Positive Teacher-student Relationships May Lead to Better Teaching

Xintong Li[1], Christi Bergin[1], Amanda A. Olsen[2]

[1] University of Missouri, USA

[2] The University of Texas at Arlington, USA

Author Note

This is a preprint, please go to https://doi.org/10.1016/j.learninstruc.2022.101581 for the published article.

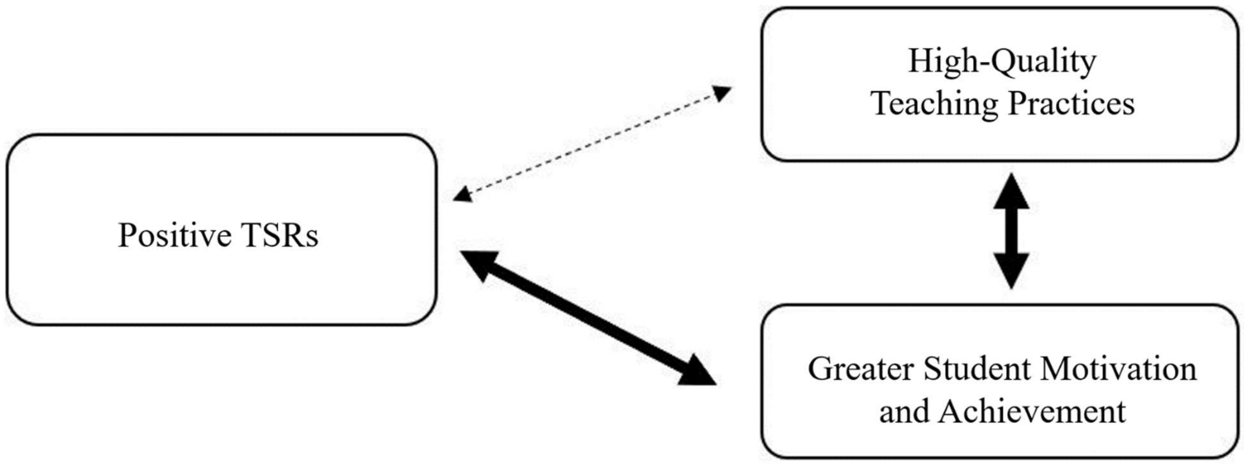
Correspondence concerning this preprint should be addressed to
Xintong Li
Assessment Resource Center
College of Education and Human Development
University of Missouri.
2800 Maguire blvd, Columbia, MO, 65201, USA.
lixint@missouri.edu

1. Introduction

Substantial research literature indicates that positive teacher-student relationships (TSRs) promote students' academic achievement. One explanation is that students are more motivated to learn when they have positive relationships with teachers (Urdan & Schoenfelder, 2006). However, another plausible explanation is that teachers engage in higher quality teaching practices when they have positive relationships with students. This is important because research on school effectiveness consistently identifies high-quality instruction as one of the largest school-based contributors to students' academic achievement (Scheerens, 2001; Thoonen et al., 2011). In the current study, we explored this alternative explanation. Figure 1 depicts the relationship among TSRs, teaching practices, and student outcomes. There is extant research evidence to support the bold arrows, which we briefly review next. Our study investigated the dashed arrow representing the bidirectionality between positive TSRs and high-quality teaching practices which, to our knowledge, no study has investigated. We used archival data from an authentic teacher evaluation system in the U.S.

Figure 1

Relationships Among Teacher-Student Relationships, Student Outcomes, and Teaching Quality



1.1 Positive Teacher-Student Relationships (TSRs) and Student Outcomes

TSRs can be characterized on multiple dimensions such as conflict, closeness, dependency, circularity, communication, or involvement (Pianta, 2001; Roorda et al., 2011; Wubbels et al., 2006). In this study we focused on positive affect, specifically caring and mutual enjoyment. When students feel teachers care about them, they work harder, engage in more challenging academic activities, behave more appropriately for the school environment, are genuinely happy to see their teacher, and meet or exceed their teacher's expectations (Allen et al., 2011; Bergin & Bergin, 2009; Hughes et al., 2008; Prewett et al., 2018; Quinn, 2017; Wentzel, 2009). As positive *Teacher-Student Relationships* motivate students to become more engaged in the classroom, learning should improve. Indeed, research confirms that positive TSRs are associated with higher grade point averages and test scores from kindergarten through high school (e.g., Baker, 2006; Cornelius-White, 2007; Curby et al., 2009; Hughes, 2011; Jia et al., 2009; McCormick et al., 2013; O'Connor & McCartney, 2007). A meta-analysis found that TSRs characterized by positive affect were strongly linked to student engagement and modestly linked to achievement (Roorda et al., 2011). Importantly, the effect size may be larger for at-risk students, suggesting that positive TSRs may help narrow the achievement gap (Gehlbach et al., 2016).

Self-determination Theory (SDT) has been used to explain a potential mechanism for the effect of positive TSRs on students' engagement and achievement (Kincade et al., 2020; Roorda et al., 2011). According to SDT, caring relationships meet a basic, innate need for relatedness or feeling connected to others (Jang et al., 2010; Ryan & Deci, 2017). When relatedness needs are met in a specific context, such as the classroom, individuals are more motivated behave in adaptive ways, engage with tasks, persist in the face of failure, and respond creatively to challenges in that context (Deci & Ryan, 2000).

1.2 Positive Teacher-Student Relationships (TSRs) and Teacher Outcomes

Just as positive TSRs may increase students' motivation, positive TSRs may also increase teachers' motivation, effort, engagement, happiness, and confidence, which in turn, may result in greater use of complex, high-impact teaching practices (van der Lans et al., 2020). Despite teachers' presumed basic need for relatedness with students in their classroom as predicted by SDT, there is a paucity of research on the effect of TSRs on teachers. Spilt and colleagues state, "there is little recognition of the internal needs that teachers themselves may have for positive, personal relationships with individual students" (Spilt et al., 2011, p. 458). Studies have found that positive TSRs may promote teachers' self-efficacy (Mashburn et al., 2006), emotional well-being (Milatz et al., 2015), and job satisfaction (Admiraal et al., 2019; Veldman et al., 2016), but they do not address how positive TSRs may be associated with quality of teaching practices.

Although research is limited in the field of education, the field of organizational psychology has addressed how positive relationships may drive adult performance. The Leader-Member Exchange (LMX) theory that is commonly applied to adult workplaces is related to

SDT (Graves & Luciano, 2013). LMX theory asserts that the quality of a relationship between a leader and a follower (akin to the relationship between a teacher and a student) predicts performance (Gerstner & Day, 1997). A meta-analysis found that in adults, the link between relationship quality and performance is likely mediated by motivation (Martin et al., 2016). The development of deep connections at work results in higher quality work, increased effort, and decreased attrition among adults (Manning, 2016). It is plausible that a similar effect occurs in schools such that positive TSRs results in teachers' increased teaching effort and performance.

1.3 Teaching Practices and Student Outcomes

Our study uses archival data from an authentic teacher growth and evaluation system called the Network for Educator Effectiveness (NEE). In this evaluation system, teachers are rated on four effortful, complex, high-impact teaching practices that are known to promote student achievement (Hattie, 2009; van de Grift et al., 2014; van der Lans et al., 2017). The association between positive TSRs and quality of teaching practices may depend on the specific teaching practices under consideration. Other related lines of research have found that teachers' self-efficacy (Burić & Kim, 2020; Holzberger et al., 2013), job satisfaction, and motivation (Klusmann et al., 2008) were associated differently with specific teaching practices. Therefore, our study examined how positive TSRs are associated with the quality of each of the four different teaching practices. We use the definition of each teaching practice as operationalized by NEE which are described next.

1.3.1 Cognitive Engagement (CE) in the Content

In the NEE system, *Cognitive Engagement* refers to active mental involvement by students in learning activities, such as meaningful processing, strategy use, concentration, and metacognition (Fredricks et al., 2004; Wang & Degol, 2014; Wang et al., 2014). Teachers engage students cognitively in the content when they use strategies such as advanced organizers, K-W-L charts, share-out, and shoulder-partner to connect instruction/activities with students' lives. This allows teachers to show relevance, use authentic examples, present a puzzling problem, and invite responses from all students.

1.3.2 Problem-Solving and Critical Thinking (PCT)

In the NEE system, *Problem-Solving and Critical Thinking* refers to skillfully applying, analyzing, synthesizing, and evaluating information to reach a conclusion or solve a problem (McCormick et al., 2015). Teachers promote their students' critical thinking by requiring students to explain or justify their thinking, evaluate others' thinking, formulate challenging questions, predict, determine what makes an argument valid, assess possible solutions, categorize problems, or create map concepts (Wirkala & Kuhn, 2011). Teachers can also give students challenging tasks that require persistent effort and various cognitive or metacognitive strategies. In typical classrooms, the promotion of critical thinking is not always appropriate because there are times when students should be practicing and over-learning skills that are foundational to higher-level processing. Yet, this complex teaching practice may be too rare in typical classrooms (Willingham, 2008).

1.3.3 Affective Engagement (AE) in the Content

In the NEE system, *Affective Engagement in the Content* refers to experiencing positive emotion during the lesson. Teachers engage students affectively in the content with strategies such as using materials and activities that students find interesting, using authentic examples, pointing out progress, communicating enthusiasm for the content, helping students set achievable but challenging goals, promoting self-efficacy, and giving students choices for classwork

(Bergin, 1999; Hidi & Renninger, 2006). As they use such strategies, teachers make lessons more enjoyable, fun, and interesting (Archambault et al., 2017). AE refers to positive emotions toward the content or lesson activities, not toward the teacher.

1.3.4 Instructional Monitoring (IM) During the Flow of the Lesson

In the NEE system, *Instructional Monitoring* refers to the teacher engaging in formative assessment for the whole class and/or individuals and taking corrective action when needed. This particular kind of formative assessment involves quick checks for understanding as the lesson is progressing (Reddy et al., 2017). The purpose is to inform the modification of teaching and learning activities in real-time to guide instruction. Strategies teachers may use include questioning, asking students to solve problems on a whiteboard, or answering spot quizzes with fist-to-five, thumbs up, or clicker techniques (Chien et al., 2016). IM has been associated with student motivation, engagement, and achievement (Simons & Klein, 2007; Zhang & Hyland, 2018).

1.4 Research Questions

Teaching is a relentless and demanding job. It is particularly demanding when teachers are expected to consistently implement effortful, complex, high-impact teaching practices. Given that both SDT and LMX theory predict that positive TSRs should contribute to increased motivation and performance among teachers, we hypothesize that variation in the quality of complex teaching practices will result from variation in TSRs. Thus, our main research question is “Do positive teacher-student relationships promote high-quality teaching practices?” In addition, we have two ancillary research questions: “Does the effect vary by teaching practice?” and “Does the effect vary by grade level?” We are interested in grade level variation because studies find that close, positive teacher-student relationships often decrease steadily from 1st grade to the end of primary school, and across middle school, before possibly stabilizing in high school (Gillen-O'Neel & Fuligni, 2013; Hughes et al., 2012). The change in TSRs across grades may alter the relationship between TSRs and the quality of instruction.

2. Methods

We addressed our research questions using archival data from a state-wide teacher growth and evaluation system, the Network for Educator Effectiveness (NEE). This system is used by 285 preK-12 school districts across the state of Missouri in the United States. NEE's teaching effectiveness measures are aligned with the Interstate Teacher Assessment and Support Consortium (InTASC) teaching standards (Council of Chief State School Officers, 2011), which apply to all subject areas and grade levels.

2.1 Participants

NEE member school districts select three to six teaching practices to focus on from a pool of 26 teaching practices based on district priorities. Thus, data for this study comes from a subset of member districts that decided to focus on the four target teaching practices of interest. As a result, our analysis of each teaching practice involves overlapping but different samples (see Table 1). The data used in the current study are based on the means of student reports aggregated at the teacher level and are nested in schools. These data include two waves of student surveys in the 2017-2018 and 2018-2019 academic years. Only teachers of regular education classrooms from 4^{th} to 10^{th} grade were included (i.e., no students in our sample indicated teacher subject areas as either English language learning or special education).

The 2019 Missouri average proficiency rate on the state test was 49% for English language arts and 42% for mathematics. In comparison, the averages of our total sample were also 49% for English language arts, but 31% for mathematics. In Missouri, 73% of k-12 students were White, non-Hispanic and 50% were eligible for free or reduced-price meals. Our total sample was 80% White and 26% eligible for free or reduced-price meals.

Table 1

Number of Teachers in Each Sample with Sample Overlap

	CE sample	PCT sample	IM sample	AE sample
CE sample	566	448	405	140
PCT sample	118	672	564	181
IM sample	110	383	733	180
AE sample	-	-	144	217
Total: 844	AE sample	CE sample	PCT sample	Overlap of four samples: 88

Note. The main diagonal shows the sample sizes of the four samples for separate analyses. The upper-right panel above the main diagonal shows the number of cases in the overlap of two teaching-practice samples. The lower left panel below the main diagonal shows the number of cases in the overlap of three teaching-practice samples. CE means cognitive engagement, PCT means problem solving and critical thinking, IM means instructional monitoring, and AE means affective engagement in the content.

2.2 Measures

2.2.1 Teaching Effectiveness & Teacher-Student Relationships

Data are based on the student survey of teaching effectiveness in the NEE database, called the Teacher Effectiveness Student Survey (TESS; see https://neeadvantage.com). TESS is modular, meaning member districts can survey students only on the teaching practices of interest to the district. When a district selects a specific teaching practice, all items associated with that teaching practice (3 to 5 items) automatically populates an online student survey.

1. *Cognitive Engagement (CE).* Four items addressed students' perception of the degree to which a teacher used strategies that cognitively engaged students in the content (α = .89).

2. *Problem-Solving and Critical Thinking (PCT).* Four items addressed students' perception of the degree to which a teacher used strategies that promote PCT (α = .91).
3. *Affective Engagement in the Content (AE).* Five items addressed students' perception of the degree to which a teacher used strategies that affectively engaged students in the content (α = .92).
4. *Instructional Monitoring (IM).* Four items addressed students' perception of the degree to which a teacher monitored learning at the individual and whole-class level as the lesson was progressing, and adjusted teaching as needed (α = .96).

On the same survey, students responded to five items that asked about their perspective on their teacher-student relationship (α = .96). Asking for student perceptions is noteworthy because a recent meta-analysis found that most studies measured TSRs from the teachers' perspective (Kincade et al., 2020). See Table 2 for a list of the items.

Table 2

Items Comprising Teaching Practices and Teacher-student Relationship Variables

Variable	Items
Cognitive Engagement (CE)	• This teacher expects us to think a lot and concentrate in this class. • This teacher's lessons make us think deeply. • This teacher's lessons make us think the whole class time. • This teacher wants us to ask questions during lessons.
Problem-Solving and Critical Thinking (PCT)	• This teacher asks "how?" and "why?" questions to make us think more. • This teacher waits a while before letting us answer questions, so we have time to think. • This teacher makes us compare different ideas or things. • This teacher makes us use what we learn to come up with ways to solve problems.
Affective Engagement in the Content (AE)	• This teacher makes lessons interesting. • This teacher points out how this topic is important to our lives. • This teacher gives us choices in our classwork. • This teacher tells us that we can all be successful if we try hard. • This teacher gets us excited about the subject.

Instructional Monitoring (IM)	• This teacher checks often to make sure we understand the lesson as we go along. • This teacher explains the lesson in different ways if we don't get it at first. • This teacher knows when we understand the lesson. • This teacher has a way to make sure everyone is learning.
Teacher-student Relationships (TSRs)	• This teacher knows me and cares about me. • Students enjoy being with this teacher. • This teacher enjoys working with students. • This teacher is friendly. • Students can talk to this teacher if they have a problem.

TESS uses a 4-point Likert scale (0 = Not true; 1 = Sort of true; 2 = True; 3 = Very true). TESS has shown good factor structure and the subscales comprising each teaching practice have a Cronbach's Alpha ranging from .77 to .90 (Grajeda et al., 2017). Latent factor scores were generated for the five measures based on the mean item scores from student ratings of their teachers, and teachers were rated by at least 5 students (very few exceptions were rated by 4). In addition, multidimensionality of teaching was distinguishable by 4^{th} – 12^{th} grade students. That is, students might rate the same teacher as better at one teaching practice than another (Grajeda et al., 2017).

2.2.2 Other Variables

We included the following variables as covariates because they may affect teaching practices and teacher-student relationships.

Years of Experience. Teachers reported their years of teaching experience. More years of experience predicts better student behavior and achievement (Ladd & Sorensen, 2017; Nye et al., 2004). Furthermore, one meta-analysis found that the effect of TSRs on achievement was larger in studies with teachers who had more years of teaching experience (Roorda et al., 2011).

Core Subject Area. Students reported whether or not a teacher primarily taught a core subject (e.g., mathematics, English language arts, science or social studies). We tested whether results were similar between core and other courses. Cohen et al. (2018) found that teaching quality is not a uniform construct across varying subjects. Compared with teachers in core subjects, teachers in non-tested subjects may use different instructional practices and are often evaluated differently (Goe & Holdheide, 2011).

Grade Level. Teacher grade level was computed as a continuous variable based on the mean grade level identified by student responses. Previous research has demonstrated that teachers at varying grade levels had different levels of self-efficacy for implementing different teaching practices (Klassen & Chiu, 2010). Student engagement, teacher behavior, and teacher-student interactions are often different across subject areas and grade levels (den Brok et al., 2004; Marks, 2000).

Demographics. Student responses on the TESS were strictly anonymous. Thus, student-level demographic data were not available. School-level demographic data were obtained from the Missouri Department of Elementary & Secondary Education (DESE). This included the percentage of students receiving free or reduced-price lunch, total student enrollment, and the percentage of White students. School level demographics predict teacher job satisfaction, stress, and burnout, which may be associated with the quality of teaching practices used (Hamre et al., 2008; Johnson et al., 2012; Kyriacou, 2001; Olsen & Huang, 2019; Perry & McConney, 2010).

Achievement. Due to students' anonymity, student-level achievement data was not available. School-level data on the state proficiency test –Missouri Assessment Program (MAP) – was obtained from the state department of education. We retrieved the percentage of students at or above the proficiency level for mathematics and English language arts and computed the mean percentage as a school-level achievement variable. See http://dese.mo.gov for more details.

2.3 Procedures

TESS was delivered online during a window of time specified by the school. An access code unique to each teacher needed to be entered to ensure students evaluated the correct teacher and each access to the survey was authorized. A proctor other than the evaluated teacher administered the survey using standard administration scripts provided by NEE. The proctor read the instructions to the students, informed them of the purpose of the survey, ensured the anonymity of their responses, emphasized the voluntary nature of the survey, and its importance for school improvement. Students were encouraged to ask questions, but the proctors were instructed not to interpret any survey items to avoid influencing student responses.

Three screening items were distributed across the survey (e.g., "I am being totally honest on this survey"), which help improve survey validity and identify inattentive responses (Cornell et al., 2012). Surveys that failed two of the three screening items were flagged for manual review. Students who indicated in a fourth screening item that they were new to the class (less than a month) were also excluded. Data use was approved by the university's Institutional Review Board. All member districts consented to using their data for research purposes.

2.4 Statistical Analysis

We hypothesized that positive teacher-student relationships (TSRs) would promote higher-quality teaching practices (TPs). To test this, we conducted four regression models with TPs as the outcomes, and TSRs, grade level, and other variables as predictors, including an interaction term between TSRs and grade level. Teachers' TPs and TSRs scores in the models and following analyses all have controlled for prior scores (see the Appendix for more method details). However, it is also possible that the direction of the effect is reversed. That is, TPs may predict TSRs, or that both are due to an unspecified confounding variable. Since an experimental design is not an option for the current inquiry due to ethical concerns, we have to rely on archival data to examine the hypothesized direction of effect, and at the same time address the consequent methodological challenges, including an alternative direction of effect (i.e., TP → TSR), potential confounding variables, an illusory halo effect, and common method variance.

To examine the direction of the effect, we employed direction dependence analysis (DDA; Li & Wiedermann, 2020b; Wiedermann et al., 2020; Wiedermann & Li, 2018), an innovative causal model selection method, as a supplementary procedure to regression analysis. DDA is a framework that consists of various statistical tests that can generate different result patterns to differentiate competing models with different directions of effect (i.e., TP → TSR or TSR → TP) and to detect potential confounding (i.e., TP ← confounder → TSR). When none of the models are selected, the results are inconclusive, which may be due to the lack of sample size. This framework was proposed to bridge the recent advancements in data science and common modeling practices among educational researchers. Compared to other methods widely used by educational researchers to make causal claims with non-experimental data, DDA has minimal theoretical assumptions; instead, it primarily holds testable distributional assumptions and requires a large sample size to achieve desirable statistical power (Li & Wiedermann, 2020b; Wiedermann & Li, 2018). Specifically, for the current study, we applied Conditional Direction Dependence Analysis (CDDA; Li & Wiedermann, 2020b) as a supplementary procedure to examine the hypothesized direction of the effect. CDDA is an extension of standard DDA and is capable of examining the direction of effect when there is an interaction between the effect of interest. This extension can also achieve more statistical power and avoid some bias in moderation models. The DDA framework has been used effectively in prevention science (Musci & Stuart, 2019; Wiedermann et al., 2020), psychology (Wendt & Bartoli, 2019), public health

(Chew et al., 2020; Pérez-Mengual et al., 2021), and education (Sebastian et al., 2018; Wiedermann et al., 2020). We conducted direction dependence analysis using the SPSS add-ons that are publicly available at https://www.ddaproject.com and followed the guidelines for the analysis and interpretation (Li et al., 2022; Li & Wiedermann, 2020a, 2020b). To control for time-invariant confounding, teachers' TP and TSR ratings in the models and following analyses all have controlled for prior scores in advance, meaning the student ratings in the following analyses were residualized variables, and therefore, prior scores are not displayed in the results. Other potential confounding can be detected by DDA procedures (Wiedermann & Li, 2019; Wiedermann & Sebastian, 2019). See the Appendix for more details about the method and its implementation.

Any halo effect should be minimal as the aggregation of student ratings were used for teacher scores (Feeley, 2006; Feldman, 1986). Moreover, various types of common method variance should also be minimal as a result of strict student anonymity, the use of mean aggregation, and districts' changing of prioritized teaching practices across the years of implementation. See the Appendix for more details about how challenges regarding the halo effect and common method variance were addressed.

Though multilevel modeling (MLM) is a preferred method as teachers were nested within schools, we used ordinary least squares (OLS) regression because DDA does not apply to MLM. When a clustering effect is present, standard errors may be biased and erroneous conclusions may be made when using OLS instead of MLM (Guo & Zhao, 2000; Hox, 1998). Therefore, we used cluster robust standard errors for parameter testing, which is considered to be a valid alternative to MLM (Berger et al., 2017; Cameron et al., 2011; Huang, 2016; Huang & Li, in press; Kauermann & Carroll, 2001).

3. Results

Table 3 shows the directional conclusions across teaching practices and grade levels. The hypothesized TSR → TP models were supported across all grade levels in the *Cognitive Engagement in the Content* (CE), *Problem-Solving and Critical Thinking* (PCT), and *Instructional Monitoring* (IM) samples, except that a confounder effect was detected between TSR and IM scores in middle schools. However, contrary to our hypothesis, *TSRs* appeared to be an outcome, rather than antecedent, of teachers' *Affective Engagement (AE) in the Content* teaching practice in secondary schools.

Table 4 shows the regression results of the supported models. TSRs and grade level were statistically significant predictors of CE, PCT, and IM across all grade levels after controlling for the covariates. The local effect sizes (Cohen's f^2; Cohen, 2013) were small ($f^2_{TSR\to CE} = .037$, $f^2_{TSR\to PCT} = .093$) to medium ($f^2_{TSR\to IM} = .179$, $f^2_{AM\to TSR} = .316$) in the four supported models. There was a grade-level main effect on teachers' CE, PCT, and IM scores. After controlling for TSRs and other covariates, teachers in higher grades tended to receive better ratings in the three teaching practices from students. Grade level did not affect TSRs after controlling for teachers' *Affective Engagement in the Content* practices and other covariates. Furthermore, the interaction term was statistically significant for *Cognitive Engagement* and *Instructional Monitoring*. That is, for students in higher grades, TSRs had a greater effect on student ratings of teachers' CE and IM practices. There was no statistically significant interaction for *Problem-Solving and Critical Thinking* and *Affective Engagement in the Content*. Figure 2 provides the interaction plots illustrating the difference in effects among 4th, 7th, and 10th grades. Due to the smaller AE sample size, statistical power may be inadequate to detect the interaction (Shieh, 2008), but this should not be the case for PCT.

Table 3

Data Supported Models across Samples and Grade Levels

Grade Levels	Direction of Effect			
	CE Sample	PCT Sample	IM Sample	AE Sample
4^{th}	TSR→CE	TSR→PCT	TSR→IM	Inconclusive
5^{th}	TSR→CE	TSR→PCT	TSR→IM	Inconclusive
6^{th}	TSR→CE	TSR→PCT	TSR→IM	Inconclusive
7^{th}	TSR→CE	TSR→PCT	Confounder	AE→TSR
8^{th}	TSR→CE	TSR→PCT	Confounder	AE→TSR
9^{th}	TSR→CE	TSR→PCT	TSR→IM	AE→TSR
10^{th}	TSR→CE	TSR→PCT	TSR→IM	AE→TSR

Note. The grade levels represent the mean of student reported grade levels. Conditional Directional Dependence Analysis (CDDA; Li & Wiedermann, 2020b) was used to examine the direction of effect between TSRs and TPs at particular mean grade levels, and for the current analysis, Hilbert Schmidt Independence Criterion (HSIC; based on 200 bootstrap resamples) and the cube-based Higher Order Correlation Difference test (ΔHOC_c; based on 1000 bootstrap resamples and 95% bootstrap confidence interval) are used. See the Appendix for more details.

Table 4

Regression Results of the Supported Models

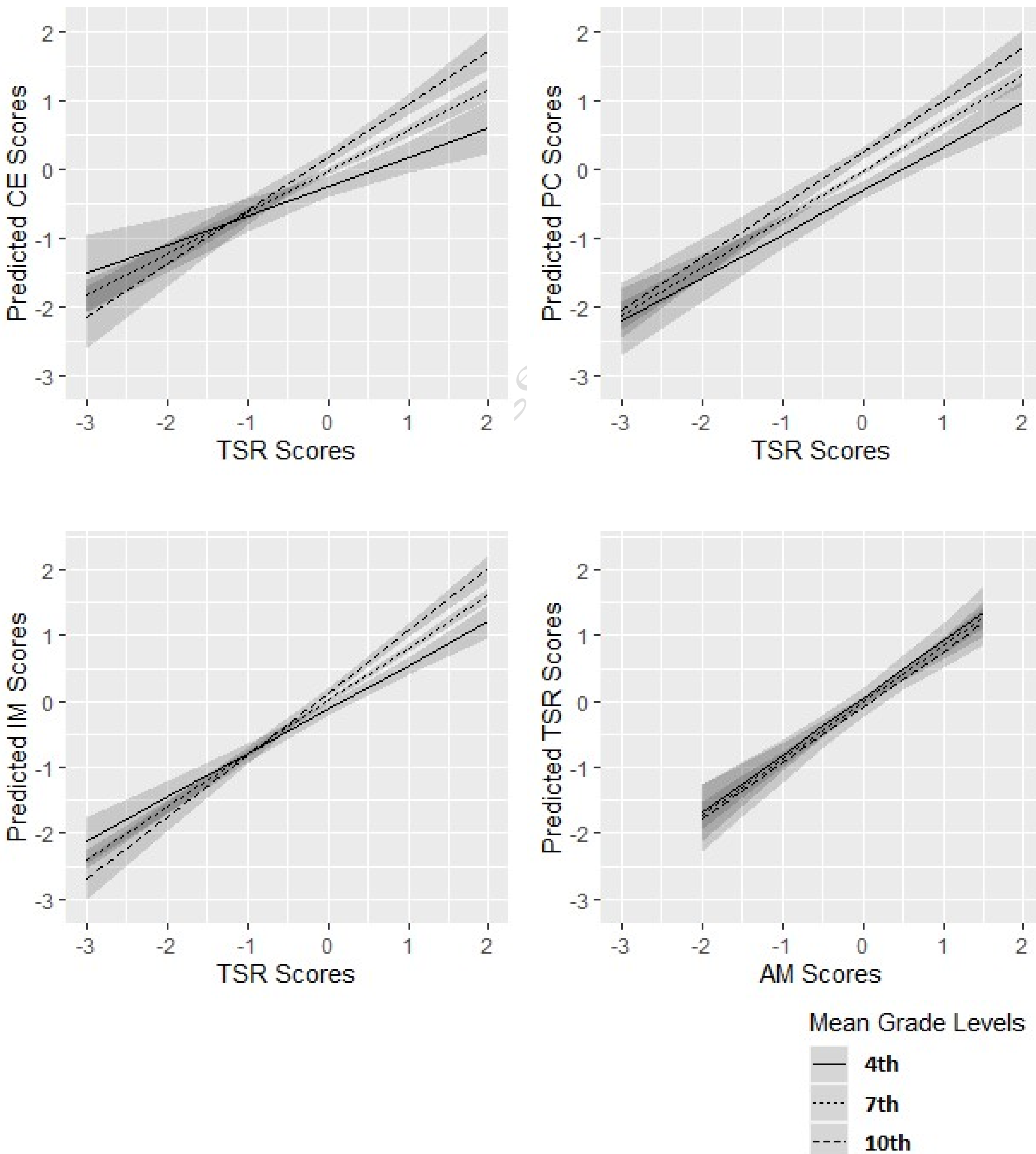
Predictors \ Outcomes	*B(SE)* Cognitive Engagement (R^2 = .384, n = 566)	Problem-solving and Critical Thinking (R^2 = .476, n = 672)	Instructional Monitoring (R^2 = .629, n = 733)	Teacher Student Relationships (R^2 = .625, n = 217)
Intercept	-.091(.367)	.530(.285)	.414(.267)	.436(.258)
Affective Engagement	–	–	–	.871(.071)***
TSR	.425(.086)***	.639(.082)***	.662(.058)***	–
Grade level	.069(.030)***	.092(.019)***	.043(.016)**	-.023(.021)
Teacher years of experience	.002(.003)	-.001(.002)	-.001(.003)	-.004(.004)
Free/Reduced lunch (%)	-.001(.002)	-.001(.002)	.002(.001)	.003(.003)
Enrollment (100ct)	-.002(.007)	-.010(.006)	-.008(.005)	.024(.009)*
White enrollment (%)	-.002(.004)	-.007(.002)***	-.005(.002)*	-.007(.003)*
MAP proficiency rate (%)	-.002(.006)	-.005(.004)	-.002(.004)	.001(.004)
Core subject	.203(.051)***	.160(.040)***	.033(.030)	.117(.047)*
TSR × Mean Grade levels	.058(.020)*	.022(.022)	.047(.016)**	–
AE × Mean Grade levels	–	–	–	-.005(.034)

Note. Both TSRs and TPs are residualized scores for 2018-2019 school year controlling both prior TSR and TP scores. Therefore, prior scores are not listed here as predictors. See Table A3 in the Appendix for the results of prior-score-only models. The parameter tests are based on adjusted cluster-robust standard errors (Cameron et al., 2011).

*** $p < .001$, ** $p < .01$, * $p < .05$

Figure 2

Predicted Outcome Scores Across Selected Mean Grade Levels



Note. Shading represents the 95% confidence intervals.

Core subject teachers, compared to teachers in other subject areas (e.g., physical and health education, fine arts, foreign languages), were rated higher in *Cognitive Engagement* and *Problem-Solving and Critical Thinking*, but similar in *Instructional Monitoring*. In the AE sample, core subject teachers tended to have better relationships with students, when controlling for other variables. Other control variables – teacher years of experience, school free or reduced lunch percentage, or school-level MAP proficiency rate – did not show any significant influence on student ratings of either TSR or teaching practices. Thus, these variables do not have a detectable effect within the scope of one school year, after controlling for the previous year's survey ratings, in this sample.

4. Discussion

Substantial research indicates that positive teacher-student relationships (TSRs) are linked to increased motivation and achievement in students. We hypothesize the same effect for teachers, although this has not been a focus of the research literature. Our hypothesis is supported by both Self-Determination Theory and Leader-Member Exchange Theory which predicts that adults increase effort and performance in work contexts where they have positive relationships. That is, positive TSRs should lead to teacher's greater use of high-impact teaching practices that are complex and effortful to implement. We explored our hypothesis using archival data from a state-wide authentic teacher growth and evaluation system, the Network for Educator Effectiveness (NEE), in Missouri, USA. Our analysis revealed three key results: (1) positive TSRs predict higher-quality instruction and this effect varies by (2) teaching practice and (3) grade level.

4.1 Positive Teacher-Student Relationships Lead to Higher-Quality Teaching

Our results, using student ratings, confirm our hypothesis that positive TSRs lead primary and secondary teachers to more effectively enact three complex teaching practices examined in this study – *Cognitive Engagement* (CE) *in the Content*, *Problem-Solving and Critical Thinking* (PCT), and *Instructional Monitoring* (IM). Our study supports previous studies that found classrooms with more positive TSRs, have teachers who are more likely to check-in, monitor, scaffold, and/or provide constructive feedback to students (Reddy & Dudek, 2014; van de Pol et al., 2010), have greater confidence in their students' abilities (Summers et al., 2017), and use better scaffolding strategies for critical thinking (Daws, 2005). We extended these studies by testing the direction of effect between positive TSRs and high-quality teaching practices. Furthermore, our results extend research using LMX theory (Gerstner & Day, 1997) from the field of organizational psychology to the field of education by demonstrating that positive relationships may lead to improved performance among adult workers.

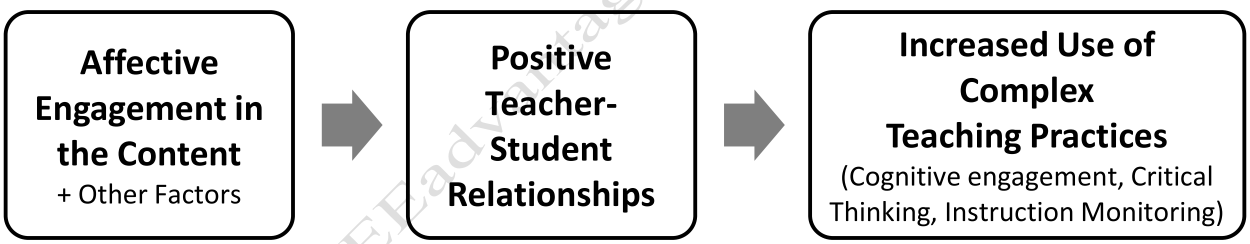
4.2 The Effect Varies by Teaching Practice

An exception to our results discussed above was use of teaching strategies to promote *Affective Engagement in the Content*. Contrary to our hypothesis, our results indicate that AE is more likely to lead to, rather than result from, positive TSRs in secondary schools. That is, when teachers put effort into making lessons interesting and enjoyable for students, and provide opportunities for autonomy in choosing learning tasks, secondary students view their teacher as more caring. This study supports previous literature indicating that making lessons interesting is perceived by students as an act of caring (Jeffrey et al., 2013; Wentzel, 1997, 2009). We extend these studies by using quantitative data to test the direction of the effect.

Our results support the logic model in Figure 3 for secondary grades. That is, teachers' use of strategies that affectively engage students in the content leads to positive TSRs. Recall that *Affective Engagement in the Content* refers to positive emotions toward the content or lesson activities, not toward the teacher. Other factors may be involved in promoting positive TSRs as well, such as being sensitive and warm, using non-coercive discipline, and behaving prosocially toward students (Bergin & Bergin, 2009; Prewett et al., 2019). When positive TSRs are established, they may motivate teachers to engage in more complex teaching practices (CE, IM, PCT) with greater frequency and higher quality.

Figure 3

Association between Teaching Practices and Teacher-Student Relationships



Given that our hypothesized TSR → AE model was rejected, we are more confident that a halo effect was not strongly present. That is, the mean aggregated student ratings of the other teaching practices were not a biased result of the halo effect from student relationships with their teachers, after prior teachers' scores were controlled. Our reasoning is that neither a consistent confounding effect nor a consistent driving force of all four teaching practices were identified. That is, neither the *general impression halo model* nor the *salient dimension halo model* was supported. See the Appendix for more details.

4.3 The Effect Varies by Grade Level

4.3.1. Main Effect

There was a statistically significant positive main effect for grade level on teaching practices while controlling for other covariates. That is, older students reported teachers more frequently used the effortful, complex, high-impact teaching practices we studied. One possible explanation is that it is easier for teachers to use *Cognitive Engagement*, *Problem-Solving and Critical Thinking*, and *Instruction Monitoring* strategies with older students because of their increased cognitive abilities. In adolescence, information processing speed, complex working memory, executive functioning (Coyle et al., 2011; Kail & Ferrer, 2007; Luciana et al., 2005; Steinberg, 2009), use of memory strategies, and vivid recall (Brainerd et al., 2004) increase. These cognitive developments allow adolescents to reason abstractly, incorporate new information, and monitor learning progress faster and easier. In addition, as course content becomes more challenging in higher grades, teachers may tend to use these complex strategies more often. Furthermore, teacher certification requires teachers in secondary grades to have deeper content expertise. As content specialists, rather than generalists, secondary teachers may engage in more complex teaching practices regarding their subject of expertise. However, there is mixed evidence on whether teacher certifications are actually associated with student outcomes (Goldhaber & Brewer, 2000; Kusumawardhani, 2017). These explanations are speculative and need to be confirmed by further research.

4.3.2 Interaction Effect

There was also a statistically significant interaction effect between TSRs and grade level for CE and IM (see Figure 2). As we mentioned above, positive teacher-student relationships tended to decrease steadily from 1st grade to the end of middle school, before stabilizing in high school (Gillen-O'Neel & Fuligni, 2012; Hughes et al., 2012). Because TSRs are lowest for teachers in higher grades, the effect of positive relationships on teaching practices may be stronger for teachers who manage to have positive relationships despite the developmental trend. Supporting this logic, a meta-analysis found a stronger effect between TSRs and achievement in secondary schools compared to primary schools, with student engagement as a mediator (Roorda et al., 2017).

In contrast, there was not a statistically significant interaction effect between TSRs and grade level for PCT and AE. This suggests that TSRs have an equal effect on *Problem-Solving and Critical Thinking* across all grade levels. A possible explanation is that PCT is a challenging and rare teaching practice (Willingham, 2008), therefore, teachers may not use it frequently during class even though they have good relationships with the students. As for *Affective Engagement in the Content*, the effect may simply be undetectable due to the small sample size.

4.4 Implications for Practice

Our results suggest that when teacher-student relationships (TSRs) are positive, teachers are more likely to engage in effortful, complex, high-impact teaching practices that are associated with student learning. Our results suggest that when schools aim to improve quality of teaching practices, they would do well to attend to improving the quality of TSRs first. One way to improve TSRs may be to use teaching strategies that affectively engage students in the content. Effects are likely to occur for all students but may be strongest for secondary students.

4.5 Limitations

A major limitation of the current study is the small size of the AE sample for primary grades, which reduced the interpretability of results. We cannot rule out a confounding variable in the relationship between *Affective Engagement in the Content* and TSRs in lower grade levels. Another limitation is that, since our archival data were collected from an authentic teacher evaluation system where student-level information was strictly anonymous, student-level measurement error cannot be fully taken into consideration, although it was attenuated by using mean aggregation (Dunn et al., 2015; Richter & Brorsen, 2006). Further, schools in the samples have somewhat lower math proficiency rates than the state average, which may be a result of selection bias into the teacher evaluation system. Therefore, the samples of the study are representative of schools with somewhat lower student performance in math

Appendix

Supplementary Statistical Analysis and Results

We hypothesized that positive teacher-student relationships would predict higher-quality teaching practices. To test this, we conducted four regression models. The outcome variables were high-quality teacher practices (*Cognitive Engagement*, *Problem-Solving and Critical Thinking*, *Instruction Monitoring*, and *Affective Engagement in the Content*). The predictor variables included TSR, teacher-level variables (taught a core subject, grade level, years of experience), and school-level variables (total enrollment, White enrollment, free lunch, MAP proficiency rate). In addition, an interaction between TSRs and grade level was included.

Contrary to our hypothesis, it is also possible that the direction of the effect is reversed. That is, Teaching practices (TPs) may predict TSRs, or that both are due to an unspecified confounding variable. Since an experimental design is not an option for the current inquiry due to ethical concerns, we have to rely on observational data to examine the hypothesized direction of effect, and at the same time address the consequential methodological challenges, including an alternative direction of effect (i.e., TPs → TSR), potential confounder, illusory halo, and common method variance.

Alternative Direction of Effect

Establishing cause-effect relationships between variables is one of the focal interests of educational researchers. Though conventionally causal direction or the direction of effect is examined with experimental or quasi-experimental studies in the educational context, we used an innovative method, Direction Dependence Analysis (DDA; Li & Wiedermann, 2020b; Wiedermann & Li, 2018; Wiedermann & von Eye, 2015) to validate the hypothesized models (TSR → TP) against models with the alternative direction of effect (TP → TSR).

In the current study, we investigated whether our hypothesized direction of effect – that positive TSRs cause higher-quality teaching practices – was supported across grade levels. That is, CDDA was used to validate the hypothesized explanatory model (i.e., TSR → TP) against the plausible alternative model (i.e., TP → TSR), with a reversed flow of causality between the two variables across grade levels.

Potential Confounder

Although we have included variables that may affect both TSRs and TP, to minimize potential confounding, we also adjusted both the TP and TSR scores for the 2018-2019 school year using the prior scores. Specifically, as CDDA requires the same set of covariates in the competing models (Li et al., 2020; Li & Wiedermann, 2020b), we use both prior scores to obtain the adjusted scores (i.e. $\boldsymbol{e_{TP}}$ and $\boldsymbol{e_{TSR}}$) shown in Eq. (A.1) and Eq. (A.2),

$$\begin{cases} TP_{2018-2019} = b_1 TP_{2017-2018} + b_2 TSR_{2017-2018} + \boldsymbol{e_{TP}} & Eq. (A.1) \\ TSR_{2018-2019} = b_3 TP_{2017-2018} + b_4 TSR_{2017-2018} + \boldsymbol{e_{TSR}} & Eq. (A.2) \end{cases}$$

where the subscripts of TP and TSR indicate the school year; b_1, b_2, b_3 and b_4 are the coefficients (b_2 *and* b_3 are not statistically significant); and $\boldsymbol{e_{TP}}$ and $\boldsymbol{e_{TSR}}$ are the residuals representing the teacher-level variance unexplained by the prior scores, and are the focal variables we used for the analysis. Using the residualized variables, the models we build are to examine the synchronous effect between TSR and TP controlling their prior scores and cross-lagged effects (Li et al., 2020). That is to say, the scope of the analysis will be narrowed within the 2018-2019 school year and potential time-invariant confounders can be greatly reduced (VanderWeele, 2019). Besides the efforts to minimize potential confounding, we also used the DDA framework to test confounding effects (Wiedermann & Li, 2019; Wiedermann & Sebastian, 2019).

Assumption Check and Implementation

Table A1

The List of Outliers Excluded from Conditional Direction Dependence Analysis

Outliers	Mahalanobis' Distance			DFBETA			Mean Grade Level | Years of Experience | Core subject (T/F)		
	CE Sample	PCT Sample	IM Sample	CE Sample	PCT Sample	IM Sample	CE Sample	PCT Sample	IM Sample
#1	36.87	61.84	48.20	-0.000	.000	.000	8 |14| F	5| 22| T	5 |22| T
#2	33.17	61.15	47.53	-0.002	-.002	.001	5 | 0 | F	6 | 2 | T	6 | 2 | T
#3	29.59	60.74	46.57	0.002	.000	.000	6 | 9 | F	4 | 1 | F	4 | 1 | F
#4	27.68	58.46	44.45	0.000	.000	.001	6 |14| F	6 | 7 | F	6 | 3 | T
#5		58.30	44.24		.000	.000		4 | 4 | T	5 | 0 | T
#6		58.23	44.20		.001	.000		6 | 3 | T	4 | 4 | T
#7		57.97	44.09		.000	.000		4 | 9 | T	6 | 7 | F
#8		57.06	44.05		.000	.000		5 | 5 | T	4 | 9 | T
#9		28.90	43.06		-.001	.000		5 | 0 | F	5 | 5 | T
#10		26.76	30.28		.001	-.001		4 | 3 | F	4 | 0 | F
#11		26.71			.001			4 |15| F	

Note. Mahalanobis' Distance is used to detect multivariate outliers ($\chi^2 \geq 26.13, df = 8, p < .001$), and no outliers were identified in the AE sample. DFBETA indicates the influence of individual cases on the main effect of Teacher-student relationships on respective teaching practices.

Following Wiedermann and Li (2018), the distribution assumptions of the adjusted focal variables were examined, and multivariate outliers were excluded from CDDA but not from the multiple regression analysis, as these cases may bias the directional conclusion but are not influential on the main effect (see Table A1 for a list of the excluded cases). Two CDDA tests were used, including the Hilbert Schmidt Independence Criterion (HSIC; Gretton et al., 2005) and the cube-based Higher-Order Correlation Difference test (ΔHOC_c; Wiedermann & Li, 2018) after we built the conventional multiple regression models and their alternative models with reversed causal directions. That is, our focal variables, adjusted TSRs and TP, served in turns as the dependent and the independent variables in pairs of competing models across each sample.

HSIC was used to examine residual-predictor independence in the regression models. A HSIC value not significantly different from zero indicates the model is more likely to have the "true" direction of effect when its directionally competing model shows a non-zero HSIC value. When HSIC values in both competing models are significantly larger than zero (i.e., a predictor-residual dependence was observed in both models) an unconsidered confounder may exist (Wiedermann & Li, 2020). The ΔHOC_c test was used to examine the distributional properties of our hypothesized dependent and independent variables after controlling other covariates. A positive value indicates a model is more likely to be directionally "true" when its competing model is not shown positive. Both tests are based on bootstrap procedures.

The Halo Effect

Halo effect (Thorndike, 1920), or illusory halo (Cooper, 1981), is a pervasive challenge when teachers are evaluated by students (Wolfe & Song, 2015). It represents the added constant (positive or negative) to all of a given rater's true scores (Feeley, 2006; Feldman, 1986). The halo effect is likely to be minimal in the current study because it can be reduced by a factor equal to the number of raters per teacher when mean aggregation is used and when raters have adequate experience with the ratees (Feeley, 2006; Feldman, 1986). However, since CDDA is capable of evaluating causal models, to probe the potential role of any remaining halo, we examined how our data may fit the three causal models (Fisicaro & Lance, 1990; Solomonson & Lance, 1997) that are often used to explain halo effect: 1) the *general impression model*, which suggests that a separate general impression affects the observed scores; 2) the *salient dimension model*, which indicates that one salient trait influences the evaluations of other traits; and 3) the *inadequate discrimination model*, which attributes the halo to the raters' failure to discriminate among conceptually distinct aspects of a ratee's behavior. Previous analysis of TESS indicated that both upper primary and secondary students do distinguish the dimensionality of teacher effectiveness (Grajeda et al., 2017), so the first and the second models are our models of interest. A confounding effect is detectible with CDDA when the *general impression model* better explains the halo, and the halo effect is big enough to bias our model estimates. Otherwise, the *salient dimension model* may be a better model to explain the data when TSRs is shown to be a driving force for all four teaching practices, though it is equally possible that the TSR→TPs hypothesis may reflect the true direction of effects.

Common Method Variance (CMV)

As both the dependence and independent variables were from the same survey by the same group of students through the same platform, common method variance (CMV) may exist and bias the variance between the constructs of TSRs and TP. CMV is the spurious variance between dependent and independent variables due to shared method and/or source (Podsakoff et al., 2003). Since we have controlled teachers' prior scores measured with the same items through the same platform, CMV due to item characteristic effects and measurement context effects were controlled. In addition, item context effects were also addressed as different districts prioritized different teaching practices over years, and therefore, the context of the item was not the same across districts and years. Moreover, mean aggregation of student ratings and strict student anonymity also helped minimize common rater effects and other measurement errors (Feeley, 2002; Podsakoff et al., 2003; Richter & Brorsen, 2006).

Results

Before addressing our research questions, we investigated the association between variables without regard to causal direction by examining the eight regression models. That is, four multiple regression models as our hypothesized models with the interaction terms between grade level and TSRs (i.e., TSR | Grade Level → TP) and four corresponding alternative models with the interaction terms between grade level and TPs(i.e., TSR | Grade Level → TP). There was a statistically significant relationship between TSRs and teaching practices in all of the eight models. Therefore, before interpreting the results of the multiple regression analyses, it is necessary to select the models with the direction of effect better supported by the data. The CDDA results are displayed in Table A2.

Cognitive Engagement (CE)

To interpret Table 4 and Table 5, look at the column for CE first. For all grades $p_{HSIC}^{(TSR \to CE)} > .05$ and $p_{HSIC}^{(CE \to TSR)} < .05$. This means that the hypothesized model (TSR → CE) is more likely to be the "true" model. Moreover ΔHOC_c supports the same conclusion for all grade levels as $\Delta HOC_c^{(TSR \to CE)} > 0$ (the zero point is below the 95% bootstrap confidence interval (BCI) in the target model), and $\Delta HOC_c^{(CE \to TSR)} \ngtr 0$ (the zero point is within or above the 95% BCI).

Problem-Solving and Critical Thinking (PCT)

Similarly, the hypothesized TSR → PCT model was also supported by both HSIC and ΔHOC_c tests across grade levels, except for 5th and 6th grades. As HSIC remains powerful across all the grades for the PCT sample, it is less likely that it consists of cases with different directions of effect (Li & Wiedermann, 2020b). That is to say, *Problem-Solving and Critical Thinking* is

Table A2

Conditional Direction Dependence Analysis Results across Grades

Grade Levels	TSR → CE HSIC	TSR → CE ΔHOC$_c$	TSR → PCT HSIC	TSR → PCT ΔHOC$_c$	TSR → IM HSIC	TSR → IM ΔHOC$_c$	TSR → AE HSIC	TSR → AE ΔHOC$_c$
Hypothesized Models								
4th	.033	.51~17.24	.003	3.87~14.72	.004	4.70~16.68	.002	2.61~23.49
5th	.004	3.23~28.15	.005	7.54~23.49	.008	8.10~22.87	.005	3.78~34.61
6th	.007	7.73~31.70	.009	11.26~27.65	.022	7.39~21.90	.012	-7.11~7.85
7th	.016	2.22~ 9.18	.016	1.77~7.91	.075**	-.88~3.10	.073***	-.82~ 3.77
8th	.050	1.16~7.76	.037	1.71~7.25	.046*	-.00~7.96	.027*	.74~9.28
9th	.033	1.68~12.72	.034	2.81~11.13	.017	1.63~16.77	.009*	-1.3~5.84
10th	.015	.92~ 6.58	.014	1.48~6.54	.008	.83~13.00	.004*	-.99~ 2.71

Grade Levels	CE→ TSR HSIC	CE→ TSR ΔHOC$_c$	PCT→ TSR HSIC	PCT→ TSR ΔHOC$_c$	IM→ TSR HSIC	IM→ TSR ΔHOC$_c$	AE→ TSR HSIC	AE→ TSR ΔHOC$_c$
Alternative Models								
4th	.074**	-.83~ 7.24	.029***	-.52~13.40	.012*	5.89~38.08	.001	3.69~38.45
5th	.027***	-.50~9.21	.049***	.90~18.05	.025***	8.55~51.97	.002	7.32~55.53
6th	.058**	-.18~1.36	.084***	2.43~17.77	.057**	7.54~39.35	.004	3.14~26.73
7th	.149***	-3.06~ 1.88	.170***	-2.28~1.88	.121***	-.88~3.10	.011	-2.04~ 1.98
8th	.174***	-2.38~1.18	.137***	-2.24~.87	.112***	-.11~5.24	.006	-.76~5.15
9th	.074**	-1.31~5.75	.049*	-1.71~1.65	.045***	3.00~13.83	.002	-1.28~3.89
10th	.032***	-1.63~ 3.89	.021*	-2.29~1.25	.015*	2.78~12.25	.001	-1.05~ 2.48

Note. Both focal variables in each model, are for 2018-2019 school year while adjusting the prior year scores; Hilbert Schmidt Independence Criterion (HSIC; based on 200 bootstrap resamples) and the cube-based Higher Order Correlation Difference test (ΔHOC; based on 1000 bootstrap resamples and 95% bootstrap confidence interval) are used in CDDA (Li & Wiedermann, 2020b)

*** $p < .001$, ** $p < .01$, * $p < .05$

less likely to be a mixture sample with some grades supporting the TSR→PCT model and the others supporting the PCT→TSR model. In addition, the interaction term in neither the hypothesized model nor the alternative model is significant. As neither scenario of moderation was the case, we can exclude the possibility of grade-level being a moderator. Therefore, as ΔHOC_c seemed not powerful enough to identify the direction of effect for 5th and 6th grades, i.e., the test returned positive ranges for both competing models, it makes more sense to use standard Direction Dependence Analysis instead. The results support our hypothesized model, $HSIC^{(TSR \to PCT)} = .022,\ p_{HSIC}^{(TSR \to PCT)} > .05;\ HSIC^{(PCT \to TSR)} = .022,\ p_{HSIC}^{(TSR \to PCT)} < .001;\ \Delta HOC_C^{(TSR \to PCT)} = 1.691, 95\%\text{BCI}\ [.29, 2.86]$.

Instructional Monitoring (IM)

The hypothesized TSR → IM model was supported by HSIC across grades except for the 7th, and 8th grades, where both $p_{HSIC}^{(TSR \to IM)}$ and $p_{HSIC}^{(IM \to TSR)}$ were smaller than .05. Therefore, an unconsidered confounder may exist between the two variables in middle schools. ΔHOC_c does not differentiate the competing models for the IM sample, as shown in Table 4 and Table 5, because ΔHOC_c results would be biased when a confounder exists (Li & Wiedermann, 2020b), as indicated by HSIC results here.

Affective Engagement in the Content (AE).

Finally, the hypothesized TSR → AE model was not supported as $p_{HSIC}^{(TSR \to AM)} <$ .05 and $p_{HSIC}^{(AM \to TSR)} >$ .05 for 7th to 10th grades but the *p*-value of both models was greater than .05 for lower grades. This means that for teachers at or above 7th grade, the alternative model was supported instead. That is, teachers' use of high-quality practices to affectively engage students may result in better TSRs in secondary schools but not in upper primary schools. The inclusive results were probably due to the smaller sample size. It is less likely that the hypothesized

TSR → AE model would be supported for teachers in lower grade levels (41.5% of cases) because HSIC is capable of identifying mixture samples consisting of two subsamples of similar sizes but reversed causal direction such that some cases in a sample support the X→Y model while the others support the Y→X model (Li & Wiedermann, 2020b). As the AE sample is smaller than suggested by Wiedermann and Li (2020) to detect a potential confounder, we cannot rule out the possibility that there was an unconsidered variable associated with both TSRs and AE for lower grade teachers. In addition, the smaller sample size contributes to unreliable ΔHOC_c test results for CDDA, because it does not have consistent performance when the sample size is smaller than 400 (Li & Wiedermann, 2020b). Similar to the PCT sample, as grade levels may not serve as a moderator, we also used standard DDA as a supplementary analysis for the AE sample. The alternative model was also favored by both tests, $HSIC^{(TSR \to AM)} = .031$, $p_{HSIC}^{(TSR \to AM)} < .05$; $HSIC^{(AM \to TSR)} = .010$, $p_{HSIC}^{(TSR \to AM)} > .05$; $\Delta HOC_C^{(TSR \to AM)} = -.463, 95\%\text{BCI}\ [-2.59, -.22]$. Therefore, we can conclude with caution that the alternative hypothesis is more likely to be "true" for *Affective Engagement* across secondary grade levels but not necessarily for lower grade levels.

Since the prior scores were partitioned out before we build models for direction dependence analysis and cluster-robust regression analysis, we examined the prior-score only models that are supported by direction dependence analysis. The results are shown in Table A3. All prior scores are significant predictors

Table A3

Results of Prior-Score Models Supported by CDDA

Predictors \ Outcomes	*B(SE)*			
	Cognitive Engagement (R^2 = .433, n = 566)	Problem-solving and Critical Thinking (R^2 = .399, n = 672)	Instructional Monitoring (R^2 = .461, n = 733)	Teacher Student Relationships (R^2 = .610, n = 217)
Intercept	.014(.029)	.065(.027)*	.049(.024)*	.019 (.039)
Prior Scores	.664(.032)***	.606(.029)***	.633 (.025)***	.702 (.038)***

Note. Outcome variables in the models are the scores for the 2018-2019 school year. Prior Scores in the models are those for the 2017-2018 school year.

*** $p < .001$, ** $p < .01$, * $p < .05$

References

Admiraal, W., Veldman, I., Mainhard, T., & van Tartwijk, J. (2019). A typology of veteran teachers' job satisfaction: Their relationships with their students and the nature of their work. *Social Psychology of Education, 22*, 337-355. https://doi.org/10.1007/s11218-018-09477-z

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science, 333*(6045), 1034-1037.

Archambault, I., Vandenbossche-Makombo, J., & Fraser, S. L. (2017). Students' oppositional behaviors and engagement in school: The differential role of the student-teacher relationship. *Journal of Child and Family Studies, 26*, 1702-1712.

Baker, J. A. (2006). Contributions of teacher-child relationships to positive school adjustment during elementary school. *Journal of School Psychology, 44*(3), 211-229. https://doi.org/10.1016/j.jsp.2006.02.002

Berger, S., Graham, N., & Zeileis, A. (2017). *Various versatile variances: An object-oriented implementation of clustered covariances in R*. Retrieved from https://cran.r-project.org/web/packages/sandwich/vignettes/sandwich-CL.pdf

Bergin, C., & Bergin, D. (2009). Attachment in the classroom. *Educational Psychology Review, 21,* 141-170. https://doi.org/10.1007/s10648-009-9104-0

Bergin, D. A. (1999). Influences on classroom interest. *Educational Psychologist, 34*(2), 87-98. doi:10.1207/s15326985ep3402_2

Brainerd, C. J., & Holliday, R. E. (2004). Behavioral measurement of remembering phenomenologies: So simple a child can do it. *Child Development, 75*(2), 505-522. https://doi.org/10.1111/j.1467-8624.2004.00689.x

Burić, I., & Kim, L. E. (2020). Teacher self-efficacy, instructional quality, and student motivational beliefs: An analysis using multilevel structural equation modeling. *Learning and Instruction*, *66*, 101302. https://doi.org/https://doi.org/10.1016/j.learninstruc.2019.101302

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics, 29*(2), 238-249. https://doi.org/10.1198/jbes.2010.07136

Chew, Q. H., Chia, F. L.-A., Ng, W. K., Lee, W. C. I., Tan, P. L. L., Wong, C. S., Puah, S. H., Shelat, V. G., Seah, E.-J. D., Huey, C. W. T., Phua, E. J., & Sim, K. (2020). Perceived stress, stigma, traumatic stress levels and coping responses amongst residents in training across multiple specialties during covid-19 pandemic-a longitudinal study. *International Journal of Environmental Research and Public Health*, *17*(18), 6572. https://doi.org/10.3390/ijerph17186572

Chien, Y.-T., Chang, Y.-H., & Chang, C.-Y. (2016). Do we click in the right way? A meta-analytic review of clicker-integrated instruction. *Educational Research Review, 17*, 1-18. https://doi.org/10.1016/j.edurev.2015.10.003

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences* (2nd ed.). Academic Press.

Cohen, J., Ruzek, E., & Sandilos, L. (2018). Does teaching quality cross subjects? Exploring consistency in elementary teacher practice across subjects. *AERA Open, 4*(3), 1-16. https://doi.org/10.1177%2F2332858418794492

Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin*, *90*(2), 218.

Cornelius-White, J. (2007). Learner-centered teacher-student relationships are effective: A meta-analysis. *Review of Educational Research, 77*(1), 113-143. https://doi.org/10.3102%2F003465430298563

Cornell, D., Klein, J., Konold, T., & Huang, F. (2012). Effects of validity screening items on adolescent survey data. *Psychological Assessment, 24*(1), 21.

Council of Chief State School Officers. (2011). InTASC model core teaching standards: A resource for state dialogue. *Washington, DC: Council of Chief State School Officers*.

Coyle, T. R., Pillow, D. R., Snyder, A. C., & Kochunov, P. (2011). Processing speed mediates the development of general intelligence (g) in adolescence. *Psychological Science, 22*(10), 1265-1269. https://doi.org/10.1177%2F0956797611418243

Curby, T. W., Rimm-Kaufman, S. E., & Ponitz, C. C. (2009). Teacher-child interactions and children's achievement trajectories across kindergarten and first grade. *Journal of Educational Psychology, 101*(4), 912-925. https://psycnet.apa.org/doi/10.1037/a0016647

Daws, J. E. (2005). Teachers and students as co-learners: Possibilities and problems. *Journal of Educational Enquiry, 6*(1), 110-125.

Deci, E. L., & Ryan, R. M. (2010). *Self-determination.* John Wiley & Sons.

den Brok, P., Brekelmans, M., & Wubbles, T. (2004). Interpersonal teacher behaviour and student outcomes. *School Effectiveness and School Improvement, 15*(3-4), 407-442. https://doi.org/10.1080/09243450512331383262

Dunn, E. C., Masyn, K. E., Jones, S. M., Subramanian, S. V., & Koenen, K. C. (2015). Measuring psychosocial environments using individual responses: An application of multilevel factor analysis to examining students in schools. *Prevention Science, 16*(5), 718-733. https://doi.org/10.1007/s11121-014-0523-x

Feeley, T. H. (2002). Comment on halo effects in rating and evaluation research. *Human Communication Research, 28*(4), 578-586. https://psycnet.apa.org/doi/10.1111/j.1468-2958.2002.tb00825.x

Feldman, J. M. (1986). A note on the statistical correction of halo error. *Journal of Applied Psychology*, *71*(1), 173.

Fisicaro, S. A., & Lance, C. E. (1990). Implications of three causal models for the measurement of halo error. *Applied Psychological Measurement*, *14*(4), 419-429.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59-109. https://doi.org/10.3102%2F00346543074001059

Gehlbach, H., Brinkworth, M. E., King, A. M., Hsu, L. M., McIntyre, J., & Rogers, T. (2016). Creating birds of similar feathers: Leveraging similarity to improve teacher-student relationships and academic achievement. *Journal of Educational Psychology, 108*(3), 342-352. https://psycnet.apa.org/doi/10.1037/edu0000042

Gerstner, C. R., & Day, D. V. (1997). Meta-analytic review of leader-member exchange theory: Correlates and construct issues. *Journal of Applied Psychology, 82*(6), 827-844. https://psycnet.apa.org/doi/10.1037/0021-9010.82.6.827

Gillen-O'Neel, C., & Fuligni, A. (2012). A longitudinal study of school belonging and academic motivation across high school. *Child Development, 84*(2), 678-692. https://doi.org/10.1111/j.1467-8624.2012.01862.x

Goe, L., & Holdheide, L. (2011). Measuring teachers' contributions to student learning growth for nontested grades and subjects. Research & Policy Brief. *National Comprehensive Center for Teacher Quality*.

Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis, 22*(2), 129-145. http://www.jstor.com/stable/1164392

Grajeda, S., Bergin, C., Jones, E., & Tsai, C. (2017). *Reliability and factor structure of a student survey of teacher effectiveness.* Paper presented at the American Educational Research Association, San Antonio.

Graves, L. M., & Luciano, M. M. (2013). Self-determination at work: Understanding the role of leader-member exchange. *Motivation and Emotion, 37,* 518-536. https://doi.org/10.1007/s11031-012-9336-z

Gretton, A., Bousquet, O., Smola, A., & Schölkopf, B. (2005, October). Measuring statistical dependence with Hilbert-Schmidt norms. In *International conference on algorithmic learning theory* (pp. 63-77). Springer.

Guo, G., & Zhao, H. (2000). Multilevel modeling for binary data. *Annual Review of Sociology, 26*(1), 441-462. https://doi.org/10.1146/annurev.soc.26.1.441

Hamre, B. K., Pianta, R. C., Downer, J. T., & Mashburn, A. J. (2008). Teachers' perceptions of conflict with young students: Looking beyond problem behaviors. *Social Development, 17*(1), 115-136. doi:10.1111/j.1467-9507.2007.00418.x

Hattie, J. (2009). The black box of tertiary assessment: An impending revolution. In L. H. Meyer, S. Davidson, H. Anderson, R. Fletcher, P.M. Johnston, & M. Rees (Eds.). *Tertiary assessment & higher education student outcomes: Policy, practice & research* (pp. 259-275). Ako Aotearoa.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*(2), 111-127. doi:10.1207/s15326985ep4102_4

Holzberger, D., Philipp, A., & Kunter, M. (2013). How teachers' self-efficacy is related to instrucitonal quality: A longitudinal analysis. *Journal of Educational Psychology, 105*(3), 774-786. http://dx.doi.org/10.1037/a0032198

Hox, J. (1998). *Multilevel modeling: When and why.* Paper presented at the Classification, Data Analysis, and Data Highways, Berlin, Heidelberg.

Huang, F. L. (2016). Alternatives to multilevel modeling for the analysis of clustered data. *The Journal of Experimental Education, 84*(1), 175-196. https://doi.org/10.1080/00220973.2014.952397

Huang, F. L., & Li, X. (in press). Using cluster robust standard errors when analyzing cluster randomized trials with few clusters. *Behavior Research Methods*.

Hughes, J. N. (2011). Longitudinal effects of teacher and student perceptions of teacher-student relationship qualities on academic adjustment. *The Elementary School Journal, 112*(1), 38-60. https://doi.org/10.1086/660686

Hughes, J. N., Luo, W., Kwok, O.-M., & Loyd, L. K. (2008). Teacher-student support, effortful engagement, and achievement: A 3-year longitudinal study. *Journal of Educational Psychology, 100*(1), 1-14. https://psycnet.apa.org/doi/10.1037/0022-0663.100.1.1

Hughes, J. N., Wu, J.-Y., Kwok, O.-m., Villarreal, V., & Johnson, A. Y. (2012). Indirect effects of child reports of teacher-student relationship on achievement. *Journal of Educational Psychology, 104*(2), 350-365. https://psycnet.apa.org/doi/10.1037/a0026339

Jang, H., Reeve, J., & Deci, E. L. (2010). Engaging students in learning activities: It is not autonomy support or structure but autonomy support and structure. *Journal of Educational Psychology, 102*(3), 588-600. https://psycnet.apa.org/doi/10.1037/a0019682

Jeffrey, A. J., Auger, R. W., & Pepperell, J. L. (2013). “If we're ever in trouble they're always there”: A qualitative study of teacher-student caring. *The Elementary School Journal*, *114*(1), 100-117. https://doi.org/10.1086/671062

Jennings, P. A., & Greenberg, M. T. (2009). The prosocial classroom: Teacher social and emotional competence in relation to student and classroom outcomes. *Review of Educational Research, 79*(1), 491-525. https://doi.org/10.3102%2F0034654308325693

Jia, Y., Way, N., Ling, G., Yoshikawa, H., Chen, X., Hughes, D., Ke, X., & Lu, Z. (2009). The influence of student perceptions of school climate on socioemotional and academic adjustment: A comparison of Chinese and American adolescents. *Child Development, 80*(5), 1514-1530. https://doi.org/10.1111/j.1467-8624.2009.01348.x

Johnson, S. M., Kraft, M. A., & Papay, J. P. (2012). How context matters in high-needs schools: The effects of teachers’ working conditions on their professional satisfaction and their students’ achievement. *Teachers College Record, 114*(10), 1-39.

Kail, R. V., & Ferrer, E. (2007). Processing speed in childhood and adolescence: Longitudinal models for examining devlopmental change. *Child Development, 78*(6), 1760-1770. https://doi.org/10.1111/j.1467-8624.2007.01088.x

Kauermann, G., & Carroll, R. J. (2001). A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association, 96*(456), 1387-1396. https://doi.org/10.1198/016214501753382309

Kincade, L., Cook, C., & Goerdt, A. (2020). Meta-analysis and common practice elements of universal approaches to improving student-teacher relationships. *Review of Educational Research, 90*(5), 710-748. https://doi.org/10.3102%2F0034654320946836

Klassen, R. M., & Chiu, M. M. (2010). Effects on teachers' self-efficacy and job satisfaction: Teacher gender, years of experience, and job stress. *Journal of Educational Psychology, 102*(3), 741-756. https://psycnet.apa.org/doi/10.1037/a0019237

Klusmann, U., Kunter, M., Trautwein, U., Ludtke, O., & Baumert, J. (2008). Teachers' occupational well-being and quality of instruction: The important role of self-regulatory patterns. *Journal of Educational Psychology, 100*(3), 702-715. https://psycnet.apa.org/doi/10.1037/0022-0663.100.3.702

Kusumawardhani, P. N. (2017). Does teacher certification program lead to better quality teachers? Evidence from Indonesia. *Education Economics, 25*(6), 590-618. https://doi.org/10.1080/09645292.2017.1329405

Kyriacou, C. (2001). Teacher stress: Directions for future research. *Educational Review, 53*, 27-35. https://doi.org/10.1080/00131910120033628

Ladd, H. F., & Sorensen, L. C. (2017). Returns to teacher experience: Student achievement and motivation in middle school. *Education Finance and Policy, 12*(2), 241-279.

Li, X., Martens, M. P., & Wiedermann, W. (2022). Conditional Direction of Dependence Modeling: Application and Implementation in SPSS. *Social Science Computer Review*. https://doi.org/10.1177/08944393211073168

Li, X., & Wiedermann, W. (2020a). *Conditional Direction Dependence Analysis in Linear Models with SPSS Macros and Custom Dialogue Implementation* (Version 1.1) [Computer software]. https://www.ddaproject.com

Li, X., & Wiedermann, W. (2020b). Conditional direction dependence analysis: Evaluating the causal direction of effects in linear models with interaction terms. *Multivariate Behavioral Research*, 1-25. https://doi.org/10.1080/00273171.2019.1687276

Li, X., Prewett, S., & Olsen, A. A. (2020). *Predictor-residual independence tests for examining causal assumptions of synchronous effects models*. Paper presented at the American Psychological Association Annual Convention, Virtual.

Luciana, M., Conklin, H. M., Hooper, C. J., & Yarger, R. S. (2005). The development of nonverbal working memory and executive control processes in adolescents. *Child Development, 76*(3), 697-712.

Manning, A. (2016). Unlocking the power of employee happiness: What top employees seek from the workplace today. *Strategic HR Review, 15*(4), 191-192. https://doi.org/10.1108/SHR-05-2016-0038

Marks, H. M. (2000). Student engagement in instructional activity: Patterns in elementary, middle, and high school years. *American Educational Research Journal, 37*(1), 153-184. https://doi.org/10.3102%2F00028312037001153

Martin, R., Guillaume, Y., Thomas, G., Lee, A., & Epitropaki, O. (2016). Leader-member exchange (LMX) and performance: A meta-analytic review. *Personnel Psychology, 69*(1), 67-121. https://doi.org/10.1111/peps.12100

Mashburn, A. J., Hamre, B. K., Downer, J. T., & Pianta, R. C. (2006). Teacher and classroom characteristics associated with teachers' ratings of prekindergartners' relationships and behaviors. *Journal of Psychoeducational Assessment, 24*(4), 367-380. https://psycnet.apa.org/doi/10.1177/0734282906290594

McCormick, M. P., O'Connor, E. E., Cappella, E., & McClowry, S. G. (2013). Teacher-child relationships and academic achievement: A multilevel propensity score model approach. *Journal of School Psychology, 51*(5), 611-624. https://doi.org/10.1016/j.jsp.2013.05.001

McCormick, N. J., Clark, L. M., & Raines, J. M. (2015). Engaging students in critical thinking and problem solving: A brief review of the literature. *Journal of Studies is Education, 5*(4), 100-113.

Milatz, A., Lüftenegger, M., & Schober, B. (2015). Teachers' relationship closeness with students as a resource for teacher wellbeing: A response surface analytical approach. *Frontiers in Psychology*, 6, 1949. https://doi.org/10.3389/fpsyg.2015.01949

Musci, R. J., & Stuart, E. (2019). Ensuring causal, not casual, inference. *Prevention Science, 20*(3), 452-456. https://doi.org/10.1007/s11121-018-0971-9

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237-257.

O'Connor, E., & McCartney, K. (2007). Examining teacher-child relationships and achievement as part of an ecological model of development. *American Educational Research Journal, 44*(2), 340-369. https://doi.org/10.3102%2F0002831207302172

Olsen, A. A., & Huang, F. L. (2019). Teacher job satisfaction by principal support and teacher cooperation: Results from the Schools and Staffing Survey. *Education Policy Analysis Archives, 27*(11), 1-27. https://doi.org/10.14507/epaa.27.4174

Pérez-Mengual, N., Aragonés-Barbera, I., Moret-Tatay, C., & Moliner-Albero, A. R. (2021). The relationship of fear of death between neuroticism and anxiety during the Covid-19 pandemic. *Frontiers in Psychiatry*, *12*(501). https://doi.org/10.3389/fpsyt.2021.648498

Perry, L., & McConney, A. (2010). School socio-economic composition and student outcomes in Australia: Implications for educational policy. *Australian Journal of Education, 54*(1), 72-85. https://doi.org/10.1177%2F000494411005400106

Pianta, R. C. (2001). *Student-Teacher Relationship Scale: Professional manual.* Odessa, FL: Psychological Assessment Resources.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879-903. https://doi.org/10.1037/0021-9010.88.5.879

Prewett, S. L., Bergin, D. A., & Huang, F. L. (2018). Student and teacher perceptions on student-teacher relationship quality: A middle school perspective. *School Psychology International, 40(*1), 66-87. https://doi.org/10.1177%2F0143034318807743

Prewett, S., Bergin, D. A., & Huang, F. L. (2019). Student and teacher perceptions on student-teacher relationship quality: A middle school perspective. *School Psychology International, 40*(1), 66-87. https://doi.org/10.1177%2F0143034318807743

Quin, D. (2016). Longitudinal and contextual associations between teacher-student relationships and student engagement: A systematic review. *Review of Educational Research, 87*(2), 345-387. https://doi.org/10.3102%2F0034654316669434

Reddy, L. A., & Dudek, C. M. (2014). Teacher progress monitoring of instructional and behavioral management practices: An evidence-based approach to improving classroom practices. *International Journal of School & Educational Psychology, 2*, 71-84.

Reddy, L. A., Dudek, C. M., & Lekwa, A. (2017). Classroom strategies coaching model: Integration of formative assessment and instructional coaching. *Theory into Practice, 56*(1), 46-55. https://doi.org/10.1080/00405841.2016.1241944

Richter, F. G. C., & Brorsen, B. W. (2006). Aggregate versus disaggregate data in measuring school quality. *Journal of Productivity Analysis, 25*(3), 279-289. https://doi.org/10.1007/s11123-006-7644-6

Roorda, D. L., Jak, S., Zee, M., Oort, F. J., & Koomen, H. M. Y. (2017). Affective teacher-student relationships and students' engagement and achievement: A meta-analytic update and test of the mediating role of engagement. *School Psychology Review, 46*(3), 239-261. https://doi.org/10.17105/SPR-2017-0035.V46-3

Roorda, D. L., Koomen, H. M. Y., Spilt, J. L., & Oort, F. J. (2011). The influence of affective teacher-student relationships on students' school engagement and achievement: A meta-analytic approach. *Review of Educational Research, 81*(4), 493-529. https://doi.org/10.3102%2F0034654311421793

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness.* The Guilford Press.

Scheerens, J. (2001). School effectiveness research. *International Encyclopedia of the Social & Behavioral Sciences,* 13567-13572. https://doi.org/10.1016/B0-08-043076-7/02438-4

Sebastian, J., Allensworth, E., Wiedermann, W., Hochbein, C., & Cunningham, M. (2018). Principal leadership and school performance: An examination of instructional leadership and organizational management. *Leadership and Policy in Schools*, 1-23. https://doi.org/10.1080/15700763.2018.1513151

Shieh, G. (2008). Detecting interaction effects in moderated multiple regression with continuous variables power and sample size considerations. *Organizational Research Methods, 12*(3), 510-528. https://doi.org/10.1177%2F1094428108320370

Simons, K. D., & Klein, J. D. (2007). The impact of scaffolding and student achievement levels in a problem-based learning environment. *Instructional Science, 35*, 41-72.

Solomonson, A. L., & Lance, C. E. (1997). Examination of the relationship between true halo and halo error in performance ratings. *Journal of Applied Psychology*, *82*(5), 665.

Spilt, J. L., Koomen, H. M. Y., & Thijs, J. T. (2011). Teacher wellbeing: The importance of teacher-student relationships. *Educational Psychology Review, 23*, 457-477.

Steinberg, L. (2009). Adolescent development and juvenile justice. *Annual Review of Clinical Psychology, 5*, 459-485. https://doi.org/10.1146/annurev.clinpsy.032408.153603

Summers, J. J., Davis, H. A., & Hoy, A. W. (2017). The effects of teachers' efficacy beliefs on students' perceptions of teacher relationship quality. *Learning and Individual Differences, 53*, 17-25.

Thoonen, E. E. J., Sleegers, P. J. C., Oort, F. J., Peetsma, T. T. D., & Geijsel, F. P. (2011). How to improve teaching practices: The role of teacher motivation, organizational factors, and leadership practices. *Educational Administration Quarterly, 47*(3), 496-536. https://doi.org/10.1177%2F0013161X11400185

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, *4*(1), 25-29.

Urdan, T., & Schoenfelder, E. (2006). Classroom effects on student motivation: Goal structures, social relationships, and competence beliefs. *Journal of School Psychology, 44*(5), 331-349. https://doi.org/10.1016/j.jsp.2006.04.003

van de Grift, W., Helms-Lorenz, M., & Maulana, R. (2014). Teaching skills of student teachers: Calibration of an evaluation instrument and its value in predicting student academic engagement. *Studies in Educational Evaluation, 43*, 150-159. https://doi.org/10.1016/j.stueduc.2014.09.003

van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher-student interaction: A decade of research. *Educational Psychology Review, 22*(3), 271-296.

van der Lans, R. M., Cremers, J., Klugkist, I., & Zwart, R. (2020) Teachers' interpersonal relationships and instructional expertise: How are they related? *Studies in Educational Evaluation, 66*, 100902. https://doi.org/10.1016/j.stueduc.2020.100902

van der Lans, R. M., van de Grift, W. J. C. M., & van Veen, K. (2017). Developing an instrument for teacher feedback: Using the Rasch Model to explore teachers' development of effective teaching strategies and behaviors. *The Journal of Experimental Education, 86*(2), 247-264. https://doi.org/10.1080/00220973.2016.1268086

VanderWeele, T. J. (2019). Principles of confounder selection. *European journal of epidemiology, 34*(3), 211-219. https://doi.org/10.1007/s10654-019-00494-6

Veldman, I., Admiraal, W., van Tartwijk, J., Mainhard, T., & Wubbels, T. (2016). Veteran teachers' job satisfaction as a function of personal demands and resources in the relationships with their students. *Teachers and Teaching, 22*(8), 913-926. https://doi.org/10.1080/13540602.2016.1200546

Wang, M.T., & Degol, J. (2014). Staying engaged: Knowledge and research needs in student engagement. *Child Development Perspectives, 8*(3), 137-143. https://doi.org/10.1111/cdep.12073

Wang, Z., Bergin, C., & Bergin, D. A. (2014). Measuring engagement in fourth to twelfth grade classrooms: The Classroom Engagement Inventory. *School Psychology Quarterly, 29*(4), 517-535. https://doi.org/10.1037/spq0000050

Wendt, G. W., & Bartoli, A. J. (2019). Understanding the psychopathy-stress association in typical developing adults: The role of emotional deficits. *Personality and Individual Differences, 149*, 296-301. https://doi.org/10.1016/j.paid.2019.06.016

Wentzel, K. R. (1997). Student motivation in middle school: The role of perceived pedagogical caring. *Journal of Educational Psychology*, *89*(3), 411. https://psycnet.apa.org/doi/10.1037/0022-0663.89.3.411

Wentzel, K. R. (2009). Students' relationships with teachers as motivational contexts. In K. R. Wenzel & A. Wigfield (Eds.), *Educational psychology handbook series. Handbook of motivation at school* (p. 301–322). Routledge/Taylor & Francis Group

Wiedermann, W., & Li, X. (2020). Confounder detection in linear mediation models: Performance of kernel-based tests of independence. *Behavior Research Methods, 52*(1), 342-359. https://doi.org/10.3758/s13428-019-01230-4

Wiedermann, W., & Li, X. (2018). Direction dependence analysis: A framework to test the direction of effects in linear models with an implementation in SPSS. *Behavior Research Methods, 50*(4), 1581-1601. https://doi.org/10.3758/s13428-018-1031-x

Wiedermann, W., Li, X., & von Eye, A. (2020). Direction dependence analysis: Statistical foundations and applications. In *Direction dependence in statistical modeling* (pp. 9-46). https://doi.org/https://doi.org/10.1002/9781119523024.ch2

Wiedermann, W., Reinke, W. M., & Herman, K. C. (2020). Prosocial skills causally mediate the relation between effective classroom management and academic competence: An application of direction dependence analysis. *Developmental Psychology*, 56(9), 1723–1735.

Wiedermann, W., & Sebastian, J. (2019). Direction dependence analysis in the presence of confounders: Applications to linear mediation models using observational data. *Multivariate Behavioral Research*, 1-21. https://doi.org/10.1080/00273171.2018.1528542

Wiedermann, W., & von Eye, A. (2015). Direction-dependence analysis: A confirmatory approach for testing directional theories. *International Journal of Behavioral Development, 39*(6), 570-580. doi:10.1177/0165025415582056

Willingham, D. T. (2008). Critical thinking: Why is it so hard to teach? *Arts Education Policy Review,* 109(4), 21-32. https://doi.org/10.3200/AEPR.109.4.21-32

Wirkala, C., & Kuhn, D. (2011). Problem-based learning in K-12 education: Is it effective and how does it achieve its effects? *American Educational Research Journal, 48*(5), 1157-1186. https://doi.org/10.3102%2F0002831211419491

Wolfe, E. W., & Song, T. (2015). Methods for monitoring and document rating quality. In H. Jiao & R. W. Lissitz (Eds.), *The next generation of testing: Common core standards, smarter-balanced, PARCC, and the nationwide testing movement*. Information Age Publishing.

Wubbels, T., Brekelmans, M., den Brok, P., & van Tartwijk, J. (2006). *An interpersonal perspective on classroom management in secondary classrooms in the Netherlands*. In C. Evertson & C. Weinstein (Eds.), Handbook of classroom management: Research, practice, and contemporary issues (pp. 1161-1191). Mahwah, NJ: Lawrence Erlbaum Associates.

Zhang, Z., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing, 36,* 90-102.