

Teacher-student relationships: The positives and negatives of assessing both perspectives

Maureen E. Brinkworth^a, Joseph McIntyre^a, Anna D. Juraschek^a, Hunter Gehlbach^{b,*}

^a Harvard University Graduate School of Education, United States

^b University of California, Santa Barbara, United States

ARTICLE INFO

Keywords:

Interest
Motivation
Secondary school
Social emotional learning
Survey design
Teacher-student relationships

ABSTRACT

Although teacher-student relationships (TSRs) lie at the heart of teaching and learning, measuring these relationships presents unique challenges. These challenges implicate how school leaders understand the connections between TSRs and student outcomes. This article addresses these challenges by describing a new approach to measuring TSRs centered around a new scale that measures the positive and negative aspects of the overall TSR from teachers' and students' perspectives. We describe the scale development process, document the measure's psychometric properties, and then use the scale to predict student outcomes. Drawing from a sample of middle and high school students ($N = 595$) and their teachers ($N = 88$) in four different schools, we found that a two-factor model best fits our items. Compared to a more traditional approach, our approach explains more variability in student outcomes and gives educators a sharper understanding of the patterning of associations between TSRs and student outcomes.

1. Introduction

Whatever else their jobs entail, school leaders are fundamentally in the business of improving student outcomes. To help students improve, schools must identify the variables of interest, develop interventions, and assess the efficacy of these interventions. Schools interested in promoting students' social emotional learning in addition to traditional outcomes such as grades, test scores, and graduation rates face an even greater challenge because of the number of distinct areas they are trying to improve. Fortunately, one aspect of students' schooling experience predicts a disproportionately large number of these outcomes: teacher-student relationships (TSRs). Pianta and Allen (2008) note that at the secondary school level "positive relationships with adults are perhaps the single most important ingredient in promoting positive youth development" (p. 24). The relative health of these relationships has the potential to impact a tremendous array of educational outcomes including students' academic achievement, affect, behavior, and motivation (Juvonen, 2006).

Because TSRs are associated with such a wide array of benefits for adolescents, they hold tremendous potential as a locus for interventions. So many of these associations exist that even if TSRs are causally related to only a fraction of the student outcomes they are correlated with, successful TSR interventions would be a tremendous boon to any school.

However, before developing interventions, secondary school leaders need confidence that their measures that will allow them to precisely understand *how* TSRs are associated with student outcomes. For example, numerous studies document that students who report more positive TSRs tend to get better grades than their counterparts with less positive TSRs (Roorda, Koomen, Split, & Oort, 2011). This finding might occur for many reasons. Students might report greater fondness for those teachers who grade them more favorably. Alternatively, students might try harder and, thus, perform better for teachers whom they like. On the other hand, this association might be driven by teachers' perceptions of the relationship. Perhaps teachers evaluate their favorite students more favorably. As another possibility, perhaps they develop an affinity towards more highly achieving students (who make them look good). If it is the teachers' perceptions that matter, students' reports of their TSRs would correlate with their grades simply because teachers' and students' perceptions of their TSRs are correlated. The critical point is this: if a school assesses TSRs in such a way that the associations between these relationships and outcomes are clear, they will have a sharper understanding from which to develop interventions. By contrast, a flawed TSR measure could encourage school leaders to develop interventions for students when they really ought to be intervening with their teachers (or vice-versa).

This manuscript introduces a measure of TSRs at the secondary level designed to sharpen school leaders' understanding of these relationships

* Corresponding author at: Gevirtz Graduate School of Education #3113, University of California, Santa Barbara 93106, United States.
E-mail address: gehlbach@ucsb.edu (H. Gehlbach).

and how they relate to important student outcomes. With this measure, we hope to provide schools with a tool and an approach to measuring TSRs that facilitates their attempts to understand, and ultimately improve these critical relationships.

1.1. Conceptualizing teacher-student relationships

Within the academic literature, TSRs are conceptualized (and therefore measured) in diverse ways. Across this variety of conceptualizations, they manifest robust associations with desirable student outcomes. The studies reviewed in the meta-analyses of Cornelius-White (2007) and Roorda et al. (2011) show that educational researchers typically operationalize TSRs as specific facets of the relationship (e.g., teacher support, holding high expectations for students, fairness, and so forth) rather than assessing the overall relationship. See Table 1 for some of the many ways these relationships are measured. Past approaches also usually focus on the positive aspects of the relationships (rather than the positive and negative). Finally, most scholars focus on one perspective (the teacher's or the students' view of the relationship, but rarely both). Despite the array of approaches, these studies consistently find robust associations with distinct academic, affective, behavioral, and motivational outcomes for students. Thus, for the purposes of this article, these reviews (a) raise questions about how to conceptualize these relationships, but (b) suggest that regardless of which aspects are emphasized, TSRs tend to be associated with positive student outcomes.

Our conceptual framework synthesizes key ideas from attachment theory (e.g., Hamre & Pianta, 2001 and their focus on issues like conflict and closeness), parent socialization (e.g., Wentzel, 2002, and her focus on control, maturity, democratic socialization, and nurturance), and self-determination theory's emphasis on autonomy support, competency development, and relatedness (e.g., Skinner, Furrer, Marchand, & Kindermann, 2008). In particular, our conceptualization emphasizes these themes of conflict, closeness/relatedness, control, autonomy support, and competence. We also supplement these foci with social psychological theory.

Similar to the attachment perspective, we view secondary school TSRs as dyadic social processes (Pianta, 1999) that involve ongoing interactions between teachers and students in classrooms. These social and learning interactions repeat and change continually based on feedback. Inevitably, certain patterns are reinforced over time. From this basic foundation, we add two elements (as shown in Fig. 1).

First, we view TSRs as more than just interactions. After any given exchange, each party might leave with divergent understandings of the interaction. Thus, we borrow from social psychological theory on relationships to include each party's *perceptions* of one another and of their interactions as key components of the relationships (Clark & Lemay, 2010). As Gable, Reis, and Downey (2003) note, "Patterns of interaction depend on the actions and reactions of both partners, and their actions and reactions depend on each individual's perceptions and interpretations of the other's behavior" (p. 100).

Second, it is important to note that students' and teachers' behaviors and perceptions are stored in their respective memories. As Loftus' (2003) research indicates, people's memories are far from immutable. Thus, teachers' and students' memories of their interactions may change over time. These revised memories might affect their respective perceptions, in turn affecting their future interactions.

In sum, we define TSRs as teachers' and students' aggregated and ongoing perceptions of one another, affect towards each other, and interactions over time; these perceptions are stored in memory and guide future interactions with the other party. Although the overall classroom climate or teachers' and students' individual personalities may influence TSRs, these constructs are distinct from TSRs.

1.2. Operationalizing a measure of teacher-student relationships

In translating this conceptualization of TSRs into a measure that proves useful for schools, we faced three key challenges. First, as noted above, these relationships comprised a myriad of factors. Past studies have operationalized TSRs in very different ways—teacher supportiveness (Goodenow, 1993), having high expectations of students (Wentzel, 2002), teacher disinterest/criticism (Murdock, 1999), teacher caring (Gregory & Weinstein, 2008), and so on. How should school leaders think about TSRs—is it some of these? All of them?

We anticipate that most school leaders will want to think about these relationships as a whole—in other words, by assessing the *overall* relationship. While the aphorism that "the whole is greater than the sum of its parts" may be overused, it is almost certainly true in the case of TSRs. Because strengths in certain areas of a TSR may compensate for weaknesses in other areas, looking at specific aspects of TSRs might be more misleading for school leaders than assessing the overall relationship. Moreover, while researchers may conduct focused, in-depth studies asking a multitude of questions about a narrow aspect of a TSR, few schools have that luxury. School assessment systems must cover numerous topics. Thus, unlike many past approaches, we examine the overall TSR by assessing a cross-section of key characteristics of TSRs within a single scale.

A second challenge is ascertaining whose perception of TSRs is of interest. TSRs are two-way streets; teachers and students construct these relationships together. Thus, members of the TSR dyad make up a 'relational unit' that may not be fully understood by tapping the perspective of a single party. Simply because a student reports liking a teacher, does not necessitate that those feelings are reciprocated. Most past scholarship takes this view theoretically. However, when it comes time to measure TSRs few studies account for both perspectives: "Missing from the literature is a description of the same child-teacher relationship from its two participants," (Pianta, Hamre, & Stuhlman, 2003, p. 218). Thus, our measure of TSRs incorporates both teachers' and students' perceptions.

A third challenge to assessing TSRs is that these relationships may be positive, negative, neither, or both. Simply because a relationship is very positive in some ways does not preclude the possibility that it may be very negative in other ways. A teacher may simultaneously feel strong positive and negative feelings towards a particular student. Furthermore, a student who does not feel particularly positively towards a teacher may or may not feel strong antipathy towards that teacher. In the study of attitudes, we know that theoretical opposites are not always neatly arrayed along a single continuum once the data are examined (Cacioppo & Berntson, 1994). We might reasonably expect the same of teachers' and students' attitudes towards one another. Thus, our measure allows for separate positivity and negativity dimensions.

In sum, we developed scale items with the goal of assessing TSRs holistically, from both teachers' and students' perspectives, and while accounting for the positive and negative aspects of TSRs independently. Although other measures, such as Pianta's (2001) Student-Teacher Relationship Scale, examine positive and negative aspects of TSRs (e.g., closeness and conflict), they do not attempt to capture the overall relationship (for example, the academic side of the relationship, respect, and encouragement, are not addressed). Thus, although other scales may have some of the same features as ours, we believe this scale would be the first at the secondary school level to assess (1) the positive and negative sides of TSRs, (2) the overall relationship, and (3) both teachers' and students' perspectives.

With these features in mind, we take three main steps to evaluate whether this approach to measuring TSRs at the secondary level makes a useful (rather than merely novel) contribution to science and practice. Because our scale design process (Gehlbach & Brinkworth, 2011) is unique in its approach to building a case for validity from the outset, we first detail this process in the hopes that it contributes to the dialogue

Table 1
TSR indicators collected through the scale development process and illustrative preliminary TSR items.

Indicators	From academic literature review (sample reference)		From focus groups and interviews		Synthesized list	Illustrative preliminary items
			Teachers	Students		
Caring (Wentzel, 1997) Communication (Johnson, 2008)			Kindness	Caring, kindness, approachability Being clear and direct, listening	Caring, liking Positive and clear communication	During class, how kind is this student/teacher when s/he talks to you? During class, how often does this student talk when you are talking (for instance, when this student is supposed to be listening?/During class, how often do you talk when this teacher is talking (i.e., when you are supposed to be listening)?
Conflict (Ang, 2005)			Conflict	Conflict	Anger/conflict	What percentage of the time do you feel upset when you interact with this student?/What percentage of the time does this teacher seem to be upset when you interact with him/her?
Criticism (Murdock, Anderman, & Hodge, 2000)			Talking back	Criticism	Disrespect, offensive, causes discomfort	How many times do you make this student feel upset in an average week?/ How many times does this teacher make you feel upset in an average week?
Democratic Interactions/Regard for Adolescent Perspective (Hamre & Pianta, 2006; Wentzel, 2002) Expectations (Muller, 2001; Murdock, 1999)			Student engagement Reasonable expectations for teachers	Interest in student opinions and concerns, enjoying students Teacher's belief in students' abilities	Engaging with student opinions Expectations	In the past week, how many times has this student spoken with you about something unrelated to class? How reasonable are this student's/teacher's expectations of you?
Fairness/Equity (Murdock, 1999)				Fairness, not favoring specific students, reasonable use of authority/power	Fairness	How unfair are the grades that you give this student in this class?/How unfair are the grades this teacher gives you in this class?
Instructional Support (Garza, 2009)			Student responsiveness	Patience, availability, teacher responsiveness	Instructional support, learning	How patient is this student/teacher when s/he listens to you?
Motivation (Martin & Dowson, 2009) Respect (Johnson, 2008; Wentzel, 2002) Trust (Roeser et al., 1996)			Interest in subject matter Respect Trust, honesty Sense of humor	Enthusiasm, encouragement Respect Trust, consistency Sense of humor, teasing, mocking	Motivation, encouragement Respect Trust/admiration Humor	How motivating is it to work with this student/teacher in class? During class, how respectful is this student/teacher towards you? How trustworthy is this student/teacher? How often does this student's/teacher's humor bother you? How friendly is this student/teacher towards you?
Warmth (Murray & Pianta, 2007)			Friendliness	Friendliness	Friendliness	

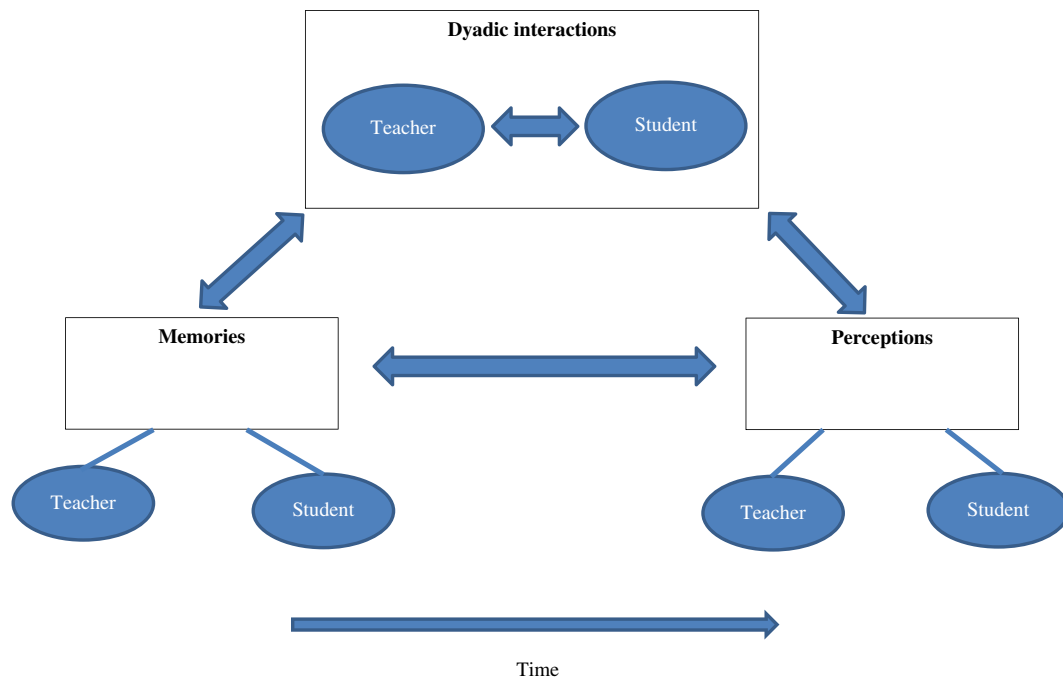


Fig. 1. A model of teacher-student relationships.

around how to best assess psychological constructs (DeVellis, 2003).

Second, after describing our participants, measures, and survey administration procedures, we tackle the question of whether the scale shows promising evidence of validity. In particular, we ask: (1) Which factor structure best fits our data? (2) how reliable are the different subscales? (3) how much evidence of measurement invariance over different levels of schooling do we find? and, (4) to what extent can we replicate robust findings from past research?

Third, because a new scale might show strong evidence of validity without contributing any new insights into how TSRs functioned, we contrasted two approaches to measuring TSRs. In one approach, we predicted a set of student outcomes using a full model which used two TSR subscales from teachers and two from students. We compared this approach against a reduced model that served as a proxy for how many studies typically approach the measurement of TSRs at the secondary level (i.e., only assessing TSR-positivity from students' perspectives). The comparison allowed us to ask: (5) Is students' TSR-positivity positively and significantly associated with all 8 outcomes in the "reduced" model? (6) does a "full" model that includes all four TSR sub-scales explain significantly more variability in the eight outcomes than the reduced model?, and (7) to what extent do the patterns of associations between students' TSR-positivity and each outcome differ between the reduced and full models?

1.3. Scale development process

In addition to addressing these challenges, we hoped to imbue our TSR scales with other attributes as well. To optimize its utility for school assessment systems, we wanted to keep the scale short. Length seemed particularly critical because we assumed teachers would frequently need to complete the survey for multiple students. Because we hoped it could assess TSRs longitudinally, the scale needed to work for middle and high school students alike. Furthermore, the language needed to be simple enough for 6th graders without patronizing teachers. To address all these constraints, and to try to enhance the measure's validity from the outset of the design process, we followed the survey design process detailed by Gehlbach and Brinkworth (2011):

- Step 1—Literature review: We reviewed relevant literature to understand the range of conceptualizations and operationalizations of TSRs as described above. This review shaped our definition, identified the key characteristics of TSRs, and helped us amass potential items for consideration.
- Step 2—Interviews and focus groups: Second, we conducted open-ended interviews and focus groups with prospective respondents to learn how they conceptualized and understood TSRs. In this way, we obtained a critical "second opinion" on which indicators were central to our construct.
- Step 3—Synthesizing a list of indicators: Third, we compared responses from these interviews and focus groups against our literature review to determine points of overlap, divergence, and disparities in terminology. By the end of this step, we had a synthesized list of indicators from previous literature as well as from teachers and students that represented the key ingredients of TSRs in secondary schools (see Table 1).
- Step 4—Developing items: Fourth, we converted our indicators into survey items in accordance with the best practices in survey design (Dillman, Smyth, & Christian, 2014). All too often schools end up using measures that increase measurement error by asking statements rather than questions, using reverse-scored items, and committing other "survey sins" (Gehlbach, 2015).
- Step 5—Expert review: Fifth, to ensure that the items still mirrored our conceptualization of TSRs, to confirm that we had not overlooked any crucial indicators, and to obtain additional evidence of construct validity (Messick, 1995), we subjected our items to an expert review. Twenty TSR experts assessed each item of the teacher TSR scale and twenty others focused on the student scales, depending on their expertise (McKenzie, Wood, Kotecki, Clark, & Brey, 1999; Rubio, Berg-Weger, Tebb, Lee, & Rauch, 2003).
- Step 6—Cognitive pre-testing: Sixth, we employed a cognitive pre-testing, or "think-aloud," procedure to see whether respondents comprehended each item as we intended and to ensure that each item triggered appropriate recollections as participants formulated responses.

After completion of the cognitive pre-testing phase, we conducted a

Table 2
Student and teacher participants for each school and for the total sample.

	School 1	School 2	School 3	School 4	Total
School description	Private, Christian, urban, 6th–12th grades	Public, suburban, 6th–8th grades	Private, Catholic, urban, 9th–12th grades	Military/vocational, urban, 9th–12th grades	
Survey mode	Paper and pencil	Paper and pencil	Web	Paper and pencil	
Student participants	(<i>n</i> = 144)	(<i>n</i> = 118)	(<i>n</i> = 198)	(<i>n</i> = 135)	(<i>N</i> = 595)
	55% Female	50% Female	57% Female	53% Female	55% Female
	19% Asian	1% Asian	2% Asian	0% Asian	5% Asian
	17% Black	3% Black	28% Black	64% Black	29% Black
	10% Hispanic	10% Hispanic	46% Hispanic	10% Hispanic	22% Hispanic
	38% White	58% White	5% White	9% White	24% White
	16% Other	28% Other	19% Other	17% Other	20% Other
Teacher participants	(<i>n</i> = 25)	(<i>n</i> = 31)	(<i>n</i> = 4)	(<i>n</i> = 28)	(<i>N</i> = 88)
	56% Female	65% Female	50% Female	61% Female	60% Female
					86% White

series of small-scale pilot tests of the measure in which we administered the items to students and teachers in their classrooms. Based on these results, we made minor additional changes.

By addressing certain aspects of content validity such as construct relevance (e.g., Steps 1–3), representativeness (e.g., Step 5), and technical quality (e.g., Steps 4 and 6) as well as substantive validity (e.g., Steps 5 and 6) into the design process (Messick, 1995), we viewed our scale as having good preliminary evidence of validity. However, assessing the validity of a new measure is an ongoing process (Gehlbach, 2015; Messick, 1995); thus, we hoped to accumulate additional indicators of other types of validity. For instance, a key question remained around the scale's structural validity—do the positive and negative aspects of TSRs require separate sub-scales or will our positive and negative indicators load onto one factor? Relatedly, we sought evidence that the scale was internally consistent. Because we wanted the measure to be useful for longitudinal research we hoped to show aspects of measurement invariance between middle and high school students. Finally, we hoped to replicate some basic findings from previous studies to show that the measure functioned in ways that were congruent with prior research.

2. Methods

To obtain data that would simultaneously provide evidence for our scale's construct validity, predictive validity, and practical utility for educators, we administered the TSR scale to students and teachers at four secondary schools in the northeastern United States and collected an array of student outcomes. At each site, we assigned each student to report on one of their teachers. Except for School 3 (where all students reported on their English teacher), we selected these teachers randomly. Specifically, we identified all of the participating teachers on each student's schedule and randomly selected one teacher for that student to report on. Teachers subsequently reported on each student who had reported on them. Thus, in our analyses students are nested within only one teacher at each school.

2.1. Participants

Five-hundred and ninety-five students and 88 teachers from four different participating schools volunteered to participate in the study. In the hopes of getting approximately 500 students so that we might have sufficient statistical power and examine some important subgroups of interest, we chose a diverse array of schools for our study—see the summary demographic characteristics of each institution and the overall sample in Table 2. Teachers in our sample represented 12 subject areas, including: Art, Biblical Studies, Computer Science, Economics, English, Math, Music, Physical Education, Science, History/Social Studies, Foreign Language, and Technology. Many

students sampled had the same teacher. The number of students nested within teacher varied widely (from 1 to 78, $M = 6.76$, $SD = 10.69$, $median = 4$).

At each school, we computed a total participation rate¹ (Hoynoski, Link, & Frankel, 2009). In other words, we calculated the fraction of students who participated out of the entire student body at each school. Participation rates were 69% (School 1), 23% (School 2), 67% (School 3), and 42% (School 4). In terms of racial composition, the students in our sample appeared to be generally representative of their larger student bodies. The percentages of White students in our samples versus their respective school populations were as follows: 38% versus 34% (School 1), 61% versus 71% (School 2), 5% versus 6% (School 3), and 9% versus 8% (School 4). In those instances where the percentages in our sample differed markedly from the larger school population (e.g., in School 3 our sample reported as 28% Black/African American whereas the school reports a 53% Black population), it was primarily due to our participants choosing more specific categories (e.g., students writing in “Haitian” in the “other” category). For teachers, we calculated participation rates as the fraction who participated out of all the teachers at each school. These rates were 83% (School 1), 67% (School 2), 19% (School 3—where we worked exclusively with all the teachers in the English department), and 74% (School 4).

2.2. Measures

In addition to having students and teachers complete our new TSRs measure, we also collected several student outcomes of interest with the goal of representing commonly studied variables across several important domains: academic, affect, behavior, and motivation. Due to different time constraints and our efforts to tailor results to the interests of the schools, we collected slightly different measures at each school (see Appendix B for complete details on the exact measures used at each school). For both teachers and students, we also collected basic demographic data (gender, race, age, grade-level, primary language spoken at home, and parents' educational level).

2.2.1. Teacher-student relationships

In each school, students and teachers completed parallel versions of our TSR measures (see Appendix A for the student and teacher versions of each subscale). Because the validity of this measure is of central interest, we describe the psychometric properties of these scales in the first part of the Results section.

¹ We preferred total participation rate to response rate because the latter fails to account for portions of the sampling frame that were omitted (e.g., students who were absent on the day that the consent forms were sent home).

2.2.2. Academic

We assessed academic outcomes through two measures. First, we examined students' grade (using a 0–100 scale) in the class that they had with the teacher of interest, i.e., their focal class. Second, after teachers assessed students' class participation (see the “behavior” measures below), they rated the contribution quality of those times when the student participated.

2.2.3. Affect

To measure students' overall affect towards school, we assessed their sense of belongingness at school. We borrowed the 4-item scale used by Roeser, Midgley, and Urdan (1996) in which students' rated statements such as “I feel like I matter in this school.” ($\alpha = 0.76$). To balance this global sense of how students fit in at their school, we also investigated their level of interest in their subject matter (3 items, $\alpha = 0.73$) through items such as, “If you could choose to take any classes you wanted to in high school, how many classes would you take in this subject?” This measure was adapted from other surveys (Maehr, 1976; Midgley et al., 2000).

2.2.4. Behavior

To assess behaviors related to students' academic performance, we collected two measures. First, teachers reported the percentage of homework that students completed and, second, we asked for their subjective impression of the frequency of students' class participation.

2.2.5. Motivation

As one measure of motivation in their focal classes, students reported how much effort they expended for class. This 5-item self-report scale included items such as, “How much effort do you put forth for this class?” ($\alpha = 0.82$). To assess students' self-efficacy in the course with their focal teacher, we adapted the scale used by Gehlbach et al. (2008). This 5-item scale assessed how confident students were with regard to different aspects of the course in question with items such as, “How confident are you that you can learn all the material presented in this class?” ($\alpha = 0.85$).

2.3. Procedures

At each school, we first described the study to teachers to ensure that most were willing to participate and obtained their consent accordingly. Next, we sent home consent forms for students and their parents/guardians to sign, usually through students' homeroom teacher. Once we confirmed our participating students, we obtained students' class schedules and then randomly selected a participating focal teacher for each student (except at School 3). Based on the randomly selected teacher, we then created individualized surveys for each student (i.e., a survey that referenced the student and focal teacher by name, as well as the specific class the student had with that teacher). A member of the research team administered the survey to students in paper and pencil format or via the web; to ensure the confidentiality of student responses, no teachers were present during the administrations. After receiving the student surveys, we sent each teacher a parallel form of the survey to complete at their own convenience. This survey included an individualized section for teachers to complete the TSR items for each student who had reported on that teacher. Teachers were paid \$4 per student who they reported on and returned the surveys to us within two weeks.

Data were collected during the fall at Schools 1 and 2 and in January just before the end of the first semester at Schools 3 and 4. At Schools 1 and 2, we also completed a follow-up data collection towards the end of the school year which was used to compute test-retest reliability.

3. Results

We approached our analyses with two overarching goals in mind. First, we hoped to provide additional evidence of the scale's validity beyond what we established through our scale development process. Second, we sought to examine whether our more intensive approach to assessing TSRs provided useful information over and above typical (and more straightforward) approaches to measuring this construct.

3.1. Scale validity results

In assessing additional evidence relevant to our scale's validity, we focused on four questions: (1) Which factor structure best fits our data? (2) how reliable are the different sub-scales? (3) how much evidence of measurement invariance over different levels of schooling do we find? (4) to what extent can we replicate prior findings such as the general decline in the health of these relationships between lower and higher grade levels (Eccles et al., 1993) or signs that teachers perceive more positive relationships with their female students and more negative relationships with their male students given differences in students' behavior (Chun & Mobley, 2010).

3.1.1. Factor structure findings

To examine whether a one- or two-factor model best fit our data, we employed confirmatory factor analyses, using Mplus version 6—to fit a model with latent TSR factors for teachers and for students simultaneously. Because the indicators for each latent variable (i.e., the TSR items) were ordered-categorical variables, we employed the *CATEGORICAL* option in Mplus. The confirmatory factor analyses were complicated by the fact that students were nested within teachers. To account for the nesting in our data, we utilized the *CLUSTER IS* option to identify “teacher” as a grouping variable. Due to this complex structure of our data, we relied on mean- and variance-adjusted weighted least squares for complex survey data (*WLSMV-complex*) estimation and utilized the *DIFF TEST* function to compare the chi-squared values of our nested models. *WLSMV-complex*, which uses a variance correction procedure, accounts for the dependence between the observations and provides corrected standard errors, confidence intervals, and coverage (Asparouhov, 2005). We used full information maximum likelihood (FIML) to address missing data. Less than 2% of responses were missing for any indicator, making the use of FIML appropriate in this case.

In testing a single factor model (model 1) and a TSR-positivity and TSR-negativity model (model 2), we found that both models failed the Chi-squared test of exact model fit. However, using fit indices (CFI and RMSEA) that are better suited to larger sample sizes, we found that each model provided a reasonable fit for our data (Kline, 2011). The chi-square difference test (Δ in $\chi^2_{(5df)} = 194.22$, $p < 0.001$), the differences in CFI (0.96 versus 0.93), and the differences in RMSEA (0.044, 90% CI [0.039, 0.048] versus 0.058, 90% CI [0.053, 0.062]²) all indicated that model 2 provided a better fit than model 1. Furthermore, although the estimated correlations between the latent positivity and negativity factors (−0.75 and −0.66 for students and teachers, respectively) are large, they are small enough to suggest the existence of two distinct but related constructs. Finally, model 1 had far more residual correlations above 0.1—which may indicate problematic misfit—as compared to model 2. This is especially important in a factor with many indicators where indices of overall model fit, such as the RMSEA and CFI, may hide severe local misfit. See Fig. 2 for the factor loadings of our final two-factor model.

² We present 90% confidence intervals for the RMSEA rather than 95% confidence intervals because guidelines for acceptable ranges of RMSEA values are typically given in terms of 90% confidence intervals. See, e.g., Kline (2011), p. 206.

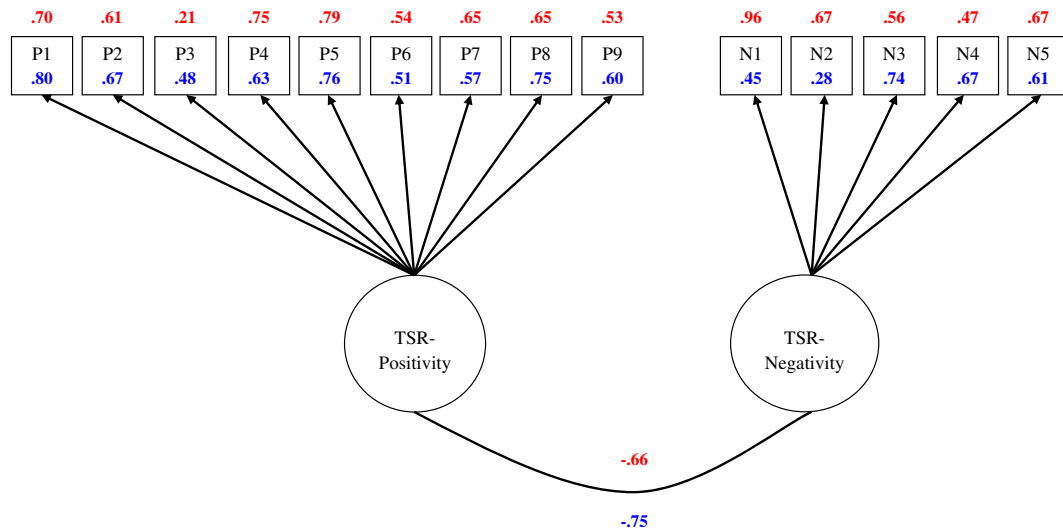


Fig. 2. The amount of variance explained by the latent construct in each indicator of the TSR scale is denoted above each indicator for teachers (in red) and inside each indicator (in blue) for students.

Table 3
Descriptive statistics and correlations for items in the Teacher-Student Relationship Scale – positivity & negativity sub-scales.

	<i>M</i>	<i>sd</i>	P1	P2	P3	P4	P5	P6	P7	P8	P9	TSR Pos. Scale	N1	N2	N3	N4	N5	TSR Neg. Scale	<i>M</i>	<i>sd</i>
P1	3.52	1.12	--	0.53	0.30	0.58	0.71	0.51	0.50	0.63	0.54	0.79	-0.45	-0.38	-0.26	-0.23	-0.30	-0.47	4.04	0.74
P2	4.01	0.94	0.62	--	0.46	0.45	0.49	0.43	0.70	0.58	0.42	0.77	-0.32	-0.19	-0.10	-0.07	-0.19	-0.26	3.76	0.91
P3	3.2	1.21	0.55	0.56	--	0.19	0.27	0.31	0.43	0.32	0.26	0.53	-0.12	-0.12	0.00	-0.07	-0.03	-0.11	3.49	0.81
P4	4.16	0.89	0.51	0.65	0.45	--	0.66	0.47	0.49	0.56	0.48	0.73	-0.62	-0.58	-0.40	-0.26	-0.44	-0.67	4.07	0.81
P5	3.04	1.42	0.75	0.54	0.52	0.47	--	0.55	0.49	0.67	0.51	0.81	-0.56	-0.52	-0.26	-0.23	-0.35	-0.57	3.81	1.04
P6	3.13	1.11	0.64	0.47	0.45	0.41	0.59	--	0.48	0.47	0.62	0.73	-0.43	-0.35	-0.31	-0.27	-0.36	-0.47	3.28	0.86
P7	3.48	1.05	0.55	0.62	0.59	0.56	0.55	0.50	--	0.61	0.48	0.79	-0.37	-0.30	-0.15	-0.13	-0.20	-0.35	3.2	0.99
P8	3.47	1.19	0.72	0.62	0.52	0.50	0.73	0.57	0.61	--	0.40	0.79	-0.42	-0.35	-0.19	-0.14	-0.24	-0.40	3.93	0.83
P9	3.85	1.05	0.69	0.47	0.43	0.49	0.60	0.57	0.49	0.59	--	0.69	-0.33	-0.30	-0.31	-0.21	-0.37	-0.41	3.61	0.71
TSR Pos. Scale	3.54	0.87	0.86	0.78	0.73	0.70	0.84	0.74	0.77	0.84	0.76	--	-0.55	-0.47	-0.29	-0.24	-0.37	-0.56	3.69	0.63
N1	2.14	1.04	-0.43	-0.30	-0.23	-0.31	-0.41	-0.36	-0.34	-0.39	-0.43	-0.46	--	0.73	0.34	0.29	0.38	0.85	1.93	1.01
N2	2.31	1.15	-0.29	-0.18	-0.18	-0.21	-0.30	-0.20	-0.22	-0.29	-0.28	-0.31	0.49	--	0.33	0.25	0.38	0.85	2.09	1.1
N3	1.4	0.85	-0.35	-0.47	-0.30	-0.56	-0.32	-0.33	-0.38	-0.38	-0.37	-0.48	0.37	0.23	--	0.47	0.61	0.65	1.32	0.56
N4	1.58	0.95	-0.45	-0.46	-0.31	-0.45	-0.40	-0.35	-0.36	-0.45	-0.42	-0.51	0.40	0.32	0.54	--	0.45	0.54	1.17	0.41
N5	1.62	0.98	-0.43	-0.39	-0.33	-0.39	-0.46	-0.38	-0.28	-0.45	-0.40	-0.50	0.40	0.35	0.46	0.60	--	0.69	1.34	0.57
TSR Neg. Scale	1.81	0.72	-0.53	-0.48	-0.36	-0.51	-0.52	-0.44	-0.43	-0.53	-0.52	-0.61	0.74	0.69	0.68	0.77	0.76	--	1.57	0.55

Notes: Student scores are on the left and below the diagonal; teacher scores are on the right and above the diagonal. Bolded values represent composite means and standard deviations. Shaded cells represent correlations involving the composite scores.

3.1.2. Reliability estimates

With the two factor model established as more appropriate, we then estimated each sub-scale's reliability. For teachers, the coefficient alpha of the positivity and negativity sub-scale scores were $\alpha = 0.90$, 95% CI [0.89, 0.91] and 0.78, 95% CI [0.75, 0.81], respectively. For students, these α s were 0.92, 95% CI [0.91, 0.93] and 0.78, 95% CI [0.75, 0.81], respectively.³ We also used semTools (Pornprasertmanit, Miller, Schoemann, & Rosseel, 2013), a package in R (R Core Team, 2013) to compute omega (ω), an estimate of reliability calculated as the model-

implied variability in the indicators attributable to the underlying factors divided by the total variability in the indicators. We found $\omega = 0.91$ and 0.86 for teachers and $\omega = 0.94$ and 0.86 for students on the positivity and negativity subscales, respectively. Finally, we computed test-retest correlations at two schools. We expect some real shift in the relationships between teachers and students over the course of the year, thus we view these estimates as lower bounds to the test-retest reliability of our scales. Among teachers, we found the test-retest correlations (unadjusted for nesting) for the positivity scale score was $r_{(218)} = 0.61$, $p < 0.001$, 95% CI [0.52, 0.69]; meanwhile $r_{(218)} = 0.60$, $p < 0.001$, 95% CI [0.51, 0.68] for the negativity scale score. For students, these test-retest correlations were $r_{(230)} = 0.55$, $p < 0.001$, 95% CI [0.45, 0.63]; and $r_{(230)} = 0.57$, $p < 0.001$, 95% CI [0.48, 0.65], respectively. We present full descriptive statistics and correlations for each item of both subscales in Table 3.

³ Here and elsewhere we report statistics which treat the indicators as interval-level data. However, all of our models treat the indicators as ordinal. We believe it is more sensible to regard the indicators as ordinal measures of the construct in question, but acknowledge that most schools may prefer to treat them as interval, which is why we present statistics such as α , and means and standard deviations of items and scale scores.

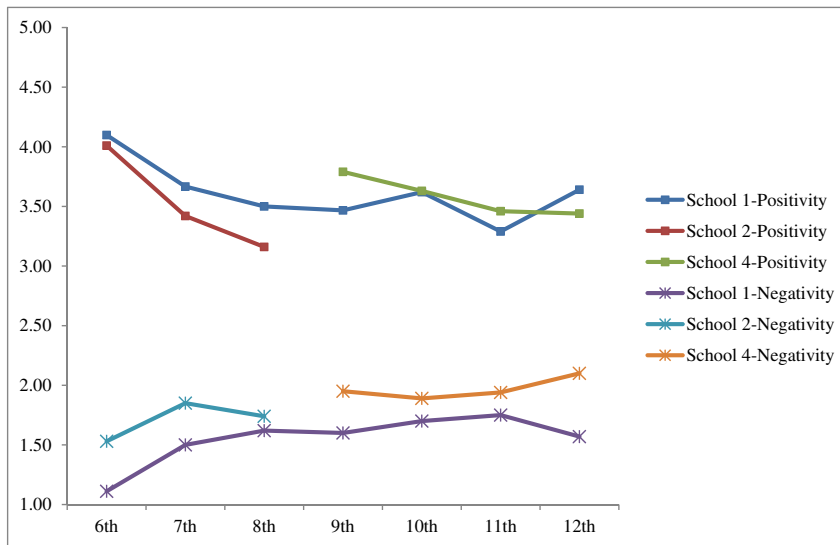


Fig. 3. Students' mean level TSR-positivity and TSR-negativity by grade at schools 1, 2, and 4.

3.1.3. Measurement invariance

Next, we investigated whether students' scores for this two factor model varied between the middle and high school students. Using a weighted least squares estimator and theta parameterization, we fit an unconstrained model in which thresholds and factor loadings were free across middle school and high schools groups, residual variances were fixed at one in both groups and factor means were fixed at zero in both groups. The fit statistics for this model were $\chi^2[688] = 1123.028$; CFI = 0.961; RMSEA = 0.046; 90% CI [0.041, 0.051].

To test for invariance between groups, we fit a model in which thresholds and factor loadings were constrained to be equal across middle and high school groups, residual variances were fixed at one and factor means were fixed at zero in the middle school group and free in the high school group. The fit indices for this model were $\chi^2[789] = 1202.98$; CFI = 0.963; RMSEA = 0.042; 90% CI [0.037, 0.047]. Using the DIFFTEST command in MPlus, the results of a χ^2 difference test between the unconstrained and constrained models were Δ in $\chi^2 = 125.77$, Δ in df = 101, $p = 0.05$.

Testing for measurement invariance using significance tests can reveal statistically significant but practically unimportant differences in the measurement properties of a model, especially when the sample is large. Therefore, we followed Cheung and Rensvold's (2002) recommendation to examine change in the model CFI caused by imposing additional constraints, and to take decreases in the CFI of -0.01 or greater as evidence of problematic non-invariance (see Chen, 2007 for an extension to models with ordinal indicators). When we imposed the constraints, the model fit improved from 0.961 to 0.963, indicating no problematic misfit.

Finally, we examined the factor loadings from the constrained and unconstrained models and found them to be extremely similar suggesting once more that the measurement model is essentially the same for middle school and high school students.

3.1.4. Scale properties and descriptive statistics

In seeking additional evidence of validity beyond the content, substantive, and structural aspects already described, we explored two aspects of external construct validity (Messick, 1995). First, we tested whether we would see the robust decline in perceived relationships across grade levels documented by Eccles and her colleagues (Eccles et al., 1993). To the extent that our scale shows a decrease in TSR-positivity and or an increase in TSR-negativity with increasing grade levels, it provides a signal that our measure is functioning in ways that are congruent with past investigations. In addition, based on previous findings that girls typically outperform boys academically

(Woolley & Grogan-Kaylor, 2006), we hypothesized that teachers would perceive more positive relationships with their female students; based on other findings from those studies showing that boys caused more behavioral problems (Chun & Mobley, 2010), we expected that teachers would perceive more negative relationships with boys.

To assess whether certain differences in TSRs emerged across particular subgroups of students, we ran a series of multi-level structural equation models that nested students within teachers using the CLUSTER IS option and controlled for school through fixed-effects. With respect to our first hypothesis—that students in earlier grades would perceive more positive TSRs than their peers in later grades—in Fig. 3, we show the means for each grade at each school. Schools 1, 2, and 4 illustrate trends that are largely consistent with our hypothesis. (To preserve the anonymity of the 4 teachers at School 3, we exclude this school in the figure). For an overall test of our hypothesis on our full data set, we used a multi-level structural equation model to regress students' TSR-positivity and negativity on students' grade level while controlling for school. Results indicated that older students reported lower levels of TSR-positivity ($\beta = -0.23$; $SE = 0.09$; 95% CI $[-0.41, -0.04]$; $p = 0.02$) than younger students. Though not significant, the differences on TSR-negativity suggested a similar trend ($\beta = 0.08$; $SE = 0.04$; 95% CI $[-0.003, 0.16]$; $p = 0.06$).

With respect to our second hypothesis—that teachers should feel more positively towards their female students and more negatively towards the boys in their classes—the results showed partial support of our conjecture. Teachers' TSR-positivity ratings were not significantly different for male and female students ($\beta = -0.20$; $SE = 0.14$; 95% CI $[-0.48, 0.08]$; $p = 0.16$), although the trend is in the direction we expected. However, teachers rated their relationship with their male students as more negative on the TSR-negativity scale ($\beta = 1.23$; $SE = 0.52$; 95% CI $[0.21, 2.26]$; $p = 0.02$).

3.2. Teacher-student relationships and student outcomes

Having developed our TSR scale and established multiple indicators of validity, we then wanted to know if the scale was actually useful for researchers and practitioners. In particular, with parallel versions for teachers and students and with TSR-positivity and negativity subscales, our instrument requires more effort than a more traditional approach such as simply asking students to complete a scale of their perceptions of their teacher. Thus, we wanted evidence that our scale provided compensatory explanatory power. We focused on three additional research questions all designed to help determine whether our approach to measuring TSRs might provide an important complement to

Table 4

Descriptive statistics and correlations for Teacher-Student Relationship Scales and student outcomes (unadjusted for nesting).

	<i>N</i>	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11
1) S_TSR +	591	3.54	0.87	–										
2) S_TSR –	591	1.81	0.72	–.61**	–									
3) T_TSR +	589	3.69	0.63	.42**	–.37**	–								
4) T_TSR –	589	1.57	0.55	–.26**	.41**	–.56**	–							
5) Class grade	473	82.46	11.02	.21**	–.26**	.28**	–.32**	–						
6) Contribution quality	437	3.55	0.80	.15**	–.10*	.42**	–.29**	.41**	–					
7) Sense of belonging	577	3.69	0.91	.26**	–.26**	.14**	–.13**	.20**	.05	–				
8) Interest	248	2.76	1.05	.54**	–.35**	.27**	–.26**	.16	.15*	.19**	–			
9) Participation frequency	437	3.26	1.17	.16**	–.06	.46**	–.14**	.22**	.47**	.12*	.22**	–		
10) Homework submission	543	84.63	16.72	.16**	–.21**	.36**	–.38**	.62**	.30**	.12**	.14	.09	–	
11) Effort	437	3.70	0.77	.51**	–.38**	.35**	–.18**	.19**	.15**	.30**	.59**	.29**	.24**	–
12) Self-efficacy	437	3.63	0.84	.47**	–.33**	.28**	–.12*	.23**	.24**	.28**	.63**	.24**	.16**	.65**

Note: S_TSR + is students' TSR-positivity; S_TSR – is students' TSR-negativity; T_TSR + is teachers' TSR-positivity; T_TSR – is teachers' TSR-negativity. All variables were arrayed on a 1 to 5 scale except for class grade and homework submission whose scales ranged from 0 to 100.

* $p < 0.05$.

** $p < 0.01$.

traditional, past approaches:

- 5) Is students' TSR-positivity positively and significantly associated with all 8 outcomes in the “reduced” model?
- 6) Does a “full” model that includes all four TSR sub-scales explain significantly more variability in the eight outcomes than the reduced model?
- 7) To what extent do the patterns of associations between students' TSR-positivity and each outcome differ between the reduced and full models?

The implicit logic of these three questions is as follows. Typically, past studies have examined TSRs and their associations with student outcomes through a single measurement approach—at the secondary level, this approach typically examines students' perceptions of the TSR and often utilizes a discrete, positive aspect of these perceptions (e.g., Wentzel, 1997 and her study of students' perceptions of teachers' caring). If we approximate this approach by looking only at students' perceptions of TSR-positivity, do we find associations with all 8 outcomes as we would expect based on previous scholarship? An affirmative answer to this question provides a sense that our overall measure of the construct functions similarly to past measures of discrete aspects of TSRs, suggests our sample is probably not overly idiosyncratic, and can be used as a reasonable proxy of some of the past investigations of TSRs at the secondary level. The real question for our new scale, then, is whether the additional sub-scales provide additional, important predictive power for these outcomes when we examine all four sub-scales simultaneously. If additional variability is explained, it suggests that the scale provides useful information beyond more traditional approaches to measuring TSRs. If the new scale shows a different pattern of associations between TSRs and student outcomes than was previously understood (e.g., if it is actually the teachers' rather than the students' perceptions that matter for a particular outcome), then the scale might sharpen researchers or practitioners understanding of the exact nature of these relationships.

Conducting the analyses to investigate these three questions again required managing complex data. Because all the key predictor variables (and four of the outcomes) are composite scales, we used structural equation models to address measurement error. Students are nested within teachers (which is a critical aspect of our research questions) and teachers are nested within school (which is peripheral to our research questions). Once again we used Mplus version 6, constructing separate latent factors for all unobserved outcomes and predictors. We treated all ordinal indicators as categorical and used cluster-robust standard errors at the teacher level, and used FIML to

address missing data. Because we were not interested in idiosyncratic differences between schools, we controlled for school using fixed effects. We present descriptive statistics for each TSR sub-scale and each of our outcomes in Table 4. Across all our models we controlled for students' race and gender, as well as teachers' gender. We did not control for teacher race because such a large majority of teachers in our sample were white.

To answer our first research question, we fit a series of structural equations models in which each outcome was regressed on the students' TSR-positivity scale (our “reduced model”). Next, we added teachers' TSR-positivity and students' and teachers' TSR-negativity to the model as latent factors and added paths from these factors to the outcome (our “full model”), then used Mplus to conduct a Wald test to test the null hypothesis that these paths were simultaneously equal to 0. To answer the third research question, we compared the regression coefficient of each outcome on students' TSR-positivity from the reduced and full models.

3.2.1. Does the reduced model predict student outcomes?

Congruent with our expectations, we found that students' perceptions of TSR-positivity were positively and significantly associated with each student outcome (see the first column of Table 5). Thus, we provide a conceptual replication of past studies that investigated these relationships by asking students about their perceptions of specific aspects of their relationship with their teacher and correlating their responses with student outcomes. Although our reduced model differs from this approach in that we take a holistic approach to assessing TSRs, the results are similar to these past findings (see Roorda et al., 2011 for examples).

3.3. Does a “full” model that includes all four TSR sub-scales explain significantly more variability in the outcomes than the reduced model?

We answered this question by fitting a model in Mplus with the outcome regressed on all four latent factors. Then, using a Wald test, we obtained a p -value associated with simultaneously constraining the paths from teachers' TSR-positivity and students' and teachers' TSR-negativity to the outcome to be equal to 0. We find evidence that this set of latent factors is associated with each outcome, even after controlling for students' TSR-positivity. This analysis suggests that failing to account for teachers' perceptions and students' TSR-negativity could prevent researchers from fully understanding how TSRs are associated with outcomes. In Table 6, we show that for several outcomes the proportion of variability explained by the full model is substantially larger than the proportion of variability explained by the reduced

Table 5

Comparison of fitted regression coefficients between full and reduced models for student outcomes.

	Reduced model	Full model			
	Student TSR positivity (SE)	Student TSR positivity (SE)	Teacher TSR positivity (SE)	Student TSR negativity (SE)	Teacher TSR negativity (SE)
Grades	2.01*** (0.44)	– 0.31 (0.96)	1.40** (0.48)	– 1.39 (1.17)	– 1.98** (0.64)
Contribution quality	0.15*** (0.05)	0.01 (0.05)	0.58*** (0.07)	0.15 (0.09)	0.01 (0.07)
Interest	0.59*** (0.05)	0.66*** (0.08)	0.05 (0.06)	0.17* (0.08)	– 0.05 (0.75)
Sense of belonging	0.30*** (0.04)	0.16 (0.10)	0.03 (0.10)	– 0.18 (0.11)	0.00 (0.13)
Class participation	0.20*** (0.05)	0.08 (0.08)	0.83*** (0.08)	0.18* (0.09)	0.37*** (0.10)
Homework submission	2.30*** (0.60)	– 0.05 (1.24)	3.63** (1.23)	0.54 (1.36)	– 3.57** (1.16)
Effort	0.57*** (0.05)	0.40*** (0.07)	0.31*** (0.08)	– 0.11 (0.09)	0.18 (0.09)
Self-efficacy	0.53*** (0.05)	0.42*** (0.10)	0.21** (0.08)	– 0.10 (0.12)	0.17 (0.11)

Note: All regression coefficients for outcomes other than grades and homework submission can be interpreted as the predicted standard deviation difference in the outcome associated with a one standard deviation difference in the latent factor. The coefficients for grade and homework submission can be interpreted as the difference in grades or percentage of homework returned, on a 100 point scale, associated with a one standard deviation difference in the latent factor. All models control for student and teacher gender and student race (White or non-White). Standard errors are reported in parentheses.

* $P < 0.05$.

** $P < 0.01$.

*** $P < 0.001$.

model.

3.4. To what extent do the patterns of associations differ between the reduced and full models?

3.4.1. Academic

Contrary to what the reduced model suggests, students' TSR-positivity was not associated with their grades (see Table 5) when also accounting for student TSR-negativity, and teacher TSR-positivity and -negativity. Instead, our full-model shows large, significant associations between both teachers' TSR-positivity and TSR-negativity and student grades. Controlling for the other variables in the model, a one standard-deviation difference in teachers' TSR-positivity predicted a 1.4 percentage-point difference in grades (on a 0 to 100 scale), while a one-unit difference in teachers' TSR-negativity predicted a nearly – 2 percentage-point difference in grades.

For the quality of students' classroom contributions, and again

Table 6

Comparison of R^2 between full and reduced models for student outcomes.

	Grades	Contribution quality	Interest	Sense of belonging	Class participation	Homework submission	Effort	Self-efficacy
R^2 (reduced model)	.11	.04	.39	.17	.14	.11	.37	.31
R^2 (full model)	.20	.15	.41	.18	.48	.24	.41	.33
ΔR^2	.09	.11	.02	.01	.34	.13	.04	.02
p -Value*	< .001	< .001	.003	.03	< .001	< .001	< .001	.04

Note: All R^2 differences represent the difference between a 6-predictor, “full” model that uses all 4 TSR sub-scales and a 3 predictor, “reduced” model that uses only students' perceptions of TSR-positivity.

* p -Value associated with constraining teachers' TSR-positivity and students' and teachers' TSR-negativity to 0 in the full model.

unlike the reduced model, teachers' TSR-positivity was the only predictor to reach significance.

3.4.2. Affect

As shown in Table 5, students' TSR-positivity was associated with their interest in the class. Surprisingly, higher values of students' TSR-negativity also predict higher levels of interest. However, because the confidence interval for this association approaches 0, this may be a chance finding.

Despite the reduced model showing an association between students' TSR-positivity and students' sense of belonging, the full model showed no such association with the outcome. However, the Wald test suggested that at least one of the regression coefficients for the TSR variables was non-zero ($t_4 = 77.8$, $p < 0.001$), even if we could not determine which one.

3.4.3. Behavior

In the full model, teachers' TSR-positivity had a strong association with students' class participation. Students' and teachers' TSR-negativity were also associated with class participation, though in an unexpected direction (see Table 5).

As Table 5 also shows, teachers' felt their relationships were significantly more positive and less negative with students who completed more homework.

3.4.4. Motivation

The more that students' felt positively about their relationship with their teachers, the harder they reported trying in class. Similarly, the more teachers felt as though they had a positive relationship with their students, the more effort the students reported putting forth in class. Both of these associations were moderately strong (see Table 5).

Likewise, the more positively students and teachers felt about their relationships, the more efficacious students felt about their class. Once again, both associations were moderately strong, though the estimated effect size of students' TSR-positivity (0.42) was twice the size of that of teachers' TSR-positivity.

4. Discussion

The relationship between teachers and students remains one of the most important aspects of adolescents' schooling experience (Pianta & Allen, 2008). As a consequence, having a precise understanding of how TSRs function, how they relate to student outcomes of interest, how they change over time, and how they might be improved through interventions are of great interest to researchers and practitioners (e.g., Gehlbach et al., 2016). However, developing a comprehensive understanding of TSRs requires measuring these relationships in different ways to illuminate different features of these relationships. At the secondary level, researchers frequently assess these relationships by asking students for their perceptions of specific aspects of TSRs, e.g., teacher caring (Wentzel, 1997), supportiveness (Goodenow, 1993), etc. This approach has produced a voluminous literature on TSRs that

illustrates how consistently these relationships are associated with a host of vital student outcomes (Roorda et al., 2011).

At the same time, the approach has left important questions unanswered: Whose perspective—teachers' or students'—of the TSR matters more for which outcomes? Are the positive and negative aspects of TSRs two ends of the same continuum or are they different dimensions? Are student outcomes predicted in different ways when one accounts for the overall relationship as opposed to discrete aspects? These questions (and many others) require a new, complementary approach to measuring this important construct.

To enable investigation of these and other important questions about TSRs, we focused on the overall relationship, emphasized the distinct perceptions of teachers and students, and allowed the positive and negative aspects of the relationships to be independent. Through a six-step survey design process, we developed a new measure of TSRs that attempted to marshal evidence of the scale's (content and substantive) validity from the outset of the scale development process. After creating the scale, we obtained additional evidence of validity, including: the factor structure of this measure, reliability estimates for the scale, and measurement invariance between middle and high school students. Finally, we showed how the scale functioned largely as would be predicted by theory with respect to expected developmental declines in the health of TSRs and gender differences in teachers' perceptions.

However, ultimately the test of whether this new approach to assessing TSRs proves a worthy complement to traditional approaches, depends more on how it predicts student outcomes—particularly whether it predicts these outcomes differently than previous approaches. We found that our approach explains more variability in student outcomes than our proxy of a more traditional approach. Most importantly, we find different patterns of associations between TSRs and student outcomes. Identifying these different patterns of associations sharpens our understanding of how TSRs are associated with student outcomes by shedding light on whose perception matters (the teachers' or the students') and whether the positivity or negativity of the relationship appears to drive the correlation. Discussion of these findings could address a number of different points. We focus on those that we think are most important for researchers and practitioners who need to assess TSRs and are deliberating over whether our new approach or a more traditional method makes the most sense.

4.1. Validity of the scale

Perhaps the most important question facing a new scale is the extent to which it shows evidence of construct validity. Because the establishment of validity is a process rather than an end state (Gehlbach, 2015), a new scale should offer a foundation of evidence that suggests that it measures what it purports to. However, this foundation will inevitably be developed further over time by other scholars conducting other studies. Between the scale development process and the results of the data collection, we are confident that our scale provides a solid base of evidence to warrant continued use by researchers and practitioners.

Past work has demonstrated that TSRs can be conceptualized and measured as discrete aspects of these relationships (e.g., Murdock, 1999; Pianta, 2001; Wentzel, Battle, Russell, & Looney, 2010). The factor structure that best fit our data suggests that measuring TSRs as an omnibus construct can also work for secondary school teachers and students. However, in our case the positive and negative dimensions of the relationship emerged as separate factors. This structure accords with recent meta-analytic findings and the suggestion that, “future research, and especially secondary school studies, should consider including negative aspects of the relationship” (Roorda et al., 2011, p. 519). An important future research direction is how many factors would emerge if an array of discrete aspects of TSRs were measured—e.g., conflict, closeness, dependency, academic support, expectations, social support, etc.

In addition, our data provide strong evidence of internal consistency

for these scale scores and evidence of moderate stability in TSR scores over the course of the school year. We interpret the practical importance of the measurement invariance results as evidence that the scales function similarly enough across middle and high school to warrant use in longitudinal studies. Finally, by largely replicating prior findings regarding grade-level and gender differences in TSRs, we find further evidence that the scales are functioning as anticipated.

As scholars conduct future studies using this scale, more validity evidence will be acquired. Through this process, we will learn the range of appropriate uses for this approach to assessing TSRs and the bounds of those uses. For now, we are confident that this measure shows sufficient evidence of validity to warrant adoption by future researchers interested in TSRs.

4.2. Utility of the new measure

Even if a scale shows compelling evidence of validity, it may not prove particularly useful in applied settings if it does not generate new insights about the construct when compared against existing approaches. We find that our full models explain more variability in each of the eight outcomes of interest and illuminate different patterns of associations as compared to our proxy for a more traditional approach.

Through our reduced model, we replicate (conceptually) many past investigations of TSRs—studies that assessed TSRs by asking students about specific, positive aspects of their relationship with their teacher and then correlated those student perceptions with student outcomes. For example, our reduced model mirrors Goodenow's (1993) finding that students who felt more support from their teachers got better grades; Wentzel's (1998) findings that students who perceived more teacher support were more interested in their class and got better grades; and Murdock and Miller's (2003) results that student perceptions of teacher caring predicted students' self-efficacy and effort. These parallels provide some assurance that our sample functioned similarly to samples in past investigations of TSRs and student outcomes.

In comparing this reduced model to our full model, it is also important to note that including all four TSR sub-scales increases the proportion of variability explained in our final models by a practically meaningful, as well as a statistically significant, amount for multiple outcomes. As Table 6 shows, the full models predict far more variation than the reduced models in student grades, contribution quality, class participation, and homework completion. Perhaps these models explain so much additional variability because the outcomes are directly affected by teachers and students to a greater degree than other outcomes. For instance, grades typically result from the students' efforts and teachers' subject evaluations of those efforts; class participation frequently requires students to raise their hands and teachers to call on them. By contrast, outcomes such as students' interest in the subject matter or sense of belonging are internal psychological states that teachers are likely to affect only indirectly.

In sum, we find that our approach to measuring TSRs offers greater overall explanatory power. Correspondingly, it seems plausible that studies conceptualizing TSRs more narrowly may underestimate its importance for some student outcomes. For example, if students' perceptions of how much a teacher respects and encourages them are the key elements in the association between TSRs and student effort, but a study measures TSRs using only teacher expectations, the correlation between TSRs and student effort might be underestimated (relative to using a measure of the overall TSR).

In addition to greater explanatory power, our approach potentially provides a more precise understanding of the nature of the associations between TSRs and student outcomes. For only three of the eight model comparisons did students' TSR-positivity remain a statistically significant predictor. Meanwhile, teachers' TSR-positivity was significantly associated with the outcomes in six out of eight cases. Both students' and teachers' TSR-positivity were significant in two out of the eight cases. We view these results as compelling evidence that scholars

should account for both teacher and student perspectives in future investigations of TSRs in order to obtain a more nuanced association of how TSRs relate to various outcomes.

We also find evidence for Roorda et al.'s (2011) suggestion that accounting for the positive and negative dimensions of these relationships separately may be important. Teachers' TSR-negativity was a significant predictor of three outcomes. Students' TSR-negativity predicted two—students' participation and interest in class. Although for participation, both students' and teachers' TSR-negativity scores showed evidence of unexpected suppressor effects (Rosenberg, 1968), thus, we are reticent to interpret this finding until future studies can demonstrate that it is more than idiosyncratic.

This sharper understanding of how TSRs are associated with student outcomes can also serve to generate important hypotheses. For instance, perhaps teachers' (as compared to students') perceptions of TSRs are more predictive of classroom outcomes in general because they (typically) hold more power within the relationship. Others might speculate that TSR-positivity better predicts "approach" outcomes such as effort but outcomes such as class participation—where students might raise their hands aggressively (an approach behavior) or hide behind a classmate (an avoidance behavior)—will typically be better predicted by TSR-positivity and negativity subscales (Elliot, 2006).

This approach to measuring and understanding TSRs may also implicate the actions taken by school leaders. This consequence is perhaps best illustrated by the associations between TSRs and grades. If a school leader wanted to improve students' grades, prior correlational studies (Goodenow, 1993; Wentzel, 1998) signal that an intervention which bolsters the positivity of students' perceptions of their relationship with their teachers might be worth testing. By contrast, our findings suggest that students' grades are more likely to go up by intervening directly with teachers to help them see their students in a more positive and/or less negative light. While researchers are (rightly) quick to point out that correlations like those between TSRs and student outcomes are not necessarily causal, school leaders have to make educated guesses about where and how to improve student outcomes (often based on correlational evidence). As the associations between TSRs and grades indicate, even correlational data from this approach to measuring TSRs can provide school leaders with clues about where and how they might want to intervene (e.g., with teachers rather than students).

4.3. Limitations and trade-offs involved in our TSR scale

In the development process, we incorporated several additional assets into our scale. The scale is short—a particularly important quality for teachers who might have to fill out the scale repeatedly. Because we demonstrate evidence of invariance between middle and high school respondents, the scale should work well for developmental research between 6th and 12th grade.

However, we necessarily had to make certain trade-offs. One cost of keeping our scale short, was that our 9-item and 5-item sub-scales do not directly represent every indicator listed in Table 1. Second, taking a holistic approach is particularly useful for answering many research questions. For other research questions taking a thinner slice of TSRs (e.g., a specific scale on teacher caring or perceived criticism) may be more appropriate. Third, we found less variability in our negativity scale than we might have liked. In particular, teachers seemed reticent to admit that they had especially negative relationships with any of their students (perhaps because a negative TSR could reflect poorly on their abilities to connect with students). However, items that we might

have otherwise reworded to get more variability had other merits. For example, few teachers responded to the question of "How unfair are you to < student's name > in class?" by using the *somewhat*, *quite*, or *extremely unfair* response anchors. Ultimately, we retained the item because the students' scores showed a much broader range. We felt that examining discordance around perceptions of fairness might be especially interesting and important for future scholars to study.

Other limitations might be addressed through future studies. For instance, including more measures from other sources (e.g., relatively objective outcomes such as test-scores and attendance and subjective measures from other sources like parent reports of their children's TSRs) will be particularly useful in providing additional tests of convergent and discriminant validity. We might also learn more about how these relationships function by having students complete this measure on multiple teachers. Such an approach would shed light on respective levels of within-student, between-teacher variance as compared to between-student, within-teacher variance in these relationships. Future studies that can document TSRs from samples who do not volunteer to participate will help with issues of representativeness—we worried that relying on volunteers might leave us with a sample of teachers (and perhaps students as well) who had disproportionately positive TSRs and/or were more interested than most teachers in working on their TSRs.

We remain hopeful that as future researchers and school leaders use this TSR scale, they may develop modifications that minimize the need to make these types of compromises and trade-offs. In addition, we hope that they might explore important issues that we could not such as whether TSRs vary by subject area, type of school, or other important factors.

5. Conclusion

We are particularly optimistic about the coming years of research on TSRs. Many exciting questions are still open for investigation: To what extent can strengths in one area of a TSR compensate for weaknesses in another area? What are the downstream consequences when students' and teachers' perceptions of their TSR are misaligned? Perhaps the most important question of all is, whether interventions that improve TSRs cause improvements in other student outcomes.

Relationships between teachers and students lie at the heart of the learning environment. Because classrooms are fundamentally social contexts (Gehlbach, 2010), improving these crucial relationships seems like a particularly promising approach to improving student outcomes. However, before scholars can test different approaches to understanding and ultimately improving TSRs, they must measure these relationships accurately. We hope that measures like this one can contribute to these new understandings and improvements.

Acknowledgments.

This research was supported by generous funding from the Harvard Graduate School of Education Dean's Venture Fund, a grant from the Spencer Foundation and an Early Career Research Award by Division 15 of the American Psychological Association. The authors are particularly grateful for the generous assistance of the participating schools and the help of Rebecca Zazove in collecting these data. We are also indebted to the scholars who participated in our expert review. Most of all, we are extremely appreciative of the students, teachers, and school principals who made this study happen.

Appendix A. Teacher-Student Relationship Scale

Student items	Teacher items
Positivity sub-scale	
1 How much do you enjoy learning from <teacher's name> ?	How much do you enjoy helping <student's name> learn?
2 How friendly is <teacher's name> towards you?	How friendly is <student's name> towards you?
3 How often does <teacher's name> say something encouraging to you?	How often do you say something encouraging to <student's name> ?
4 How respectful is <teacher's name> towards you?	How respectful is <student's name> towards you?
5 How excited would you be to have <teacher's name> again next year?	How excited would you be to have <student's name> again next year?
6 How motivating are the activities that <teacher's name> plans for class?	How motivating does <student's name> find the activities that you plan for class?
7 How caring is <teacher's name> towards you?	How caring is <student's name> towards you?
8 How much do you like <teacher's name> 's personality?	How much do you like <student's name> personality?
9 Overall, how much do you learn from <teacher's name> ?	Overall, how much does <student's name> learn from you?
Negativity sub-scale	
1 How often do you ignore something <teacher's name> says?	How often does <student's name> ignore something you say?
2 During class, how often do you talk when <teacher's name> is talking (for instance, when you are supposed to be listening)?	During class, how often does <student's name> talk when you are talking (for instance, when <student's name> is supposed to be listening)?
3 How often does <teacher's name> say something that offends you?	How often do you say something that offends <student's name> ?
4 How unfair is <teacher's name> to you in class?	How unfair are you to <student's name> in class?
5 How angry does <teacher's name> make you feel during class?	How angry do you make <student's name> feel during class?

Note: Response anchors were arrayed along five points. For example: Not at all/Slightly/Somewhat/Quite a bit/A tremendous amount; Not at all friendly/Slightly friendly/Somewhat friendly/Quite friendly/Extremely friendly; Almost never/Once in a while/Sometimes/Frequently/Almost all the time; or Almost nothing/A little bit/Some/Quite a bit/A great deal.

Appendix B. Student outcome measures*Academic*

- 1) Student grades: Collected at Schools 1, 3, and 4
 - a. School 1 – student self-report
 - b. School 3 – teacher reported
 - c. School 4 – teacher reported
- 2) Contribution quality: Teacher ratings of the quality of students' class participation: Collected at Schools 2, 3, and 4
When *Student X* participates in class, how would you rate the quality of his/her contributions?
Response anchor:

Far below average	Below average	Average	Above average	Far above average
-------------------	---------------	---------	---------------	-------------------

Affect

- 3) Interest in subject matter: Collected at Schools 2 and 4
How interesting do you find your ____ class?
Response anchor:

Not at all interesting	Slightly interesting	Somewhat interesting	Quite interesting	Extremely interesting
------------------------	----------------------	----------------------	-------------------	-----------------------

If you could choose to take any classes you wanted to in high school, how many classes would you take in this subject?
Response anchor:

No classes	A few classes	Some classes	Quite a few classes	A lot of classes
How likely are you to go into a _____-related career? Response anchor:				
Not at all likely	Slightly likely	Somewhat likely	Quite likely	Extremely likely
4) Sense of belonging: Collected at all 4 schools. I feel like I belong in this school. I feel like I am successful in this school. I feel like I matter in this school. I do not feel like I am important in this school. (reversed) Response anchor:				
Not at all true of me	Slightly true of me	Somewhat true of me	Quite true of me	Extremely true of me

Note: Scale taken from [Roeser et al. \(1996\)](#).

Behavior

5) Homework completion: Teacher reports of percentage of homework completed; collected at all 4 schools.
Approximately what percentage of the assigned homework does *Student X* complete fully? ____
6) Teacher ratings of frequency of class participation: Collected at Schools 2, 3, 4
How frequently does *Student X* participate in class?
Response anchor:

Almost never	Once in a while	Sometimes	Frequently	Almost all the time
-----------------	--------------------	-----------	------------	------------------------

Motivation

7) Students' self report of their *effort*: Collected at Schools 2, 3, 4
How much effort do you put forth for this class?
When *Teacher X* is speaking, how much effort do you put into trying to pay attention?
How much effort do you put into getting involved in class discussions?
How much effort do you put into your homework for this class?
How much effort do you put into in-class activities?
Response anchor:

Almost no effort	A little bit of effort	Some effort	Quite a bit of effort	A great deal of effort
8) Students' self report of their <i>self-efficacy</i> : Collected at Schools 2, 3, 4. How confident are you that you can do the hardest work that is assigned in this class? How confident are you that you can learn all the material presented in this class? When complicated ideas are presented in this class, how confident are you that you can understand them? How confident are you that you will remember what you learned in this class next year? How confident are you that you can complete all the work that is assigned in this class? Response anchor:				
Not at all confident	Slightly confident	Somewhat confident	Quite confident	Extremely confident

Note: Scale adapted from ([Gehlbach et al., 2008](#)).

References.

- Ang, R. P. (2005). Development and validation of the Teacher-Student Relationship Inventory using exploratory and confirmatory factor analysis. *Journal of Experimental Education*, 74(1), 55–73. <http://dx.doi.org/10.3200/jexe.74.1.55-74>.
- Asparouhov, T. (2005). Sampling weights in latent variable modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 12(3), 411–434. http://dx.doi.org/10.1207/s15328007sem1203_4.
- Gehlbach, H. (2010). The social side of school: Why teachers need social psychology. *Educational Psychology Review*, 22(3), 349–362. <http://dx.doi.org/10.1007/s10648-010-9138-3>.
- Gehlbach, H. (2015). Seven survey sins. *The Journal of Early Adolescence*, 35, 883–897. <http://dx.doi.org/10.1177/0272431615578276>.
- Gehlbach, H., Brown, S. W., Ioannou, A., Boyer, M. A., Hudson, N., Niv-Solomon, A., & Janik, L. (2008). Increasing interest in social studies: Social perspective taking and self-efficacy in stimulating simulations. *Contemporary Educational Psychology*, 33(4), 894–914. <http://dx.doi.org/10.1016/j.cedpsych.2007.11.002>.
- Gehlbach, H., & Brinkworth, M. E. (2011). Measure twice, cut down error: A process for enhancing the validity of survey scales. *Review of General Psychology*, 15(4), 380–387. <http://dx.doi.org/10.1037/a0025704>.
- 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. <http://dx.doi.org/10.1037/edu0000042>.
- Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, 115(3), 401–423. <http://dx.doi.org/10.1037/0033-2909.115.3.401>.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464–504.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233–255.
- Chun, H., & Mobley, M. (2010). Gender and grade-level comparisons in the structure of problem behaviors among adolescents. *Journal of Adolescence*, 33(1), 197–207. <http://dx.doi.org/10.1016/j.adolescence.2009.03.010>.
- Clark, M. S., & Lemay, E. P., Jr. (2010). Close relationships. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Vol. Eds.), *Handbook of social psychology* (5th ed.). Vol. 2. *Handbook of social psychology* (pp. 898–940). Hoboken, NJ US: John Wiley & Sons Inc.
- Cornelius-White, J. (2007). Learner-centered teacher-student relationships are effective: A meta-analysis. *Review of Educational Research*, 77(1), 113–143. <http://dx.doi.org/10.3102/003465430298563>.
- DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Newbury Park, CA: Sage.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th ed.). Hoboken, NJ: John Wiley & Sons.
- Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Mac Iver, D. J. (1993). Development during adolescence: The impact of stage-environment fit on young adolescents' experiences in schools and in families. Special issue: Adolescence. *American Psychologist*, 48(2), 90–101. <http://dx.doi.org/10.1037/0003-066X.48.2.90>.
- Elliot, A. J. (2006). The hierarchical model of approach-avoidance motivation. *Motivation and Emotion*, 30(2), 111–116.
- Gable, S. L., Reis, H. T., & Downey, G. (2003). He said, she said: A quasi-signal detection analysis of daily interactions between close relationship partners. *Psychological Science*, 14(2), 100–105.
- Garza, R. (2009). Latino and white high school students' perceptions of caring behaviors: Are we culturally responsive to our students? *Urban Education*, 44(3), 297–321.
- Goodenow, C. (1993). Classroom belonging among early adolescent students: Relationships to motivation and achievement. *The Journal of Early Adolescence*, 13(1), 21–43. <http://dx.doi.org/10.1177/0272431693013001002>.
- Gregory, A., & Weinstein, R. S. (2008). The discipline gap and African Americans: Defiance or cooperation in the high school classroom. *Journal of School Psychology*, 46(4), 455–475. <http://dx.doi.org/10.1016/j.jsp.2007.09.001>.
- Hamre, B. K., & Pianta, R. C. (2001). Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development*, 72(2), 625–638.
- Hamre, B. K., & Pianta, R. C. (2006). Student-teacher relationships. In G. G. Bear, & K. M. Minke (Eds.), *Children's needs III: Development, prevention, and intervention* (pp. 59–71). Washington, DC US: National Association of School Psychologists.
- Hoynoski, B., Link, M., & Frankel, M. (2009). *Measuring total participation: An alternative metric to response rate*. Paper presented at the 64th Annual Conference of the American Association for Public Opinion Research, Hollywood, FL.
- Johnson, B. (2008). Teacher-student relationships which promote resilience at school: a micro-level analysis of students' views. *British Journal of Guidance and Counselling*, 36(4), 385–398.
- Juvonen, J. (2006). Sense of belonging, social bonds, and school functioning. In P. A. Alexander, & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 655–674). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.
- Loftus, E. F. (2003). Make-believe memories. *American Psychologist*, 58(11), 867–873. <http://dx.doi.org/10.1037/0003-066X.58.11.867>.
- Maehr, M. L. (1976). Continuing motivation: An analysis of a seldom considered educational outcome. *Review of Educational Research*, 46(3), 443–462.
- Martin, A. J., & Dowson, M. (2009). Interpersonal relationships, motivation, engagement, and achievement: Yields for theory, current issues, and educational practice. *Review of Educational Research*, 79(1), 327–365. <http://dx.doi.org/10.3102/0034654308325583>.
- McKenzie, J. F., Wood, M. L., Kotecki, J. E., Clark, J. K., & Brey, R. A. (1999). Establishing content validity: Using qualitative and quantitative steps. *American Journal of Health Behavior*, 23(4), 311–318.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749. <http://dx.doi.org/10.1037/0003-066X.50.9.741>.
- Midgley, C., Maehr, M. L., Huda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., ... Urdan, T. (2000). Patterns of adaptive learning study. Retrieved from http://www.umich.edu/~pals/PALS%202000_V13Word97.pdf.
- Muller, C. (2001). The role of caring in the teacher-student relationship for at-risk students. *Sociological Inquiry*, 71(2), 241–255.
- Murdock, T. B. (1999). The social context of risk: Status and motivational predictors of alienation in middle school. *Journal of Educational Psychology*, 91(1), 62–75. <http://dx.doi.org/10.1037/0022-0663.91.1.62>.
- Murdock, T. B., Anderman, L. H., & Hodge, S. A. (2000). Middle-grade predictors of students' motivation and behavior in high school. *Journal of Adolescent Research*, 15(3), 327–351. <http://dx.doi.org/10.1177/0743558400153002>.
- Murdock, T. B., & Miller, A. (2003). Teachers as sources of middle school students' motivational identity: Variable-centered and person-centered analytic approaches. *The Elementary School Journal*, 103(4), 383–399. <http://dx.doi.org/10.1086/499732>.
- Murray, C., & Pianta, R. C. (2007). The importance of teacher-student relationships for adolescents with high incidence disabilities. *Theory Into Practice*, 46(2), 105–112.
- Pianta, R. C. (1999). *Enhancing relationships between children and teachers*. Washington D.C.: American Psychological Association.
- Pianta, R. C. (2001). *Student-teacher Relationship Scale (STRS): Professional manual*. Lutz, FL: Psychological Assessment Resources.
- Pianta, R. C., & Allen, J. P. (2008). Building capacity for positive youth development in secondary school classrooms: Changing teachers' interactions with students. In M. Shinn, & H. Yoshikawa (Eds.), *Toward positive youth development: Transforming schools and community programs* (pp. 21–39). Oxford: Oxford University Press.
- Pianta, R. C., Hamre, B. K., & Stuhlman, M. (2003). Relationships between teachers and children. In W. M. Reynolds, & G. E. Miller (Vol. Eds.), *Handbook of psychology: Educational psychology*. 7. *Handbook of psychology: Educational psychology* (pp. 199–234). Hoboken, NJ: John Wiley & Sons, Inc.
- Pornprasertmanit, S., Miller, P., Schoemann, A., & Rosseel, Y. (2013). *semTools: Useful tools for structural equation modeling: R package version 0.4-0*. Retrieved from <http://CRAN.R-project.org/package=semTools>.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>.
- Roeser, R. W., Midgley, C., & Urdan, T. C. (1996). Perceptions of the school psychological environment and early adolescents' psychological and behavioral functioning in school: The mediating role of goals and belonging. *Journal of Educational Psychology*, 88(3), 408–422.
- Roorda, D., Koomen, H., Split, 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. <http://dx.doi.org/10.3102/0034654311421793>.
- Rosenberg, M. (1968). *The logic of survey analysis*. New York: Basic Books.
- Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, 27(2), 94–104.
- Skinner, E., Furrer, C., Marchand, G., & Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? *Journal of Educational Psychology*, 100(4), 765–781. <http://dx.doi.org/10.1037/a0012840>.
- Wentzel, K. R. (1997). Student motivation in middle school: The role of perceived pedagogical caring. *Journal of Educational Psychology*, 89(3), 411–419. <http://dx.doi.org/10.1037/0022-0663.89.3.411>.
- Wentzel, K. R. (1998). Social relationships and motivation in middle school: The role of parents, teachers, and peers. *Journal of Educational Psychology*, 90(2), 202–209.
- Wentzel, K. R. (2002). Are effective teachers like good parents? Teaching styles and student adjustment in early adolescence. *Child Development*, 73(1), 287–301. <http://dx.doi.org/10.1111/1467-8624.00406>.
- Wentzel, K. R., Battle, A., Russell, S. L., & Looney, L. B. (2010). Social supports from teachers and peers as predictors of academic and social motivation. *Contemporary Educational Psychology*, 35(3), 193–202. <http://dx.doi.org/10.1016/j.cedpsych.2010.03.002>.
- Woolley, M. E., & Grogan-Kaylor, A. (2006). Protective family factors in the context of neighborhood: Promoting positive school outcomes. *Family Relations*, 55(1), 93–104. <http://dx.doi.org/10.1111/j.1741-3729.2006.00359.x>.