teachers who felt accountable, particularly in environments with relatively low psychological safety, may simply have been putting their heads down and focusing on the task at hand—improving instruction, which then yielded positive benefits in terms of student achievement. As Grissom and colleagues’ (2014) research on teacher attitudes pre- and post-NCLB found, even with the increase in pressure that teachers felt along with the number of hours worked, NCLB did, in general, enhance teachers’ sense of control in the classroom as well as intentions to remain with their school. Thus, this external accountability system, NCLB, may have enhanced felt accountability for teachers in some schools, which, over time, helped them improve student performance. Still, additional research is needed because this prior study did not examine teachers’ felt accountability (or psychological safety); as Grissom et al. (2014: 432) concluded, “more direct measures of accountability pressure may yield more nuanced results.”

We suggest that future research should investigate exactly what workers are attending to (or not attending to) when working in environments under duress,

when the pressure to change is high and the stakes are ever important but also increasingly difficult to attain. To delve deeper into the findings from the present research, qualitative research on the nature of worker conversations in different work climates is warranted. Future research could help us understand whether and when worker conversations unwittingly undermine the core task at hand, which, in this case, is to focus on instruction. The results from the present study cannot resolve these questions about underlying mechanisms because we did not have data on teacher conversations; however, they do raise the surprising possibility that psychological safety, on its own and over time, can actually hurt, rather than help, organization-level performance. Further, they corroborate prior suggestions (e.g., Deng et al., 2019; Edmondson, 2008) that psychological safety and felt accountability should be considered in tandem in management research.

Task Type and the Role of Interdependence

Although we did not initially set out to study the role of interdependence in understanding psychological safety, our findings suggest that this area may also be worthy of further investigation. The majority of research on psychological safety has examined teams, which are, by definition, contexts that entail at least some degree of interdependence among workers (Hackman, 2002). Indeed, Deng and colleagues (2019:1139) argued that interdependence, and even the mere presence of others in a workgroup, can yield two possible pathways for psychological safety: interdependence can “heighten the importance of a safe environment in reducing fear of failure” or can “impose a certain level of accountability to prevent effort withdrawal,” leading to their conclusion that it would be a “logical next step for future research [on psychological safety] to examine the role of interdependence.”

We agree with this recommendation and recognize that it might be even more important in contexts such as ours that do not involve natural forms of interdependence. As discussed, the task of teaching has traditionally involved very little natural interdependence, given the egg-crate structure of most U.S. schools, whereby teachers typically close their doors and teach with little opportunity to share their practice (Little, 1990; for a review, see Vangrieken et al., 2017). Further, teachers’ work is rarely “in the presence of others,” beyond their own students. Thus, the classic “mere presence of others” explanation (Zajonc, 1968) or the idea that “monitoring” may be associated with the presence of others (Deng et al., 2019) does not easily apply here. Given our findings, we question whether interdependence and felt

accountability might be substitutes in the psychological-safety-pathway equation—that is, whether felt accountability is especially important in contexts with little natural interdependence. In the present study, we did not juxtapose interdependent with independent work; all of the teachers were working in a similar context in which they had their own classes. However, future research could examine this distinction directly. For example, future research could compare, in a field experiment, ways of bolstering interdependence among people working on similar tasks that were not inherently interdependent to examine whether and how interdependence interacts with psychological safety and felt accountability to influence performance over time.

Beyond Learning Outcomes

In the extensive portfolio of research by Edmondson and colleagues, the outcomes of interest regarding psychological safety have largely been about learning behaviors, including risk-taking behavior and creativity. This focus has aligned well with the origins of psychological safety research and its embeddedness in the organizational learning literature (e.g., Argyris, 1982). Although some research has shifted to examine how psychological safety affects performance outcomes, the arguments have remained largely focused on mechanisms associated with adult learning (e.g., Edmondson, 2019). To extend prior research, we suggest that future psychological safety research consider, both theoretically and empirically, a variety of outcomes beyond learning to more fully understand when, and to what end, psychological safety is “helpful.”

In our study context, we speculate that one possible reason why schools performed best when teachers reported working in environments with relatively low psychological safety and high felt accountability is that this kind of workplace climate was conducive to adult learning, such that teachers were more innovative and figured out entirely new instructional techniques that then improved student performance. Although we do not have data to corroborate this interpretation of our findings regarding teacher learning behaviors, it would align well with prior psychological safety research. Another possibility, however, and one worthy of future exploration, is that something quite different could be going on here: perhaps, instead, this study could indicate the ways in which certain conditions enable teachers to focus on executing what they *already* know “works” rather than trying out something entirely new and different. Our study does not capture the typical types of learning outcomes studied in previous psychological safety studies (e.g., ingenuity or

Author’s Voice:

What is the social relevance of your research?



creativity); thus, our study may have captured a different way in which psychological safety can be “helpful,” when paired over time with felt accountability, opening up new avenues for research.

Foundational research by Hackman and Oldham (1980) on job design and workplace climate suggested that three “intermediate criteria” ultimately influence work effectiveness: (a) the level of effort brought to the task, (b) the amount of knowledge and skill brought to the task, and (c) the appropriateness of the task-performance strategies used (Hackman & Oldham, 1980: 170). Deng and colleagues’ (2019) study of how psychological safety affects work motivation maps onto the first condition, whereas the majority of studies on psychological safety have focused on the second—the ways in which psychological safety promotes the learning of new skills and knowledge. We suggest that our study may bring into the mix the third factor; namely, the extent to which psychological safety and felt accountability, together, help workers focus on appropriate task-*performance* strategies. Felt accountability might serve as a forcing or focusing mechanism—in a positive sense—helping workers attend to those task-performance strategies that are most appropriate, rather than, perhaps, learning entirely new ones.

Time

In our analyses, the two-way interaction between psychological safety and felt accountability was not significant, yet once we included time the resulting three-way interaction was significant, such that the effects of the two-way interaction became accentuated over time. Some of the inherent conceptual complexities of time as a construct may explain why this was the case. Of the six conceptual dimensions of time identified by George and Jones (2000), the interaction effect of psychological safety and felt accountability on performance reflects the “spirals and intensity” dimension.¹⁶ Per this dimension, the intensity of a phenomenon sometimes “increases in an upward direction, or decreases in a downward direction, nonlinearly and exponentially, sometimes

¹⁶ The other five dimensions are as follows: (a) the past, future, and present, and the subjective experience of time; (b) time aggregations; (c) duration of steady states and rates of change; (d) incremental versus discontinuous change; and (e) frequency, rhythms, and cycles.

over a short period of time” (George & Jones, 2000: 665). The “duration of steady states and rates of change” dimension, which focuses on the question of how long a particular state lasts or is stable over time, may have implications for spiral and intensity (George & Jones, 2000) but is not as direct a match for our phenomena—namely psychological safety and felt accountability. Understanding which dimension(s) of time are salient matters because it has implications for understanding the very nature of the phenomena in question.

The intensification of our phenomena is visually depicted in Figure 1, which highlights that, over time, the effect of the interaction of psychological safety and felt accountability becomes stronger—in this case, more negative. If we compare the left side of the figure (Times 1 to 2) with the right side of the figure (Times 2 to 3), we can see that the drop in each line is greater on the right side, meaning that the interaction effect intensifies over time. Of course, how to examine the spirals and intensity dimension of time appropriately is the subject of ongoing debate (George & Jones, 2000), and we acknowledge that our approach represents just one way of doing so. Nonetheless, by beginning to consider the nuances of time in the present study, we were able to elucidate that the intensity of the combined effect of psychological safety and felt accountability accelerated over time. We hope that future theory-building about psychological safety will further this inquiry by exploring questions including whether psychological safety and felt accountability spiral independently over time, as well as whether and why the relationships between constructs might spiral over time (George & Jones, 2000).

Implications for Practice

Our exploratory research suggests several practical implications for managers, especially when considering the role of workplace climate and how it might help or hinder organizational performance. In particular, our findings regarding the dark side of psychological safety were surprising and suggest that managers must be mindful of the impulse to simply provide a psychologically safe space for voice or to reduce barriers to speaking up. Free or open dialogue in and of itself may not produce the results intended. Simply opening up this kind of space to create a high psychological safety workplace might make individuals feel too “comfortable,” as Edmondson (2008) surmised. In addition, such a space could lead to unproductive “complaining” that might undermine performance. Indeed, our findings suggest that managers need to be thoughtful about how psychological safety might also have a dark side if it is not well-managed.

Thus, we recommend that managers thoughtfully open up the inquiry space, recognizing possible tripwires that could emerge. As many have suggested, the conditions that enable positive work outcomes must be “managed” (e.g., Edmondson, 2008). Indeed, managers could be served well by offering scaffolding to support that inquiry space through vehicles such as thoughtful agenda-setting, creating collective incentives, or designing meetings around problems of practice identified by those closest to the work itself (for an example of this approach in education, see Boudett, City, & Murnane, 2013). These actions might help keep the felt accountability high among workers, while also creating adequate space for psychological safety to emerge.

Limitations

Although our research context, the largest public school district in the United States, is a strength of our study, it also poses some limitations. This context is rife with conflict and calls for change. Under increasing scrutiny, schools fail to enjoy the “logic of confidence” suggested by Meyer and Rowan (1977) decades ago. Although it is in such contexts that education scholars and practitioners have turned their attention to levers for change that are “softer”—such as workplace climate—here, by studying such a strong situation, we may have also limited the generalizability of our findings. Therefore, although many private-sector domains also face increasing scrutiny, examining the impact of psychological safety and felt accountability, together, on organizational performance and over time in a wider range of contexts would be helpful.

We are also limited by the nature of our data. Specifically, because the NYCDOE designed the surveys, we were limited by the items included. For organization-level psychological safety, although we were able to secure approval from the teachers’ union to include our questions about psychological safety in the surveys, these items were included in only the electronic, and not the paper, version of the Time 1 survey, which thus limited our Time 1 sample to those teachers who had access to the electronic survey. Nonetheless, with over 25,000 teacher respondents at Time 1 and over double that at Times 2 and 3, we had a robust sample to analyze and from which to draw inferences. Further, we note that our study was limited by the range of psychological-safety levels in our sample. The best-performing schools in our sample—with “low” psychological safety and “high” felt accountability—still had psychological safety levels of about 2.6 on a 4-point scale. Thus, to more fully understand the impact of lower levels of psychological safety, future research

would benefit by selectively sampling for contexts in which psychological safety varies more widely, particularly organizations with low psychological safety in absolute terms. For organization-level felt accountability, we were fortunate to see just how closely the survey's six accountability-related items matched theory regarding felt accountability, such that we could reasonably consider them as a scale to measure this construct. However, we acknowledge the importance of expanding upon our discovery-oriented research by employing a more robust scale for felt accountability at the organization level in future work.

Additionally, the NYCDOE did not allow us to track individual teacher responses over time due to very stringent protocols regarding the confidentiality of these data. Nonetheless, we suggest that our type of data—teachers' reports aggregated to the organization level—is a solid reflection of what each school felt like as a place of work, regardless of the turnover of individual teachers. Our research focuses on capturing the culture and climate in schools, as reported by the teachers working at those schools at that particular time. This focus is supported by our ICC results, in that they show low variability in our focal constructs (psychological safety and felt accountability) within schools, meaning our measures, indeed, captured organization-level constructs.

Further, although we were fortunate to be able to examine our questions across three waves of multilevel data spanning three years, we were also limited by not having even more years in our data collection. Although we expect that time might accelerate the effect of the interaction between psychological safety and felt accountability on school performance, given that these constructs had positive relationships with outcomes in prior research, time could instead have been “slacking” the effect. In other words, the joint impact of psychological safety and felt accountability on school performance may need a period longer than three years for its true effects on performance to be revealed. As such, our study cannot rule out this slacking effect over time (e.g., Preacher, Curran, & Bauer, 2006). As is often the case, future research could benefit from a longer-term investigation allowing us to better understand the nature of the proposed time effect associated with the factors studied here.

Finally, we are limited in the ways in which we were able to study organizational performance in this study. We selected AYP as our focal dependent variable in large part because it was, practically speaking, the only performance measure available at the time that was psychometrically robust enough for use in our longitudinal analyses. Other performance measures of NYC schools exist but are very

problematic by virtue of such issues as measurement imprecision, instability from year to year due to changing standards, inconsistency across schools, and even lack of objectivity (due to such processes as political pressure) as to be rendered unusable for longitudinal analysis purposes (Di Carlo, 2012). Moreover, AYP was the only salient performance measure to employees in this context (i.e., teachers). In other words, failing to meet AYP had very real consequences for schools and teachers, and those consequences worsened over time if they failed from year to year (Smith & Kovacs, 2011). Thus, as discussed, schools and teachers felt considerable pressure around meeting this standard (Rubin, 2011). In spite of its limitations, this salience was not true of any other performance outcomes and made AYP the best option for use as our dependent variable.

CONCLUSION

This study provides the first investigation into when psychological safety is helpful for organizational performance over time. It is the first to empirically explore the boundary conditions of psychological safety with respect to one additional aspect of workplace climate that has been suggested multiple times in prior research but never empirically examined—felt accountability. We examined the fundamental question of when psychological safety is helpful by exploring the relationship between psychological safety, felt accountability, and organizational performance over time with a large data set of thousands of teacher responses in hundreds of schools over a three-year timeframe, using advanced multilevel modeling techniques.

Our discoveries highlight a new role for psychological safety, one not previously found in the management literature. Here, psychological safety never acted alone, or even positively, with regard to organizational performance, particularly over time. Rather, relative to others in the sample, schools with relatively low psychological safety combined with relatively high felt accountability demonstrated the best performance outcomes over time. Thus, much like a pinch of yeast—that is, a small dose that is neither nonexistent nor overly abundant—psychological safety acted as a catalyst in our study context, amplifying the positive effects of felt accountability on organizational performance over time. Further, psychological safety was never, on its own, positive; rather, it served as a critical amplifier in the quest to improve performance. The fact that psychological safety worked best when coupled with other levers, such as felt accountability, is both a new finding and one that offers many different opportunities for theorizing and empirical research. Our findings clearly highlight the role of felt accountability, and thus the

ways in which psychological safety not only acts as a catalyst but also may need to be directed—with felt accountability—so that workers feel safe speaking up and discussing matters that directly impact organizational performance.

We are hopeful that our discoveries regarding the conditions under which psychological safety helps improve performance at the organization level pique the interests of scholars studying psychological safety in organizations, felt accountability, and the role of time as they relate to how workplace climate may influence organizational performance. Further, we hope that the theoretical implications of this work for research on topics such as the role of attention, the interdependency of work, outcomes beyond learning, and time enhance our understanding of when psychological safety is indeed helpful with respect to organizational performance. Finally, we hope that our study encourages scholars to explore, as well as apply, management and organizational theory to the pressing needs and challenges associated with education—a context that is filled with organizations that must be well-managed and is ripe for change.

REFERENCES

- Argyris, C. 1982. *Reasoning, learning, and action: Individual and organizational*. San Francisco, CA: Jossey-Bass.
- Balfanz, R., & Legters, N. 2005. *The graduation gap: Using promoting power to examine the number and characteristics of high schools with high and low graduation rates in the nation and each state, CSOS Policy Brief*. Baltimore, MD: Johns Hopkins University, Center for Social Organization of Schools.
- Bamberger, P. A., & Pratt, M. G. 2010. Moving forward by looking back: Reclaiming unconventional research contexts and samples in organizational scholarship. *Academy of Management Journal*, 53: 665–671.
- Barrett, B. D. 2009. No Child Left Behind and the assault on teachers' professional practices and identities. *Teaching and Teacher Education*, 25: 1018–1025.
- Bentler, P. M. 1990. Comparative fit indexes in structural models. *Psychological Bulletin*, 107: 238–246.
- Bliese, P. D. 2000. Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*: 349–381. San Francisco, CA: Jossey-Bass.
- Boudett, K. P., City, E. A., & Murnane, R. J. 2013. *Data wise: A step-by-step guide to using assessment results to improve teaching and learning*. Cambridge, MA: Harvard Education Press.
- Bresman, H., & Zellmer-Bruhn, M. 2013. The structural context of team learning: Effects of organizational and team structure on internal and external learning. *Organization Science*, 24: 1120–1139.
- Browne, M. W., & Cudeck, R. 1992. Alternative ways of assessing model fit. *Sociological Methods & Research*, 21: 230–258.
- Bunderson, J. S., & Boumgarden, P. 2010. Structure and learning in self-managed teams: Why “bureaucratic” teams can be better learners. *Organization Science*, 21: 609–624.
- Busse, C., Kach, A. P., & Wagner, S. M. 2016. Boundary conditions: What they are, how to explore them, why we need them, and when to consider them. *Organizational Research Methods*, 20: 574–609.
- Castilla, E. J. 2015. Accounting for the gap: A firm study manipulating organizational accountability and transparency in pay decisions. *Organization Science*, 26: 311–333.
- Chan, D. 1998. Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology*, 83: 234–246.
- Deng, H., Leung, K., Lam, C. K., & Huang, X. 2019. Slack-ing off in comfort: A dual-pathway model for psychological safety climate. *Journal of Management*, 45: 1114–1144.
- Di Carlo, M. 2012. *The stability and fairness of New York City's school ratings*. Washington, DC: Albert Shanker Institute.
- Dinkes, R., Cataldi, E., Lin-Kelly, W., & Snyder, T. 2009. *Indicators of school crime and safety: 2008*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, US Department of Education, and Bureau of Justice Statistics, Office of Justice Programs, US Department of Justice.
- Dobrow, S. R., & Weisman, H. 2021. Only time will tell: Conducting longitudinal research of careers. In J. Tosti-Kharas & W. Murphy (Eds.), *Handbook for research methods in careers*: 164–190. Cheltenham, U.K.: Edward Elgar Publishing.
- Donaldson, M. L., Johnson, S. M., Kirkpatrick, C. L., Marinell, W. H., Steele, J. L., & Szczesiul, S. A. 2008. Angling for access, bartering for change: How second-stage teachers experience differentiated roles in schools. *Teachers College Record*, 110: 1088–1114.
- Edmondson, A. C. 1999. Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44: 350–383.
- Edmondson, A. C. 2003. Speaking up in the operating room: How team leaders promote learning in interdisciplinary action teams. *Journal of Management Studies*, 40: 1419–1452.
- Edmondson, A. C. 2008. The competitive imperative of learning. *Harvard Business Review*, 86. doi:10.1300/J026v23n02_03.

- Edmondson, A. C. 2019. *The fearless organization: Creating psychological safety in the workplace for learning, innovation, and growth*. Hoboken, NJ: John Wiley & Sons.
- Edmondson, A. C., Bohmer, R., & Pisano, G. 2001. Speeding up team learning. *Harvard Business Review*, 79: 125–134.
- Edmondson, A. C., & Lei, Z. 2014. Psychological safety: The history, renaissance, and future of an interpersonal construct. *Annual Review of Organizational Psychology and Organizational Behavior*, 1: 23–43.
- Edmondson, A. C., Higgins, M. C., Singer, S., & Weiner, J. M. 2016. Understanding psychological safety in health care and education organizations: A comparative perspective. *Research in Human Development*, 13: 65–83.
- Elmore, R. F. 2004. *School reform from the inside out: Policy, practice, and performance*. Cambridge, MA: Harvard Education Press.
- Frazier, M., Fainshmidt, S., Klinger, R., Pezeshkan, A., & Vracheva, V. 2017. Psychological safety: A meta-analytic review and extension. *Personnel Psychology*, 70: 113–165.
- Frink, D. D., & Klimoski, R. J. 2004. Advancing accountability theory and practice: Introduction to the human resource management review special edition. *Human Resource Management Review*, 14: 1–17.
- Frink, D. D., Hall, A. T., Perryman, A. A., Ranft, A. L., Hochwarter, W. A., Ferris, G. R., & Royle, T. M. 2008. Meso-level theory of accountability in organizations. *Research in Personnel and Human Resources Management*, 27: 177–245.
- Frink, D. D., & Klimoski, R. J. 1998. Toward a theory of accountability in organizations and human resource management. *Research in Personnel and Human Resources Management*, 16: 1–51.
- George, J. M., & Jones, G. R. 2000. The role of time in theory and theory building. *Journal of Management*, 26: 657–684.
- Glick, W. H. 1985. Conceptualizing and measuring organizational and psychological climate: Pitfalls in multilevel research. *Academy of Management Review*, 10: 601–616.
- Grissom, J. A., Nicholson-Crotty, S., & Harrington, J. R. 2014. Estimating the effects of No Child Left Behind on teachers' work environments and job attitudes. *Educational Evaluation and Policy Analysis*, 36: 417–436.
- Hackman, J. R. 1992. Group influences on individuals in organizations. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology*, vol. 3 (2nd ed.): 199–267. Palo Alto, CA: Consulting Psychologists Press.
- Hackman, J. R. 2002, July. New rules for team building. *Optimize*, 2779: 50–62.
- Hackman, J. R., & Oldham, G. R. 1980. *Work redesign*. Reading, MA: Addison-Wesley.
- Hall, A. T., Frink, D. D., Ferris, G. R., Hochwarter, W. A., Kacmar, C. J., & Bowe, M. G. 2003. Accountability in human resources management. In C. A. Schriesheim & L. L. Neider (Eds.), *New directions in human resource management*: 29–63. Greenwich, CT: Information Age Publishing.
- Hall, A. T., Royle, M. T., Brymer, R. A., Perrewé, P. L., Ferris, G. R., & Hochwarter, W. A. 2006. Relationships between felt accountability as a stressor and strain relations: The neutralizing role of autonomy across two studies. *Journal of Occupational Health Psychology*, 11: 87–99.
- Hall, A. T., Frink, D. D., & Buckley, M. R. 2017. An accountability account: A review and synthesis of the theoretical and empirical research on felt accountability. *Journal of Organizational Behavior*, 38: 204–224.
- Hess, F. M., & Petrilli, M. J. 2006. *No Child Left Behind primer*. New York, NY: Peter Lang Publishing.
- Higgins, M., Ishimaru, A., Holcombe, R., & Fowler, A. 2012. Examining organizational learning in schools: The role of psychological safety, experimentation, and leadership that reinforces learning. *Journal of Educational Change*, 13: 67–94.
- Hochbein, C., Mitchell, A. M., & Pollio, M. 2013. Gamed by the system: Adequate yearly progress as an indicator of persistently low-achieving school performance. *NASSP Bulletin*, 97: 270–289.
- Hu, L., & Bentler, P. M. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6: 1–55.
- Kozlowski, S. W. J., & Klein, K. J. 2000. A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*: 3–90. San Francisco, CA: Jossey-Bass.
- Leana, C., & Pil, F. 2006. Social capital and organizational performance: Evidence from urban public schools. *Organization Science*, 17: 353–366.
- LeBreton, J. M., & Senter, J. L. 2008. Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, 11: 815–852.
- Lenzenweger, M. F., Johnson, M. D., & Willett, J. B. 2004. Individual growth curve analysis illuminates stability and change in personality disorder features. *Archives of General Psychiatry*, 61: 1015–1024.
- Linn, R. L., Baker, E. L., & Betebenner, D. W. 2002. Accountability systems: Implications of requirements of the No Child Left Behind act of 2001. *Educational Researcher*, 31: 3–16.
- Little, J. W. 1990. The persistence of privacy: Autonomy and initiative in teachers' professional lives. *Teachers College Record*, 91: 510–536.

- Lortie, D. C. 1975. *Schoolteacher: A sociological study*. Chicago, IL: University of Chicago Press.
- Mackey, J. D., Brees, J. R., McAllister, C. P., Zorn, M. L., Martinko, M. J., & Harvey, P. 2018. Victim and culprit? The effects of entitlement and felt accountability on perceptions of abusive supervision and perpetration of workplace bullying. *Journal of Business Ethics*, 153: 659–673.
- Medina, J. 2010. State's exams became easier to pass, education officials say. *New York Times*: A18.
- Meyer, J. W., & Rowan, B. 1977. Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, 83: 340–363.
- Mitchell, T. R., & James, L. R. 2001. Building better theory: Time and the specification of when things happen. *Academy of Management Review*, 26: 530–547.
- Morgeson, F. P., & Hofmann, D. A. 1999. The structure and function of collective constructs: Implications for multilevel research and theory development. *Academy of Management Review*, 24: 249–265.
- Morrison, H. E., Morrison, J. D., & Bedford, S. 2007. Beyond numbers: Making AYP one student at a time. *Principal Leadership*, 7: 26–31.
- Morrow, S. L., McGonagle, A. K., Dove-Steinkamp, M. L., Walker, C. T., Marmet, M., & Barnes-Farrell, J. L. 2010. Relationships between psychological safety climate facets and safety behavior in the rail industry: A dominance analysis. *Accident; Analysis and Prevention*, 42: 1460–1467.
- Nathanson, L., Cole, R., Kemple, J. J., Lent, J., McCormick, M., & Segeritz, M. 2013. *New York city school survey 2008–2010: Assessing the reliability and validity of a progress report measure*. Research Alliance for New York City Schools. New York, NY: New York University.
- National Center for Education Statistics. 2020, July 24. Public high school graduation rates. Retrieved from https://nces.ed.gov/programs/coe/indicator_coi.asp.
- New York City Department of Education. 2019, June. DOE data at a glance. Retrieved from <https://www.schools.nyc.gov/about-us/reports/doe-data-at-a-glance>.
- New York State Office of the Governor. 2015. The state of New York's failing schools. Retrieved from <https://www.governor.ny.gov/sites/default/files/atoms/files/NYSFailingSchoolsReport.pdf>
- No Child Left Behind Act of 2001*. Pub. L. No. 107–110.
- Orfield, G., & Frankenberg, E. 2014. Increasingly segregated and unequal schools as courts reverse policy. *Educational Administration Quarterly*, 50: 718–734.
- Paige, R. 2006. No Child Left Behind: The ongoing movement for public education reform. *Harvard Educational Review*, 76: 461–473.
- Patil, S. V., Vieider, F., & Tetlock, P. E. 2014. Process versus outcome accountability. In M. Bovens R. E. Goodin & T. Schillemans (Eds.), *The Oxford handbook of public accountability*: 69–89. Oxford, U.K.: Oxford University Press.
- Pierce, J. R., & Aguinis, H. 2013. The too-much-of-a-good-thing effect in management. *Journal of Management*, 39: 313–338.
- Preacher, K. J., Curran, P. J., & Bauer, D. J. 2006. Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, 31: 437–448.
- Raudenbush, S. W., & Bryk, A. S. 2002. *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Reardon, S. F. 2011. The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In G. J. Duncan & R. J. Murnane (Eds.), *Whither opportunity? Rising inequality, schools, and children's life chances*: 91–116. New York, NY: Russell Sage Foundation.
- Reardon, S. F., Kalogrides, D., & Shores, K. 2019. The geography of racial/ethnic test score gaps. *American Journal of Sociology*, 124: 1164–1221.
- Riddle, W. & Kober, N. 2011. State policy differences greatly impact AYP numbers. Washington, DC: Center on Education Policy.
- Rubin, D. 2011. The disheartened teacher: Living in the age of standardisation, high-stakes assessments, and No Child Left Behind (NCLB). *Changing English*, 18: 407–416.
- Ryan, T. P. 2009. *Modern regression models* (2nd ed.). Hoboken, NJ: John Wiley & Sons.
- Sass, T. R., Hannaway, J., Xu, Z., Figlio, D. N., & Feng, L. 2012. Value added of teachers in high-poverty schools and lower poverty schools. *Journal of Urban Economics*, 72: 104–122.
- Schein, E. H., & Bennis, W. G. 1965. *Personal and organizational change through group methods: The laboratory approach*. New York, NY: Wiley.
- Schlenker, B., Britt, T., Pennington, J., Murphy, R., Doherty, K., & Kintsch, W. 1994. The triangle model of responsibility. *Psychological Review*, 101: 632–652.
- Schulte, M., Cohen, N. A., & Klein, K. J. 2012. The coevolution of network ties and perceptions of team psychological safety. *Organization Science*, 23: 564–581.
- Shapiro, E. S., Solari, E., & Petscher, Y. 2008. Use of a measure of reading comprehension to enhance prediction on the state high stakes assessment. *Learning and Individual Differences*, 18: 316–328.
- Singer, J. D., & Willett, J. B. 2003. *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.

- Smith, J. M., & Kovacs, P. E. 2011. The impact of standards-based reform on teachers: The case of “No Child Left Behind.” *Teachers and Teaching*, 17: 201–225.
- Song, Y., Liu, Y., Shi, J., & Wang, M. 2017. Use of proactive socialization tactics and socialization outcomes: A latent growth modeling approach to understanding newcomer socialization process. *Academy of Management Discoveries*, 3: 42–63.
- Spillane, J. P., Camburn, E. M., Pustejovsky, J., Pareja, A. S., & Lewis, G. 2008. Taking a distributed perspective: Epistemological and methodological tradeoffs in operationalizing the leader-plus aspect. *Journal of Educational Administration*, 46: 189–213.
- Taylor, J., Stecher, B., O’Day, J., Naftel, S., & Le Floch, K. C. 2010. *State and local implementation of the No Child Left Behind act, volume IX—Accountability under NCLB: Final report*. Washington, DC: US Department of Education.
- Tetlock, P. E., & Mellers, B. A. 2011. Structuring accountability systems in organizations: Key trade-offs and critical unknowns. In B. Fischhoff & C. Chauvin (Eds.), *Intelligence analysis: Behavioral and social scientific foundations*: 249–270. Washington, D.C.: The National Academies Press.
- Tucker, A. L., Nembhard, I. M., & Edmondson, A. C. 2007. Implementing new practices: An empirical study of organizational learning in hospital intensive care units. *Management Science*, 53: 894–907.
- U.S. Census Bureau. 2018. *Annual survey of school system finances (FY 2016)*. Washington, DC: US Department of Commerce.
- U.S. Department of Education. 2006. *National assessment of Title I interim report: Executive summary*. Washington, DC: Institute of Education Sciences.
- Usher, A. 2012. *AYP results for 2010–11—May 2012 update*. Washington, DC: Center on Education Policy.
- Vangrieken, K., Grosemans, I., Dochy, F., & Kyndt, E. 2017. Teacher autonomy and collaboration: A paradox? Conceptualising and measuring teachers’ autonomy and collaborative attitude. *Teaching and Teacher Education*, 67: 302–315.
- Vroom, V. H. 1964. *Work and motivation*. New York, NY: Wiley.
- Wallace, J. C., Johnson, P. D., Mathe, K., & Paul, J. 2011. Structural and psychological empowerment climates, performance, and the moderating role of shared felt accountability: A managerial perspective. *Journal of Applied Psychology*, 96: 840–850.
- Wang, M., Liao, H., Zhan, Y., & Shi, J. 2011. Daily customer mistreatment and employee sabotage against customers: Examining emotion and resource perspectives. *Academy of Management Journal*, 54: 312–334.
- Weick, K. E. 1976. Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, 21: 1–19.
- Weiner, J. M. 2014. Disabling conditions: Investigating instructional leadership teams in action. *Journal of Educational Change*, 15: 253–280.
- Winerip, M. 2011. 10 years of assessing students with scientific exactitude. *New York Times*: A24.
- Wong, G. Y., & Mason, W. M. 1985. The hierarchical logistic regression model for multilevel analysis. *Journal of the American Statistical Association*, 80: 513–524.
- Wright, T. A. 1997. Time revisited in organizational behavior. *Journal of Organizational Behavior*, 18: 201–204.
- York-Barr, J., & Duke, K. 2004. What do we know about teacher leadership? Findings from two decades of scholarship. *Review of Educational Research*, 74: 255–316.
- Zaheer, S., Albert, S., & Zaheer, A. 1999. Time scales and organizational theory. *Academy of Management Review*, 24: 725–741.
- Zajonc, R. B. 1968. Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology Monographs*, 9: 1–27.



Monica C. Higgins (monica_higgins@gse.harvard.edu) is the Kathleen McCartney Professor of Education Leadership at the Harvard Graduate School of Education, where she teaches and conducts research on leadership and organizational change. Her research has examined developmental networks and leadership development, and, more recently, social entrepreneurship in education.

Shoshana R. Dobrow (s.r.dobrow@lse.ac.uk) is an assistant professor of management at the London School of Economics. Her research examines careers—particularly the meaning of work, the sense of calling, and developmental mentoring networks—often utilizing longitudinal approaches.

Jennie Weiner (jennie.weiner@uconn.edu) is an associate professor of educational leadership at the Neag School of Education at the University of Connecticut. Her scholarship focuses on educational leadership and organizational change in underperforming schools and districts, gender and racial bias in educational leadership, and educational infrastructure at local, district, and state levels.

Haiyang Liu (h.liu34@lse.ac.uk) is an assistant professor of management at the London School of Economics and Political Science. His research focuses on leader personality, employee well-being, and new technology.



APPENDIX A

DEFINITIONS AND FORMULAE FOR MULTILEVEL MODELS

We estimated five core multilevel logistic growth models, Models 1–5, for our outcome variable, organizational performance (AYP). We combined the Levels 1 and 2 models for all models into composite models as follows (per Singer & Willett, 2003):

- (1) *Baseline model*: This model includes all control variables plus the time variable (Year). These control variables are tested as predictors of the initial log odds ratio of the outcome (i.e., the intercept):

$$\ln \left[\frac{\hat{p}(\text{AYP}_{ij} = 1)}{1 - \hat{p}(\text{AYP}_{ij} = 1)} \right] = \hat{\beta}_{00} + \hat{\beta}_{01}\text{Physical Safety}_{ij} + \hat{\beta}_{02}\text{Prior Performance} + \hat{\beta}_{03}\text{Racial Demographics} + \hat{\beta}_{04}\text{Socioeconomic Demographics} + \hat{\beta}_{05}\text{Limited English Proficiency}_{ij} + \hat{\beta}_{06}\text{Documented Disability}_{ij} + \hat{\beta}_{07}\text{School Type : Middle School Dummy} + \hat{\beta}_{08}\text{School Type : High School Dummy} + \hat{\beta}_{09}\text{School Type : Other School Type} + \hat{\beta}_{010}\text{Size} + \hat{\beta}_{011}\text{Teacher Experience}_{ij} + \hat{\beta}_{012}\text{Year}_{ij}.$$

Here $\ln \left[\frac{\hat{p}(\text{AYP}_{ij} = 1)}{1 - \hat{p}(\text{AYP}_{ij} = 1)} \right]$ is the log odds ratio (logit) of the predicted value of the probability of achieving AYP (= 1) for school j at time i , and $\hat{\beta}_{00}$ is the estimated intercept (the estimated value of the outcome when the predictor $\text{Year}_{ij} = 0$; namely, when the school year is 2008–2009). $\hat{\beta}_{01}$ through $\hat{\beta}_{011}$ are the estimated coefficients for the 11 control variables. $\hat{\beta}_{012}$ is the estimated coefficient for the slope, which quantifies the estimated amount of linear change in the outcome per year.

- (2) *Adding main effects of core constructs as predictors of intercept*: This model includes all control variables, psychological safety, and felt accountability as predictors of the initial log odds ratio of the outcome (i.e., the intercept), plus the time variable (Year). The fitted full model equation is

$$\ln \left[\frac{\hat{p}(\text{AYP}_{ij} = 1)}{1 - \hat{p}(\text{AYP}_{ij} = 1)} \right] = \hat{\beta}_{00} + \hat{\beta}_{01}\text{Physical Safety}_{ij} + \hat{\beta}_{02}\text{Prior Performance} + \hat{\beta}_{03}\text{Racial Demographics} + \hat{\beta}_{04}\text{Socioeconomic Demographics} + \hat{\beta}_{05}\text{Limited English Proficiency}_{ij} + \hat{\beta}_{06}\text{Documented Disability}_{ij} + \hat{\beta}_{07}\text{School Type : Middle School Dummy} + \hat{\beta}_{08}\text{School Type : High School Dummy} + \hat{\beta}_{09}\text{School Type : Other School Type} + \hat{\beta}_{010}\text{Size} + \hat{\beta}_{011}\text{Teacher Experience}_{ij} + \hat{\beta}_{012}\text{Year}_{ij} + \hat{\beta}_{013}\text{Psychological Safety}_{ij} + \hat{\beta}_{014}\text{Felt Accountability}_{ij}.$$

Here, the new $\hat{\beta}_{013}$ and $\hat{\beta}_{014}$ are the estimated coefficients for the two focal predictor variables.

- (3) *Adding interactions of core constructs as predictors of intercept*: This model includes all control variables, psychological safety, felt accountability, and the interaction of psychological safety and felt accountability as predictors of the initial log odds ratio of the outcome (i.e., the intercept), plus the time variable (Year). The fitted full model equation is

$$\ln \left[\frac{\hat{p}(\text{AYP}_{ij} = 1)}{1 - \hat{p}(\text{AYP}_{ij} = 1)} \right] = \hat{\beta}_{00} + \hat{\beta}_{01}\text{Physical Safety}_{ij} + \hat{\beta}_{02}\text{Prior Performance} + \hat{\beta}_{03}\text{Racial Demographics} + \hat{\beta}_{04}\text{Socioeconomic Demographics} + \hat{\beta}_{05}\text{Limited English Proficiency}_{ij} + \hat{\beta}_{06}\text{Documented Disability}_{ij} + \hat{\beta}_{07}\text{School Type : Middle School Dummy} + \hat{\beta}_{08}\text{School Type : High School Dummy} + \hat{\beta}_{09}\text{School Type : Other School Type} + \hat{\beta}_{010}\text{Size} + \hat{\beta}_{011}\text{Teacher Experience}_{ij} + \hat{\beta}_{012}\text{Year}_{ij} + \hat{\beta}_{013}\text{Psychological Safety}_{ij} + \hat{\beta}_{014}\text{Felt Accountability}_{ij} + \hat{\beta}_{015}\text{Psychological Safety}_{ij} \times \text{Felt Accountability}_{ij}.$$

Here, the new $\hat{\beta}_{0.15}$ is the estimated coefficient for the interaction effect of the two focal predictors.

- (4) *Adding main effects of core constructs as predictors of slope*: This model includes all control variables, psychological safety, felt accountability, and the interaction of psychological safety and felt accountability as predictors of the initial log odds ratio of the outcome (i.e., the intercept), the time variable (Year), plus psychological safety and felt accountability as predictors of change in (i.e., the slope of) the log odds ratio of the outcome over time. The fitted full model equation is

$$\ln \left[\frac{\hat{p}(AYP_{ij} = 1)}{1 - \hat{p}(AYP_{ij} = 1)} \right] = \hat{\beta}_{00} + \hat{\beta}_{01}\text{Physical Safety}_{ij} + \hat{\beta}_{02}\text{Prior Performance} + \hat{\beta}_{03}\text{Racial Demographics} + \hat{\beta}_{04}\text{Socioeconomic Demographics} + \hat{\beta}_{05}\text{Limited English Proficiency}_{ij} + \hat{\beta}_{06}\text{Documented Disability}_{ij} + \hat{\beta}_{07}\text{School Type : Middle School Dummy} + \hat{\beta}_{08}\text{School Type : High School Dummy} + \hat{\beta}_{09}\text{School Type : Other School Type} + \hat{\beta}_{010}\text{Size} + \hat{\beta}_{011}\text{Teacher Experience}_{ij} + \hat{\beta}_{012}\text{Year}_{ij} + \hat{\beta}_{013}\text{Psychological Safety}_{ij} + \hat{\beta}_{014}\text{Felt Accountability}_{ij} + \hat{\beta}_{015}\text{Psychological Safety}_{ij} \times \text{Felt Accountability}_{ij} + \hat{\beta}_{016}\text{Psychological Safety}_{ij} \times \text{Year}_{ij} + \hat{\beta}_{017}\text{Felt Accountability}_{ij} \times \text{Year}_{ij}.$$

Here, the new $\hat{\beta}_{016}$ and $\hat{\beta}_{017}$ are the estimated coefficients for the two-way interactions of each of the two focal predictors and time.

- (5) *Full model: Adding interactions of core constructs as predictors of slope:* This model includes all control variables, psychological safety, felt accountability, and the interaction of psychological safety and felt accountability as predictors of the initial log odds ratio of the outcome (i.e., the intercept), the time variable (Year), plus psychological safety, felt accountability, and their interaction as predictors of change (i.e., the slope of) in the log odds ratio of the outcome over time. The fitted full model equation is

$$\ln \left[\frac{\hat{p}(AYP_{ij} = 1)}{1 - \hat{p}(AYP_{ij} = 1)} \right] = \hat{\beta}_{00} + \hat{\beta}_{01}\text{Physical Safety}_{ij} + \hat{\beta}_{02}\text{Prior Performance} + \hat{\beta}_{03}\text{Racial Demographics} + \hat{\beta}_{04}\text{Socioeconomic Demographics} + \hat{\beta}_{05}\text{Limited English Proficiency}_{ij} + \hat{\beta}_{06}\text{Documented Disability}_{ij} + \hat{\beta}_{07}\text{School Type : Middle School Dummy} + \hat{\beta}_{08}\text{School Type : High School Dummy} + \hat{\beta}_{09}\text{School Type : Other School Type} + \hat{\beta}_{010}\text{Size} + \hat{\beta}_{011}\text{Teacher Experience}_{ij} + \hat{\beta}_{012}\text{Year}_{ij} + \hat{\beta}_{013}\text{Psychological Safety}_{ij} + \hat{\beta}_{014}\text{Felt Accountability}_{ij} + \hat{\beta}_{015}\text{Psychological Safety}_{ij} \times \text{Felt Accountability}_{ij} + \hat{\beta}_{016}\text{Psychological Safety}_{ij} \times \text{Year}_{ij} + \hat{\beta}_{017}\text{Felt Accountability}_{ij} \times \text{Year}_{ij} + \hat{\beta}_{018}\text{Psychological Safety}_{ij} \times \text{Felt Accountability}_{ij} \times \text{Year}_{ij}.$$

Here, the new $\hat{\beta}_{018}$ is the estimated coefficient for the three-way interaction of the two focal predictors and time.