

What Can Student Perception Surveys Tell Us About Teaching? Empirically Testing the Underlying Structure of the Tripod Student Perception Survey

Tanner LeBaron Wallace
University of Pittsburgh, Pittsburgh
Benjamin Kelcey
University of Cincinnati
Erik Ruzek
University of Virginia

We conducted a theory-based analysis of the underlying structure of the Tripod student perception survey instrument using the Measures of Effective Teaching (MET) database (N = 1,049 middle school math class sections; N = 25,423 students). Multilevel item factor analyses suggested that an alternative bifactor structure best fit the Tripod items, and preliminary evidence suggests that both the general responsiveness and the classroom management-specific dimensions are positively associated with teacher value-added scores. In our discussion, we consider the distinct characterizing features of adolescents as raters of teaching, the implications for teacher professional learning opportunities, and key areas for future research.

TANNER LEBARON WALLACE is an associate professor of applied developmental psychology and co-chair of the Motivation Center (MC) at the University of Pittsburgh, 230 S. Bouquet Street, WWPH 5948, Pittsburgh, PA 15260; e-mail: twallace@pitt.edu. She studies adolescents' interpretations of instructional interactions and how these affect learning outcomes.

BENJAMIN KELCEY is an associate professor in the Quantitative Research Methodologies Program at the University of Cincinnati. His research interests focus on causal explanation and inference and measurement methods within the context of classrooms and schools.

ERIK RUZEK is a research assistant professor at the Center for Advanced Study of Teaching and Learning at the University of Virginia. His research integrates information on students and teachers from different measurement traditions, including classroom observations, student reports, teacher reports, and direct assessments. His work combines data from these different sources to describe how students' motivation, engagement, and learning are influenced by their experiences in classrooms.

KEYWORDS: multilevel confirmatory factor analysis, student perceptions, teacher evaluation

Advancing theories of effective teaching on the basis of adolescent meaning-making (within the context of a teacher evaluation system) necessitates several acts of translation. First, adolescents' daily experiences of exposure to particular teacher actions that affect student outcomes must be translated into high-quality survey items. Then, translation must proceed from the data generated from those items into informed individualized professional learning opportunities for teachers. Indeed, policy-relevant findings and guidelines for teaching are predicated on issues of validity—actually knowing the specific nature of what is being measured—so that clear practices and policies to support the improvement of teaching can be designed and implemented. It is well known that “scores averaged across an ill-defined assortment of items offer no basis for knowing what is being measured” (Marsh & Roche, 1997, p. 1187). Understanding how theories of effective teaching should or should not inform the selection, use, and improvement of teacher evaluation measures, such as student surveys, has important implications for practice given that in the present accountability context, measurement incentivizes how teachers teach. Underdeveloped measures may lead to ill-informed pedagogical choices.

As of publication, 45 U.S. states had developed statewide teacher evaluation systems. Common reform features include the use of multiple measures of teacher performance (e.g., observational instruments, students' standardized test scores, teacher portfolios) as well as teacher rating scales that denote a range of distinct performance levels. While variability exists in the details of what data are gathered and how those data are used to model student growth, a noteworthy and recent addition to the assessment of effective secondary teaching is the inclusion of student perception surveys.

In the present paper we use data from the large, U.S.-based Measures of Effective Teaching (MET) database to examine the underlying structure of the Tripod student perception survey (Ferguson, 2010). In part because of the publicity garnered by the MET project, the Tripod student perception survey has become among the most popular student perception surveys on the U.S. market. We estimate that during the 2015–2016 school year approximately 1,400 schools administered the Tripod. The 36-item Tripod survey is organized around seven distinct theoretical domains of classroom instruction, or “the 7Cs.” The components of effective teaching that the Tripod purports to measure (e.g., care, challenge, confer) are central to many theoretical models of effective teaching and effective teachers, such as the concepts of warm demander (Kleinfeld, 1975; Ware, 2006) and culturally responsive teaching (Gay, 2010).

Two decades ago, Marsh and Roche (1997) argued that student evaluation of teaching instruments failed to provide an ample assessment of

theoretically sound, multiple dimensions of teaching quality, in ways that undermined the use of such instruments to inform diagnostic feedback. As such, well-articulated theories of effective teaching, such as Ferguson's (2010) 7Cs, productively inform building an evidence base around the underlying explanatory mechanisms that link teacher action with adolescent outcomes. More specifically, attending to adolescent perceptions as a central aspect of theory-building acknowledges adolescents as active co-constructors of learning ecologies with the agency to, as Nakkula and Toshalis (2006) describe, rightfully resist educational experiences they perceive to be inauthentic or unsafe.

In 2014, Schweig took an atheoretical, exploratory approach to examining the multilevel factorial structure of the Tripod using results from an administration of the survey to California students ($N = 6,386$ students; $N = 349$ class sections) in one urban school district during the 2010 school year. Results from that exploratory analysis suggested "five factors at the within level, and two factors at the between level" (pp. 270–271). Our study extends this and other prior research on the Tripod instrument by empirically testing the underlying structure of the Tripod student perception survey through the lenses of predominant theories of effective teaching. This investigation is made possible due to the fact that despite Tripod items being developed in relation to a specific theory of effective teaching, the items cover a diverse range of classroom experiences described in rather general terms. Therefore, student endorsements of Tripod items can help us learn about student assessments of effective teaching more broadly. Our investigation brings to bear plausible and theoretically supported factor structures concerning the Tripod instrument by considering multiple theoretical factor structures and the extent to which there is comparative and absolute support for each structure across multiple school districts.

Student Perceptions of the Learning Environment

Prior research has demonstrated that student perceptions of the learning environment can be both reliable (e.g., Fauth, Decristan, Rieser, Klieme, & Büttner, 2014; Wagner, Göllner, Helmke, Trautwein, & Lüdtke, 2013) and predictive of learning (e.g., Kane & Cantrell, 2010; Kane & Staiger, 2012). Broadly speaking, student perceptions of the learning environment are likely indicative of the motivational aspects of classrooms (McCaslin, 2009; Spearman & Watt, 2013) for the fact that student perceptions of a classroom environment, constituted in youth-adult interactions, are the primary mechanism through which adolescents assent-to-learn in classrooms (Brophy & Good, 1974; Erickson, 1987; Erickson et al., 2007).

If students do not form a positive connection with their teacher, it is within their control to minimally learn core content or refuse to learn anything at all (Wallace & Chhuon, 2014). Thus, interactions between a student

and his or her teacher are the source material for student-teacher connections, and these connections can strongly influence student learning (Davis, 2003). Given the importance of student perceptions of the learning environment, student reports of the quality of interactions and processes within a classroom are potentially an important measurement strategy for evaluating and developing teachers. Beyond providing firsthand impressions of the quality of student-teacher interactions and classroom processes, student reports offer a promising way to evaluate teaching because they can be used to measure theoretically informed and practically important dimensions of instruction; draw on the perspectives of multiple students, making survey measures potentially more reliable; and provide more efficient assessments of quality compared with alternative, resource-intensive assessments such as classroom observations (Turner & Meyer, 2000).

Students as Raters of Teaching Quality

Despite the advantages of student surveys, student raters are fundamentally different than the trained adult observers who use formal observational instruments to gather data relevant to teaching quality. Prior research has found low to moderate agreement between teacher ratings and student ratings of instruction (e.g., Desimone, Smith, & Frisvold, 2010; Kunter & Baumert, 2006). Unlike adult observers who undergo rigorous training and certification processes to establish their skill at consistently differentiating among complicated, theoretically proposed domains of instructional practice, students receive no training prior to data collection. As such, the implicit conceptualization of expertise is quite distinct between adult observers and student survey takers.

While adult observers gain expertise by mastering and applying the content of adult-created frameworks of effective teaching, student observers possess naturally acquired expertise through their lived, everyday experiences in classrooms. For example, students are often experts in “contingencies of their social environments” (McCaslin & Hickey, 2001, p. 137) and, therefore, likely attune to the nuances of peer interaction that are conditioned indirectly, but powerfully, by a teacher’s instructional practice (Gest & Rodkin, 2011). Classrooms, by their very nature, are participatory, and it is probable that ongoing experiences of participation shape assessments of teaching quality in ways that are fundamentally distinct from a nonparticipant, outside observer.

Likewise, students’ firsthand experiences of participation in instructional activities situate adolescents’ assessments of teaching quality as originating primarily from the perspective of a learner versus that of the teacher. From a situative perspective, learning may best be characterized as the development of more effective participation in the practices of inquiry and discourse around collaborative sense-making and reasoning (Greeno,

1998). Thus, adolescents may be particularly sensitive to the affordances and constraints that support their effective participation, particularly interactions and experiences that are most salient to the continual renegotiation of resources for learning. These sensitivities to teachers' distribution of participation opportunities (i.e., framing of who gets to participate) and the meaningfulness of the work students are being asked to complete may influence how an adolescent rater interprets or makes sense of the significance of instructional interactions and subsequently endorses survey items.

Theories of Effective Teaching

Teaching is a complex and multidimensional practice with overlapping and simultaneous teacher actions occurring continuously over time (Carlisle, Kelcey, Berebitsky, & Phelps, 2011). Like practitioners in other clinical practices, such as counseling or medicine, teachers must enact specific competencies in real time under dynamic and evolving contexts. Teaching, however, is unique from other clinical practices in that it takes place publicly in front of up to 30 developing individuals. Moreover, teaching is a practice aimed at promoting positive development among a diverse set of persons entering at different starting points with unique histories both within school and out of school.

Within this context of complexity, theoretical models of teaching have emerged to summarize and advance particular conceptualizations of effective teaching. One aspect of teacher evaluation systems that often goes unchallenged is how particular theories of effective teaching are instantiated in the instruments used in such systems. At present, four alternative multidimensional models of effective teaching are prominent in the evaluation instruments used by many districts in the United States. As noted earlier, Tripod items were developed on the basis of one specific conceptualization of effective teaching. Yet, these items cover a diverse range of experiences in classrooms described in rather general terms. Thus, testing the underlying structure of the Tripod student perception survey instrument permits testing potential alternative conceptualizations of teaching from the student perspective. Below we review prominent conceptualizations of effective teaching informing our investigation of the underlying structure of the Tripod instrument.

One theoretically proposed model of effective teaching comprises a two-dimensional structure with factors representing academic press and social support for learning (Ferguson & Danielson, 2014; Lee & Smith, 1999; Lee, Smith, Perry, & Smylie, 1999). Academic press is defined as the "normative emphasis on academic success" (Lee et al., 1999, p. 10), which motivates a rigor and focus on fostering conditions so that children can learn what is expected of them by the standards of society (Ferguson & Danielson, 2014). Social support for learning is defined as the instructional supports that

are both relational and social that provide students with a sense of trust, confidence, and psychological safety (Ferguson & Danielson, 2014; Lee et al., 1999).

An alternative, theoretically proposed model of effective teaching comprises a three-dimensional structure with dimensions corresponding to classroom organization, instructional support, and emotional support (Pianta & Hamre, 2009; see also Kunter & Baumert, 2006 for a similar, but differently worded formulation). Classroom organization is a teacher's ability to help students organize their attention and behavior toward academic activities and goals, and strong classroom organization is often materialized in the classroom through clear and consistent routines (Emmer & Strough, 2001). Instructional support distinguishes between the classroom activities and interactions that simply promote fact knowledge versus those that promote an understanding of how facts are organized, conditioned, and interconnected with one another (Mayer, 2002). The quality and nature of teacher-provided feedback are also critical components of instructional support (Kulik & Kulik, 1998). Emotional support has long been understood to be critical to child and adolescent development and motivation and is evidenced by teachers' dependability, their demonstration of genuine concern for and care about students, a desire to understand students' feelings and points of views, and respect for students (Patrick, Anderman, & Ryan, 2004).

Another theoretically proposed model of effective teaching is that which informs the Tripod (Ferguson, 2010) and comprises a seven-dimensional structure. These dimensions include the following: (a) care, (b) confer, (c) captivate, (d) clarify, (e) consolidate, (f) challenge, and (g) control. *Care* describes how a teacher creates interpersonal relationships that support perceptions of emotional closeness and belonging (Ferguson, 2010). Recent conceptualizations of care within the adolescent-adult relationship context have acknowledged the asymmetry of status that exists between a teacher and a student and the importance of understanding care in terms of the identities being ascribed to the student via experiencing the classroom environment in terms of coherency with deeply personal, evolving identities that students hold (Toshalis, 2015; Wallace & Chhuon, 2014). *Confer* describes how a teacher solicits students' points of view and invites students to express themselves (Ferguson, 2010). Support for the importance of student-centered approaches to instructional interactions comes from many sources but perhaps most centrally from autonomy support theory, which prescribes following student leads and incorporating student input into instruction (e.g., Reeve & Jang, 2006). Three dimensions—*captivate*, making material interesting and relevant to students; *clarify*, diagnosing students' particular gaps in knowledge and having multiple ways to explain ideas; and *consolidate*, helping students organize content knowledge in preparation for future learning (Ferguson, 2010)—all describe current conceptualizations of

ambitious pedagogy. For example, Thompson, Windschitl, and Braaten (2013) identify the following four sets of practices: selecting big ideas/models, working with students' ideas, working with disciplinary ideas, and pressing for explanation. High-quality enactments of all of these would entail the features described by the former three Tripod dimensions. *Challenge*, pressing students to work and to think hard, attends to the role of high expectations in student learning (Ferguson, 2010). Investigations of the relation between teachers' expectancies and student academic outcomes have demonstrated weak to moderate associations, suggesting that teacher expectancy behavior is achievement relevant (Friedrich, Flunger, Nagengast, Jonkmann, & Trautwein, 2015; Good, 1987; Noguera, 2003). *Control* describes how a teacher manages her classroom and, in particular, maintains effective communication and focus (Ferguson, 2010). Recent conceptualizations of classroom management have emphasized relational factors such as prioritizing students' perceptions of trust, respect, and connectedness (e.g., Milner & Tenore, 2010), while keeping central a notion of minimizing disruptions so that teachers can teach and students can focus on academic work (e.g., Cartledge, Lo, Vincent, & Robinson-Ervin, 2015). In such a framework, evaluating a teacher's competence in managing a classroom has focused simultaneously on positive concepts like respect and negative concepts like disruptive behavior (Schweig, 2014).

The last proposed model conceptualizes teaching as comprising both general and specific dimensions. Recent research by Hamre, Hatfield, Pianta, and Jamil (2014) suggests that there are generic properties of interactive behavioral exchanges between students and teachers as well as properties that are specific to role, intent, and content of particular instructional interactions. This bifactor-like theory of effective teaching proposes a model wherein a general element of instructional quality related to responsivity influences all teacher-student interactions and, further, a set of domain-specific elements related to motivational supports, classroom management, and cognitive facilitation influence particular interactions dependent upon the specifics of the interaction (Hamre et al., 2014). The general element of instructional quality related to responsivity can also be understood as how successful a teacher is at establishing and maintaining a psychologically safe classroom for a diverse group of students (Wanless, 2016; Williams, Woodson, & Wallace, 2016). Lee's (1992, 2001) cultural modeling theory advances an implicit bifactor-like theory of effective teaching wherein teachers possess a general sensitivity to the importance of culturally based interaction styles but use specific kinds of scaffolding that prepare adolescents for future learning by using prior cultural knowledge as a foundation during instructional interactions. The bifactor-like theory of teaching is also supported by recent research on the importance of teachers' social and emotional competencies for effective enactment of more complex instructional practices (e.g., Jennings & Greenberg, 2009; Roeser, Skinner, Beers, &

Jennings, 2012), specifically mounting evidence that teachers interpret and respond differentially to similar types of classroom demands dependent upon the stress response, or an in-the-moment appraisal of whether the available resources are commensurate with the demand (McCarthy, Lineback, & Reiser, 2014).

The Current Study

In the present study, we address policy-relevant research questions related to supporting the measurement and improvement of teaching. We investigated the dimensionality of a popular U.S. secondary student perceptions instrument, the Tripod student perception survey (Ferguson, 2010), using multilevel item factor analyses. Our aim was to examine whether different conceptualizations of effective teaching were reflected in students' endorsements of Tripod items. To do so, we examined the structural validity of the Tripod using a novel sample and probing the extent to which there is empirical support for the underlying dimensions proposed by theory.

We considered five factor structures: (a) a baseline model comprising a unidimensional structure with one overall latent factor; (b) a two-dimensional structure with factors representing press and support (Ferguson & Danielson, 2014; Lee & Smith, 1999); (c) a three-dimensional structure with dimensions corresponding to classroom organization, instructional support, and emotional support (Pianta & Hamre, 2009); (d) a seven-dimensional structure representing the conceptual organization used by the Tripod developers (Ferguson, 2010); and (e) a bifactor-like structure consisting principally of a general responsiveness dimension that partially informs responses on all items and a classroom management dimension that governs responses on items surrounding classroom management issues (Hamre et al., 2014; Lee, 1992, 2001). Further, we examined correlations between our final Tripod measurement model and ratings using the Classroom Learning Assessment Scoring System-Secondary (CLASS-S; Pianta, La Paro, & Hamre, 2008), and we used multilevel structural equation modeling to examine the predictive validity of the Tripod as it relates to teacher value-added scores.

Methods

Supported by the Bill and Melinda Gates Foundation, the MET database is the largest dataset of classroom teaching ever collected in the United States. The MET database contains a variety of data sources related to teaching collected over a 2-year period (AY 2009–2010 and AY 2010–2011) in the classrooms of more than 2,500 fourth- through ninth-grade teachers working in 317 schools located in six large school districts in the United States (Charlotte-Mecklenburg Schools, North Carolina; Dallas Independent

School District, Texas; Denver Public Schools, Colorado; Hillsborough County Public Schools, Florida; Memphis City Schools, Tennessee; New York City Department of Education, New York).

Sample and Data Collection

Recruitment for study participation began at the district level with districts having prior connections to the Gates Foundation targeted for inclusion. Once a district agreed to participate, special education schools, alternative schools, community schools, autonomous dropout and pregnancy programs, returning education schools, vocational schools that did not teach academic courses, and any other schools that had team teaching or other structural features that made it impossible to assign responsibility for a student's learning to a single, specific teacher were excluded (White & Rowan, 2014). Of the eligible schools, principals decided whether to participate. Teachers in participating schools were then recruited. The participation of particular teachers and the specific class sections to be observed determined the student sample for the MET database. Once a teacher met all eligibility criteria and consented to be in the study, efforts were made to include all her students in the sampled class section. In all districts but one, a process of passive consent provided parents the opportunity to remove their child from the study. Given the sampling procedures, the students in the MET database were, as the MET User Guide describes, "included in the study simply as a result of all these prior opportunistic processes" (White & Rowan, 2014, p. 23).

Here, we limit our analyses to sixth ($n = 388$), seventh ($n = 337$), and eighth grade ($n = 324$) mathematics class sections. The students represented in these class sections were equally split among male and female students, with 40% of students self-identifying as Latinx, 31% self-identifying as African American, and 27% self-identifying as European American. The remaining 2% of students self-identified with other races and ethnicities. Our analysis focuses on mathematics classrooms because of the well-documented potential for threatening environments to exist for students of color in science, technology, engineering, and math classrooms (e.g., McGee and Martin, 2011) and, thus, the resulting need for greater understanding of effective teaching in these classrooms.

Measures

Student Perception Survey

Adolescents' perceptions of their mathematics teacher and adolescent self-report of personal experiences in their mathematics classrooms were measured using the Tripod scale items developed by Ferguson (2010) and other survey items authored by MET researchers. The 36-item Tripod scale includes questions as follows: "My teacher seems to know if something is

bothering me” and “In this class, we learn to correct our mistakes.” Each item was rated on a 5-point Likert scale ranging from *totally untrue* to *totally true*. Negatively worded items of the Tripod survey were reverse coded. See Table 1 for a listing of all 36 Tripod items. Despite the ordered nature of responses, prior analyses concerning the Tripod instrument and other student reports of teacher behavior that have used similar Likert items have largely treated student responses as if they are continuous. However, research has indicated that such approaches, which treat ordinal data as if they had metric properties, may lead to incorrect results and inferences (e.g., Jöreskog & Moustaki, 2006).

Classroom Assessment Scoring System-Secondary (CLASS-S)

This 10-dimension observational instrument (Pianta et al., 2008) assesses student-teacher interactions and is organized into three higher-level domains. The emotional support domain comprises positive climate, negative climate (in which a low score is desirable), teacher sensitivity, and regard for student perspectives. The classroom organization domain comprises behavior management, productivity, and instructional learning formats. The instructional support domain comprises content understanding, analysis and problem solving, instructional dialogue, and quality of feedback. Each dimension of the CLASS-S is scored on a 7-point scale, with 1–2 representing low scores, 3–5 representing moderate scores, and 6–7 representing high scores. Current or former teachers trained by MET researchers scored the videos. Working under a “scoring leader,” each video coder participated in several practices, including joint review sessions and the double scoring of videos, to ensure reliable and valid coding (White & Rowan, 2014).

Teachers’ Value-Added Scores in Mathematics

Because teachers in the dataset came from different states, MET researchers standardized test scores (mean of 0 and a standard deviation of 1) for each district, subject, year, and grade level. The value-added model predicts a student’s end-of-year score on the state mathematics assessment accounting for that student’s test score in that subject from the prior year, a set of student characteristics used as controls (which varied depending on what was available by school district but typically included student demographics, free or reduced-price lunch, English-language learner status, and special education status), and the mean prior test score and mean student characteristics in the specific course section or class that the student attended (for more details, see Kane & Cantrell, 2010).

Data Analysis

Different approaches are available to model student responses to survey items. One approach ignores existing theory and associated theoretically

Table 1
Alternative Factor Structures With Item Assignment at the Within Level

Survey Item	Three-Factor			Two-Factor			Bifactor		
	Tripod Original	Instructional Support	Emotional Support	Classroom Organization	Support	Press	General	Specific	Negative
My teacher in this class makes me feel that s/ he really cares about me.	Care		▣		▣		▣		
My teacher seems to know if something is bothering me.	Care		▣		▣		▣		
My teacher really tries to understand how students feel about things.	Care		▣		▣		▣		
Student behavior in this class is under control.	Control			▣		▣	▣		
I hate the way that students behave in this class.	Control			▣		▣	▣		▣
Student behavior in this class makes the teacher angry.	Control			▣		▣	▣		▣
Student behavior in this class is a problem.	Control			▣		▣	▣		▣
My classmates behave the way my teacher wants them to.	Control			▣		▣	▣		▣
Students in this class treat the teacher with respect.	Control			▣		▣	▣		▣
Our class stays busy and does not waste time.	Control			▣		▣	▣		▣
If you don't understand something, my teacher explains it another way.	Clarify	▣			▣		▣		
My teacher knows when the class understands and when we do not.	Clarify	▣			▣		▣		
When s/he is teaching us, my teacher thinks we understand even when we don't.	Clarify	▣			▣		▣		▣
My teacher has several good ways to explain each topic that we cover in this class.	Clarify	▣			▣		▣		▣

(continued)

Table 1 (continued)

Survey Item	Tripod Original	Three-Factor			Two-Factor			Bifactor	
		Instructional Support	Emotional Support	Classroom Organization	Support	Press	General	Specific	Negative
My teacher explains difficult things clearly.		▣			▣				
My teacher asks questions to be sure we are following along when s/he is teaching.	Challenge	▣				▣			
My teacher asks students to explain more about answers they give.	Challenge	▣				▣			
In this class, my teacher accepts nothing less than our full effort.	Challenge		▣			▣			
My teacher doesn't let people give up when the work gets hard.	Challenge		▣			▣			
My teacher wants us to use our thinking skills, not just memorize things.	Challenge	▣				▣			
My teacher wants me to explain my answers—why I think what I think.	Challenge	▣				▣			
In this class, we learn a lot almost every day.	Challenge			▣					
In this class, we learn to correct our mistakes.	Challenge	▣				▣			
This class does not keep my attention—I get bored.	Captivate	▣					▣		▣
My teacher makes learning enjoyable.	Captivate		▣						
My teacher makes lessons interesting.	Captivate		▣						
I like the ways we learn in this class.	Captivate								
My teacher wants us to share our thoughts.	Confer	▣							
Students get to decide how activities are done in this class.	Confer		▣						

(continued)

Table 1 (continued)

Survey Item	Tripod Original	Three-Factor			Two-Factor			Bifactor	
		Instructional Support	Emotional Support	Classroom Organization	Support	Press	General	Specific	Negative
My teacher gives us time to explain our ideas.	Confer			▣	▣			▣	
Students speak up and share their ideas about classwork.	Confer		▣		▣			▣	
My teacher respects my ideas and suggestions.	Confer	▣			▣			▣	
My teacher takes the time to summarize what we learn each day.	Consolidate	▣			▣			▣	
My teacher checks to make sure we understand what s/he is teaching us.	Consolidate	▣			▣			▣	
We get helpful comments to let us know what we did wrong on assignments.	Consolidate	▣			▣			▣	
The comments that I get on my work in this class help me understand how to improve.	Consolidate	▣			▣			▣	

proposed dimensions and summarizes the responses in a way that yields the best correlation with value-added scores or in a way that summarizes the observed variation (e.g., through averages or principal components). This data-driven approach prioritizes statistical associations as opposed to advancing theoretical explanation or supporting causal inference. If the goal is to predict the most variation in the value-added scores, then this approach is often reasonable.

An alternative approach is to try to understand the underlying factor structure that drives students' responses. Here the starting point is existing theory, and the statistical modeling of students' responses focuses on assessing theories using the data. In this approach, understanding the underlying core causal forces that drive student responses to survey items is the central priority. If the goal is to understand the factor structure and advance theory, then predictive value is important but is not the determining criterion for which model is better.

In this study, we take the latter approach. We investigate whether adolescents' responses to perception survey items support theoretically proposed models of teaching through multilevel item factor analyses within a structural equation modeling framework. We first examined the extent to which empirical evidence supported the multidimensional structures of effective teaching proposed in the literature. Fundamentally, we were interested in understanding what the Tripod items might be measuring in order to appreciate the extent to which student surveys might be useful not only in evaluating teaching but also in helping to improve teaching.

Within this context, our analyses considered five alternative factor structures. The first model we considered was a unidimensional structure in which each Tripod item loaded onto a single general latent factor. The second model was a two-dimensional structure with factors representing press and support. The third model was a three-dimensional structure with dimensions corresponding to classroom organization, instructional support, and emotional support. The fourth model was a seven-dimensional structure comprising the dimensions of care, confer, captivate, clarify, consolidate, challenge, and control. The final model was a type of multilevel double structure bifactor model (e.g., Rijmen, 2013) consisting of a general responsiveness dimension driving all items with a secondary or specific dimension for classroom management-specific items. See Table 1 for item assignment in each model.

Our consideration of a fifth alternative factor structure (i.e., the bifactor) was motivated by the insufficient fit of each of the structures found in the literature to the current data. As we shall see, assessments of the first four structures persistently resulted in factors that were highly correlated and in the presence of a distinct and prominent classroom management factor. On the basis of these results, we reviewed existing theory and previous empirical investigations and identified an alternative bifactor-like structure

that considers a general factor that informs student responses to all items along with a classroom management factor that additionally contributes to student responses on classroom management items.

Bifactor models have been rediscovered as a viable alternative structural representation of multiple types of higher order and multidimensional factor structures (Reise, 2012). Our implementation adopts a type of extended or double-structure bifactor model that extends the conventional bifactor model by drawing on a general factor and two types of secondary factors—a secondary factor governing classroom management items and a secondary factor tracking response style for negatively worded items (e.g., Cai, Yang, & Hansen, 2011; Gibbons & Hedeker, 1992; Rijmen, 2013). Although the classroom management factor constitutes a substantively meaningful construct, we introduced the negatively worded items nuisance factor, which we refer to as a “testlet” factor, across all of the aforementioned factor structures to accommodate the increased dependence among these items.¹ Such dependence does not generalize across contexts and thus is not of direct interest. However, ignoring these effects can lead to violations of the local item independence assumption (Sliter & Zickar, 2014). In this way, our proposed model imposes a type of double structure bifactor organization.

To empirically examine the fit of each factor structure to the data, we drew on multilevel (two-tier) item factor models (e.g., Cai, 2010). Inspection of the intraclass correlation coefficients suggested substantial classroom-level variance in the item responses (see Table 2).² We used a multilevel graded response model formulation, such that

$$\begin{aligned}
 P\left(Y_{isc}^f = k\right) &= P\left(Y_{isc}^f \geq k\right) - P\left(Y_{isc}^f \geq k+1\right) P\left(Y_{isc}^f = k\right) \\
 &= \frac{1}{1 + \exp\left(-\left[a_i^{fc} \theta_c^f + a_i^{fs} \theta_{sc}^f - d_i^{k-1}\right]\right)} - \frac{1}{1 + \exp\left(-\left[a_i^{fc} \theta_c^f + a_i^{fs} \theta_{sc}^f - d_i^k\right]\right)} \quad (1)
 \end{aligned}$$

Here, Y_{isc}^f is the rating given by student s in classroom c on Tripod item i belonging to factor f , and a_i^{fc} and a_i^{fs} represent the classroom and student-level loading parameters for item i on factor f with θ_c^f as the classroom-level factor and θ_{sc}^f as the student-level factor. Let K represent the number of categories on which items are graded (seven), with k as a specific category, and let d_i^1, \dots, d_i^{k-1} be a set of $K - 1$ ordered item thresholds. To identify the scale, the distribution of each factor was set to be normally distributed with a mean of 0 and variance of 1.

Analogous to conventional multilevel regression, our multilevel factor analyses split students’ observed judgments into two complementary components. The first component consisted of classroom-level factors (θ_c^f) that describe students’ collective judgments of effective teaching in their

classrooms. The second analytic component contained student-level factors (θ_{sc}^f) that captured student-specific deviations within classrooms. That is, these student-specific deviations describe how an individual student's judgments differ from the collective judgments of his or her classroom peers. Similar to scores based on value-added models and classroom observations, the focus of our study rests on the classroom-level factors that differentiate teachers in terms of their persistent effectiveness across all students rather than the residual discrepancies among students within classrooms.

For each model that included multiple dimensions at each level, we specified multidimensional versions of Equation 1 to capture the implied student and classroom structure. In these models, we specified $\theta^f \sim MVN(0, \Sigma)$ with Σ as covariance matrix and freely estimated the covariances among the dimensions within a level. For the bifactor-like model, the general teaching quality dimension was set to be orthogonal to the classroom management-specific dimension. This specification captured the unique influence of teachers' classroom management-specific practice (i.e., variance over and above the general responsivity factor). Similarly, in each instance, we designed the negatively worded testlet factor to be orthogonal to all remaining factors implied by a given structure.

To assess the relative or predictive fit among the models, we drew on full-information maximum marginal likelihood estimation (e.g., Cai, 2010). Using the resulting log-likelihoods, we compared the models through two information criteria: Akaike information criterion (AIC) and Bayesian information criterion (BIC). These predictive fit criteria assess the extent to which the observed model fit might replicate in future equivalent samples and are population based rather than sample based (Kline, 2010).

To supplement the relative comparisons among models, we further assessed their fit through multiple absolute, incremental, and relative fit statistics. To assess the absolute and incremental fit, we drew on fit statistics based on a limited information, mean- and variance-adjusted, weighted least squares estimator (e.g., Maydeu-Olivares & Joe, 2005). Using this estimator, we first assessed fit using the χ^2 test of model fit. The χ^2 test of model fit assesses how close the observed values are to the values expected under the fitted model. Second, we measured fit using the root mean square error of approximation (RMSEA). RMSEA is a parsimony-adjusted absolute fit index that estimates the amount of error of approximation per model degree of freedom while taking sample size into account. To measure incremental fit, we examined the comparative fit index (CFI). CFI assesses the relative improvement in fit in a model compared with the null model with no item associations.

Under the maximum likelihood estimator, we subsequently examined predictive validity. A guiding principle in theories of effective teaching is that instruments should measure quality as it relates to student achievement. As a result, a primary benchmark for the validity of student ratings is their

Table 2
Descriptive Statistics for Tripod Student Perception Survey Items

Item Number	Item Text	<i>M</i>	<i>SD</i>	ICC
1	My teacher in this class makes me feel that s/he really cares about me.	3.65	1.25	.19
2	My teacher respects my ideas and suggestions.	3.71	1.15	.14
3	If you don't understand something, my teacher explains it another way.	4.04	1.08	.15
4	Student behavior in this class is under control.	3.36	1.26	.20
5	I hate the way that students behave in this class. ^a	2.51	1.34	.17
6	Student behavior in this class makes the teacher angry. ^a	3.08	1.33	.22
7	My teacher asks questions to be sure we are following along when s/he is teaching.	4.39	0.93	.11
8	My teacher wants us to share our thoughts.	3.69	1.19	.16
9	My teacher knows when the class understands and when we do not.	3.83	1.11	.12
10	My teacher asks students to explain more about answers they give.	4.13	0.97	.11
11	Students get to decide how activities are done in this class.	2.31	1.07	.15
12	When s/he is teaching us, my teacher thinks we understand even when we don't. ^a	2.45	1.23	.11
13	Student behavior in this class is a problem. ^a	2.70	1.29	.24
14	This class does not keep my attention—I get bored. ^a	2.62	1.35	.12
15	My teacher takes the time to summarize what we learn each day.	3.50	1.23	.14
16	My teacher seems to know if something is bothering me.	3.06	1.32	.16
17	My teacher checks to make sure we understand what s/he is teaching us.	4.10	1.05	.16
18	My teacher gives us time to explain our ideas.	3.67	1.13	.16
19	Students speak up and share their ideas about classwork.	3.55	1.18	.12
20	My teacher has several good ways to explain each topic that we cover in this class.	3.92	1.09	.18
21	In this class, my teacher accepts nothing less than our full effort.	4.02	1.06	.11
22	My teacher makes learning enjoyable.	3.49	1.29	.26
23	My teacher really tries to understand how students feel about things.	3.48	1.22	.18
24	My teacher doesn't let people give up when the work gets hard.	4.04	1.09	.14
25	My teacher makes lessons interesting.	3.49	1.26	.24
26	My teacher wants us to use our thinking skills, not just memorize things.	4.10	1.02	.09

(continued)

Table 2 (continued)

Item Number	Item Text	<i>M</i>	<i>SD</i>	ICC
27	My classmates behave the way my teacher wants them to.	3.10	1.22	.25
28	Students in this class treat the teacher with respect.	3.56	1.17	.30
29	We get helpful comments to let us know what we did wrong on assignments.	3.69	1.19	.13
30	My teacher wants me to explain my answers—why I think what I think.	4.07	1.03	.11
31	Our class stays busy and does not waste time.	3.50	1.17	.20
32	In this class, we learn a lot almost every day.	4.00	1.04	.15
33	My teacher explains difficult things clearly.	3.86	1.13	.17
34	The comments that I get on my work in this class help me understand how to improve.	3.70	1.16	.13
35	I like the ways we learn in this class.	3.83	1.03	.21
36	In this class, we learn to correct our mistakes.	4.08	1.02	.14

Note. Item order matches order in administered survey. We report the intraclass correlation coefficients (ICC) of items only as a heuristic indicator of the multilevel structure of the items. Subsequent analyses treat the items as categorical because they were administered using Likert-style ordinal response.

^aReverse coded.

efficacy in predicting student achievement. For this reason, we assessed the correlation between Tripod and value-added scores in mathematics. To do so, we further extended the aforementioned equations to incorporate structural components using the latent factors as predictors of the average student achievement gains in each class:

$$Y_c = \pi_0 + \sum_{f=1}^F \pi_f \theta_c^f + \varepsilon_c \quad (2)$$

We continue with the preceding notation and introduce Y_c as the value-added score on state standardized achievement tests for classroom c , π_0 as the intercept, π_f as the regression coefficients capturing the relation between Tripod dimensions and value-added, and ε_c as the error term.

Results

The results are presented in two sections. In the first section, we report the results of factor analyses. In the second section, we outline results examining the validity of the factors by reporting associations of the two classroom-level dimensions from the bifactor model with other data sources—adolescent self-report of personal experiences in the classroom, an observational assessment of effective teaching, and student learning.

Investigating the Dimensionality of the Ratings

Our analysis of the null model (that assumed no item associations) suggested that the RMSEA for this null model was unusually low (<0.10). One consequence of such a low RMSEA for the null model is that incremental indices such as CFI/Tucker-Lewis Index (TLI) may not be particularly informative because even when a proposed model improves the RMSEA to levels that are typically considered to provide evidence of reasonable fit (e.g., 0.05; MacCallum, Browne, & Sugawara, 1996), such indices will often be practically limited to values less than common cutoff values (e.g., <0.95 for CFI/TLI).³ As a result, we report the incremental fit index CFI but do not incorporate these indices into further assessments of the models. All results refer to the classroom-level unless otherwise noted.

Because the Tripod designers propose that the survey measures seven conceptual dimensions of classroom instruction, we first report the results pertaining to this structure. Despite probing multiple estimation and identification strategies, the seven-dimensional model did not reach a permissible solution (see Table 3). An assessment of the correlations among dimensions using simple classroom-level item averages for each dimension suggested that correlations were high (see Table 4). These results suggested that the

Table 3
Comparison of One-, Two-, and Three-Dimension Measurement Models

Model	χ^2	df	RMSEA	CFI ^a	SRMR (W/B) ^b	Log-Likelihood	AIC	BIC
Null	201126	1260	.10	—	.35/.71			
1D	103480	1188	.07	.49	.06/.12	-719067	1438576	1440289
2D	75329	1181	.06	.63	.05/.10	Did not converge		
3D	68720	1177	.06	.66	.05/.10	-729527 ^c	1459508 ^c	1461267 ^c
7D	Did not converge					Did not converge		
Bifactor	31490	1169	.04	.85	.03/.06	-712138	1424746	1426567

Note. RMSEA = root mean square error of approximation; CFI = comparative fit index; SRMR = standardized root mean square residual; W/B = within/between; AIC = Akaike information criterion; BIC = Bayesian information criterion. χ^2 , RMSEA, CFI, and TLI are based on the weighted least squares estimator adjusted for means, variance, and clustering; log-likelihood, AIC, and BIC are based on the maximum likelihood estimator. Information criteria are based on 30, 31, and 33 degrees of freedom for the one-, two-, and three-dimensional models, respectively.

^aThis criterion is not particularly informative because the null model RMSEA was so low.

^bIndices presented as within/between.

^cModel converged but to an unstable solution, suggesting the solution may not be unique, likely caused by the model being overly parametric (i.e., too many factors).

Table 4
Correlations Among Seven Factors Using Simple Item Averages

	Dimension						
	Care	Confer	Captivate	Clarify	Consolidate	Challenge	Control
Instructional							
Care	1						
Confer	.83	1					
Captivate	.85	.82	1				
Clarify	.82	.78	.84	1			
Consolidate	.85	.84	.84	.88	1		
Challenge	.78	.79	.80	.84	.85	1	
Control	.42	.50	.54	.44	.45	.59	1

indicators were unable to distinguish among classrooms along seven separate dimensions.

We next fit a three-factor model in which the three latent factors at the student- and classroom-levels comprised survey items that were reflective of the conceptual dimensions of instructional support, classroom organization, and emotional support (see Table 1 for the assignment of items into each of these domains). With the exception of the RMSEA, fit indices suggested that

the fit of the three-dimensional model was marginal. Furthermore, assessment of the associations among latent factors indicated high correlations. The classroom organization factor was strongly associated with the instructional support factor ($r = .79, p < .05$) and the emotional support factor ($r = .77, p < .05$). The emotional support factor and the instructional support factor were even more strongly associated ($r = .97, p < .05$). These results suggested that a more parsimonious structure might better suit the observed data.

Given that the seven- and three-factor models have very high correlations between the latent factors, we estimated more parsimonious structures using a two-factor model representing academic press and support and the baseline, one-factor model (see Table 1 for the specific items assigned to each of these factors). Despite the reduction in the number of factors, the correlation between the two factors was exceptionally high ($r = .86, p < .05$). Results concerning fit for the one- and two-dimensional models also showed worse absolute, incremental, and relative fit as compared to the three-dimensional model.

Finally, we assessed the proposed bifactor model. Fit indices largely fell within acceptable ranges, and comparative assessments of relative fit indices demonstrated substantial improvements over the previous models (Table 3). With evidence supporting the plausibility of the proposed bifactor structure, we also evaluated the strength with which the Tripod items reflected each of the latent dimensions, using the loading parameters (as shown in Table 5). Loading parameters conceptually capture the strength of the correlation between responses on a particular item and underlying dimensions and, thus, describe how well an item can differentiate among classrooms in terms of their underlying quality. The results suggested that items were appreciably and significantly correlated with the underlying dimension(s). Collectively, the resulting estimates of the loading parameters indicated that most items were able to differentiate fairly well among different latent levels on each dimension.

General Responsivity and Classroom Management-Specific Associations With Adolescent Self-report of Personal Experiences in the Classroom

At the classroom level, adolescents' self-reports of the positivity of the classroom climate were strongly associated with the general responsivity factor. For example, "this class is a happy place for me to be" (.82), "for a new student, this class would be a good one to join" (.72), "being in this class makes me angry" (-.67), and "I feel stressed in this class" (-.61). Adolescents' self-reports of effort were weakly to moderately correlated with the general responsivity factor: for example, "I take it easy and do not try my best" (-.14), and "I stop trying when the work gets hard" (-.32). Across all items, self-reports of personal experiences in the classroom

Table 5
Loading Parameters of the Bifactor Model

	General Factor		Specific Factor		Negative Testlet	
	Within	Between	Within	Between	Within	Between
My teacher in this class makes me feel that s/he really cares about me.	1.9 (0.03)	1.11 (0.04)				
My teacher seems to know if something is bothering me.	1.34 (0.02)	0.78 (0.03)				
My teacher really tries to understand how students feel about things.	1.94 (0.03)	1.04 (0.03)				
If you don't understand something, my teacher explains it another way.	1.91 (0.03)	0.94 (0.03)				
My teacher knows when the class understands and when we do not.	1.52 (0.02)	0.07 (0.03)				
When s/he is teaching us, my teacher thinks we understand even when we don't.	0.82 (0.02)	0.53 (0.02)				
My teacher has several good ways to explain each topic that we cover in this class.	2.12 (0.03)	1.16 (0.04)				
My teacher explains difficult things clearly.	2.19 (0.03)	1.08 (0.04)				
This class does not keep my attention—I get bored.	1.15 (0.02)	0.8 (0.03)			0.68 (0.03)	—
My teacher makes learning enjoyable.	2.1 (0.03)	1.67 (0.05)				
My teacher makes lessons interesting.	2.17 (0.03)	1.57 (0.05)				
I like the ways we learn in this class.	1.57 (0.02)	1.21 (0.04)				
My teacher wants us to share our thoughts.	1.06 (0.02)	0.63 (0.02)				
Students get to decide how activities are done in this class.	0.68 (0.02)	0.67 (0.02)				
My teacher gives us time to explain our ideas.	1.99 (0.03)	0.90 (0.03)				
Students speak up and share their ideas about classwork.	1.37 (0.02)	0.69 (0.03)				
My teacher respects my ideas and suggestions.	2.13 (0.03)	0.95 (0.03)				

(continued)

Table 5 (continued)

	General Factor		Specific Factor		Negative Testlet	
	Within	Between	Within	Between	Within	Between
My teacher takes the time to summarize what we learn each day.	1.61 (0.02)	0.66 (0.03)				
My teacher checks to make sure we understand what s/he is teaching us.	2.36 (0.03)	1.00 (0.04)				
We get helpful comments to let us know what we did wrong on assignments.	1.82 (0.03)	0.78 (0.03)				
The comments that I get on my work in this class help me understand how to improve.	1.88 (0.03)	0.83 (0.03)				
My teacher asks questions to be sure we are following along when s/he is teaching.	1.33 (0.02)	0.53 (0.03)				
My teacher asks students to explain more about answers they give.	1.15 (0.02)	0.32 (0.02)				
In this class, my teacher accepts nothing less than our full effort.	1.37 (0.02)	0.48 (0.02)				
My teacher doesn't let people give up when the work gets hard.	1.61 (0.03)	0.74 (0.03)				
My teacher wants us to use our thinking skills, not just memorize things.	1.54 (0.02)	0.49 (0.03)				
My teacher wants me to explain my answers—why I think what I think.	1.42 (0.02)	0.42 (0.02)				
In this class, we learn a lot almost every day.	1.63 (0.03)	0.65 (0.03)				
In this class, we learn to correct our mistakes.	2.03 (0.03)	0.81 (0.03)				

(continued)

Table 5 (continued)

	General Factor		Specific Factor		Negative Testlet	
	Within	Between	Within	Between	Within	Between
Our class stays busy and does not waste time.	1.01 (0.02)	0.44 (0.04)	0.65 (0.02)	0.77 (0.03)		
Student behavior in this class is under control.	0.98 (0.02)	0.72 (0.05)	1.44 (0.03)	1.08 (0.03)		
I hate the way that students behave in this class.	0.14 (0.02)	0.28 (0.04)	1.13 (0.03)	1.00 (0.03)	1.06 (0.04)	—
Student behavior in this class makes the teacher angry.	0.4 (0.02)	0.61 (0.04)	1.03 (0.03)	0.96 (0.03)	0.98 (0.03)	—
Student behavior in this class is a problem.	0.5 (0.03)	0.63 (0.06)	1.55 (0.03)	1.37 (0.04)	1.15 (0.04)	—
My classmates behave the way my teacher wants them to.	1.26 (0.03)	0.99 (0.06)	1.84 (0.04)	1.29 (0.04)		
Students in this class treat the teacher with respect.	1.15 (0.03)	1.13 (0.06)	1.41 (0.03)	1.21 (0.04)		

Note. Maximum likelihood estimates of loadings were obtained using the typical logit link standard errors in parentheses.

were weakly correlated (.00–.31) with the classroom management–specific factor.

General Responsivity and Classroom Management–Specific Associations With Adult Assessments of Effective Teaching

A second set of correlational analyses were carried out to assess how the general responsivity and classroom management–specific factors related to the three domains (instructional support, emotional support, and classroom organization) of the CLASS-S (Pianta et al., 2008), an adult-scored observational measure of teachers’ instruction based largely upon student-teacher interactions as the source material for effectiveness. Due to missing data, this correlational analysis could be conducted on slightly more than half of all classrooms in our initial sample. The classroom management–specific factor had the strongest relative association to the classroom organization domain of the CLASS-S (.19) but not to CLASS-S emotional support domain. In contrast, the general responsivity factor was unrelated to the classroom organization domain but had the strongest relative association with the emotional support domain of the observational instrument (.20). The general responsivity factor was weakly correlated with CLASS-S instructional support domain (.14), and the classroom management–specific factor was unrelated (–.01).

General Responsivity and Classroom Management–Specific Associations With Student Learning

Finally, we examined the predictive validity of the bifactor solution. The results, reported at the classroom level (i.e., standardized at the classroom level), suggested that both the general responsivity factor and classroom management–specific factor were positively and significantly associated with teachers’ value-added measures based on the state mathematics test (standardized coefficients were .25 and .25, respectively). More specifically, a 1 standard deviation increase in either dimension was associated with a one-quarter standard deviation increase in teachers’ value-added scores. Because these bifactor dimensions were orthogonal by design, a 1 standard deviation increase in both dimensions would be associated with an overall .50 standard deviation increase in teachers’ value-added scores. Importantly, the correlation between teacher value-added and the classroom management–specific factor suggests that this factor is not a method effect due to differential wording of items.

Discussion

Adolescent students are unique reporters of classroom interactions. Without observing a teacher’s instructional practice through the lens of an

adult-developed theory of effective teaching, adolescents likely draw on their indigenous expertise of what makes them feel safe, respected, and competent. In other words, adult raters using observational instruments are trained (i.e., given descriptive or example anchors for the different scoring levels) in order to calibrate the assessment of effectiveness in relation to a particular theory of effective teaching. Adolescent raters using perception surveys, like the Tripod, however, are not subjected to a training process by which such explicit anchors are shared with raters in a systematic way.

Initial classroom-level analyses of Tripod data from the MET database indicated that student perceptions of a given teacher's strengths and weaknesses are fairly consistent across the different classrooms of students they teach (classroom-to-classroom correlations ranged from .58 to .68, depending on the dimension) and that aggregate student responses on a subset of Tripod items from one class section are significantly associated with student achievement gains in other class sections taught by the same teacher ($r = .219$ different section, $r = .235$ previous year; Kane & Cantrell, 2010). While these results are intriguing because of the sheer volume of classrooms from which the data were drawn, these studies did not explore the theory of teaching implied by the items comprising the Tripod instrument. We posit that the high correlations among the theoretical factors found in the literature on the Tripod measure (e.g., ranging from .65 to .95 in Raudenbush & Jean, 2014) and in our analyses as well as our evidence concerning the relative and absolute fit of the proposed factor structures suggest that those theoretical factors previously posed in the literature are largely implausible in the current sample and are fundamentally incomplete. That is, the theories found in the literature appear to be ill-defined and offer little basis for understanding how student surveys and theories of effective teaching intersect to inform teaching, teacher development, and assessments of teaching.

It is unclear whether the original 7Cs that describe the Tripod instrument were intended to capture seven distinct dimensions on which students can reliably discriminate among teachers or whether the 7Cs were merely intended to be more heuristic domains that map out important aspects of teaching. The initial investigations, both published and unpublished, of the Tripod have made important contributions, yet they are incomplete in many ways (e.g., Raudenbush & Jean, 2014; Schweig, 2014). Our study extends these investigations through a complementary but different perspective that endeavors to delineate the specific theoretical nature of the constructs being measured so as to inform teacher development and the appropriate use of student perception surveys.

Our results replicated the previous findings that student perceptions of teacher behavior and the learning environment measured in the Tripod track classroom features that relate to teachers' value-added scores. That said, what specific construct the general responsiveness factor instantiates remains somewhat vague and indefinite. We produced preliminary evidence that

the social and emotional experiences of students in a class section may be particularly salient source material influencing student endorsements when completing the Tripod survey. Evidence for this claim includes the moderate, but relatively strongest, association of the general responsiveness factor with the emotional support domain of the CLASS-S as well as the strong association of the general responsiveness factor with adolescents' self-reports of the positivity of the classroom climate.

On the basis of evidence generated across studies of the Tripod, including the one reported here, it seems reasonable to suggest that adolescents' interpretations of a teacher's enactment of particular classroom management actions contribute additional and unique information about the quality of a classroom experience above and beyond general responsiveness. We contribute further evidence for this claim given that in this study, the classroom management-specific factor was moderately associated with the classroom organization domain of the CLASS-S but unrelated to the emotional support domain of the CLASS-S.

The level of cooperation among students and their teacher may provide student information on teacher effectiveness in the most unequivocal manner (Gregory & Ripski, 2008). For example, teachers can proactively support students' autonomous completion of tasks or seek compliance through controlling feedback (Wallace, Sung, & Williams, 2014). The role of appraisals and coping resources highlighted in a transactional model of teacher stress (e.g., McCarthy et al., 2014) may explain teachers' differential responses to specific events in classrooms. In turn, adolescents may well be sensitive to the resulting variation in cooperative interactions on the basis of the concrete particulars of the enactment of classroom management. For example, a student texting during the teacher's launching of a task may be experienced by a teacher as either (a) an energizing challenge to regain the student's attention or (b) a trigger for frustration and anger. This variability may be captured by the Tripod instrument through items such as "Student behavior in this class makes the teacher angry." This item may instantiate general teacher responsiveness in terms of adolescents' broad perceptions of the teacher's disposition toward students and expression of negative emotions. This item may simultaneously instantiate adolescents' specific perceptions of the teacher's explicit public appraisals of student misbehavior as either a productive challenge that requires adapting instructional strategies or confirmatory evidence of deficit-based thinking about students. The critical role of such public appraisals to establishing and maintaining cooperative interactions is well documented in the culturally responsive classroom management literature (e.g., Weinstein, Tomlinson-Clarke, & Curran, 2004). Indeed, the work of culturally responsive teachers is a labor-intensive process that requires dedication and reflection of teachers to develop and refine such practices over time (Cholewa, Amatea, West-Olatunji, & Wright, 2012).

One unique feature of Tripod survey items comprising the classroom management–specific factor, relative to the rest of the survey items, is the extent to which they reference students rather than teachers. Indeed, four of the seven survey items ask students to reflect on other students in their classroom, not on the teacher. Particularly in adolescence, peer relationships become more salient in the classroom (Martin & Dowson, 2009; Wentzel, 2005). Adolescents' feelings of being connected to their classroom peers are associated with positive behavioral and motivational outcomes, including their self-efficacy (Ryan & Patrick, 2001), achievement motivation (Nelson & DeBacker, 2008; Ruzek, Hafen, Allen, Gregory, Mikami & Pianta, 2016), expectancies for success (Goodenow, 1993), engagement (Furrer & Skinner, 2003; Shernoff, Ruzek, & Sinha, 2016), and school interest (Wentzel, Battle, Russell, & Looney, 2010). To the extent that students answer these survey items thinking only about their peers, then these items may also reflect nuances in the peer context of the classroom and could provide unique assessments of peer effects as related to classroom management.

Without further research on the validity of the proposed bifactor model and theory-based augmentations of it, using this particular student survey measure as a basis for guiding teacher professional development efforts is limited. This is most especially true for the general factor, which is an amalgamation of all items on the survey instrument. If a teacher scores low on the Tripod's general responsivity factor, then what specific aspects of her instructional practice should an instructional coach or mentor attend to? Specificity is highly valuable in these professional development contexts because it provides guidance for improvement efforts. This is less of a concern with the classroom management–specific factor, which has specificity around the amount of respect students and teachers show each other, the extent to which the class remains productively focused on learning, and the degree to which peer behavior is perceived as psychologically safe. Low scores in the classroom management–specific factor are thus truly diagnostic and could likely be bolstered with focused efforts to improve the tenor of classroom discourse and the nature of peer interaction.

If surveys are to be used in efforts to evaluate and improve teachers' instructional practice, then it is critical to gain an understanding of what the surveys are, in fact, measuring about such practice. What unique information do survey results add to the other instruments or measures of teaching quality (e.g., value-added scores, observational protocols)? The results from this study suggest that the Tripod may provide unique information on classroom management. Interestingly, in prior research on the Tripod, the Control subscale (typically measured as a simple average of the individual items) has been among the best predictors of other measures of instructional quality (Ferguson & Danielson, 2014; Kane & Cantrell, 2010), so it is

perhaps not surprising that our results suggest that these items provide valuable information about teaching. Similarly interesting is the fact that this information is provided by students, not adults, suggesting that students' views of classroom management are an important indicator of a high-functioning classroom.

While we have generated preliminary evidence for the plausibility of a bifactor-like theory of effective teaching that builds on recent research (e.g., Hamre et al., 2014; Lee 1992, 2001), a critical, but yet unknown, aspect of students' assessments of teacher quality gathered via surveys is whether adolescent reports of teaching quality may (a) detect substantive differences in teaching quality that have implications for our theories of effective teaching or (b) be influenced by rater effects (e.g., extreme response styles) due to the fact that adolescents are not trained to view and assess the classroom in any particular manner so by default draw on their naturally occurring interpretations to endorse items. Without an empirically based understanding of whether or how either of these two factors influence student perception survey results, it remains unclear how such data should be appropriately used in teacher evaluation and training efforts. Accordingly, the magnitude of the theoretical and practical implications of large-scale investigations of effective teaching necessitates thoughtful investigations of the construct validity of the data produced for the purposes of teacher evaluation.

Conclusion

To use student surveys in teacher evaluation systems (such as those being developed in the United States), policymakers and administrators must understand, at a minimum, the degree to which the survey items function as indicators of distinct aspects of instruction believed to be important—that is, we must amass evidence that indicators are linked to experienced aspects of teaching that drive effective teaching. In addition to exploring what Clausen (2002; as cited in Kunter & Baumert, 2006) calls “perspective-specific validities,” we must examine additional questions such as whether student raters have the same interpretation of survey items across different contexts (allowing for the comparison of teacher instructional quality across classrooms, schools, districts, or even states) and why and how student characteristics influence interpretations of items. Only in the past decade have researchers begun to assess such aspects of student ratings of instruction (e.g., Fauth et al., 2014; Kunter, Tsai, Klusmann, Brunner, Krauss, & Baumert, 2008; Wagner et al., 2013), a necessity if these measures are to be used in high-stakes teacher evaluation systems. So while the general consensus among researchers is that student surveys are useful evaluation instruments—especially when used in combination with other evaluation tools—questions remain about what these surveys capture about teaching practice, and our results (e.g., the modest fit of several factorial

structures we tested) suggest that this concern is particularly relevant to the Tripod.

Despite these unresolved issues, U.S. states and school districts increasingly rely upon student perception surveys in the mix of measures they use to evaluate teachers' instructional quality. In the United States, one of the most popular student surveys is the Tripod student self-perception survey (Ferguson, 2010), which was the focus of the present investigation. To our knowledge, this study is the first to systematically investigate the multidimensionality of the Tripod student perception survey using student-level data from the MET database.

Notes

¹We conducted all of our analyses with and without the negative testlet factor; results were substantively the same in each instance, but fit was consistently better for models that included a negative testlet factor. For the bifactor model, items included in the negative testlet were part of both the general and specific factor.

²We report the intraclass correlation coefficients of items only as a heuristic indicator of the multilevel structure of the items. Subsequent analyses treat the items as categorical because they were administered using Likert-style ordinal response.

³For instance, if we rewrite the expression for TLI as a function of the RMSEA, we have

$$TLI = \frac{\chi_{null}^2/d_{null} - \chi_{my}^2/d_{my}}{\chi_{null}^2/d_{null} - 1} = \frac{(1 + R_{null}^2(n-1)) - (1 + R_{my}^2(n-1))}{R_{null}^2(n-1)} = \frac{R_{null}^2 - R_{my}^2}{R_{null}^2},$$

where $RMSEA = R = \sqrt{\frac{\chi^2 - d}{d(n-1)}}$, χ^2 is the model based chi-square, d represents the degrees of freedom, n is the sample size, $null$ represents the null model, and my represents the proposed model. If the RMSEA from the null model is $R_{null} = .10$ (as is the case in our study) and a proposed model reduces the RMSEA from a null of .10 to levels typically described as very good, such as .05, the resulting TLI would be only $(.10^2 - .05^2)/.10^2 = .75$. In this way, CFI/TLI are not particularly informative when the null RMSEA is low (e.g., <.15).

References

- Brophy, J. E., & Good, T. L. (1974). *Teacher-student relationships: Causes and consequences*. Oxford, England: Holt, Rinehart & Winston.
- Cai, L. (2010). A two-tier full-information item factor analysis model with applications. *Psychometrika*, 75(4), 581–612. doi:10.1007/s11336-010-9178-0
- Cai, L., Yang, J., & Hansen, M. (2011). Generalized full-information item bifactor analysis. *Psychological Methods*, 16, 221–248.
- Carlisle, J., Kelcey, B., Berebitsky, D., & Phelps, G. (2011). Embracing the complexity of instruction: A study of the effects of teachers' instruction on students' reading comprehension. *Scientific Studies of Reading*, 15, 409–439. doi:10.1080/10888438.2010.497521
- Cartledge, G., Lo, Y., Vincent, C. & Robinson-Ervin, P. (2015). Culturally responsive classroom management. In C. Evertson & C. Weinstein (Eds.), *Handbook of classroom management* (2nd Ed.). New York, New York: Routledge Publications.

- Cholewa, B., Amatea, E., West-Olatunji, C. A., & Wright, A. (2012). Examining the relational processes of a highly successful teacher of African American children. *Urban Education, 47*(1), 250–279. doi:10.1177/0042085911429581
- Davis, H. A. (2003). Conceptualizing the role and influence of student-teacher relationships on children's social and cognitive development. *Educational Psychologist, 38*(4), 207–234. doi:10.1207/S15326985EP3804_2
- Desimone, L., Smith, T., & Frisvold, D. (2010). Survey measures of classroom instruction: Comparing student and teacher reports. *Educational Policy, 24*(2), 267–329. doi:10.1177/0895904808330173
- Emmer, E. T., & Strough, L. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist, 36*(2), 103–112. doi:10.1207/S15326985EP3602_5
- Erickson, F. (1987). Transformation and school success: The politics and culture of educational achievement. *Anthropology and Education Quarterly, 18*(4), 335–356. doi:10.1525/aeq.1987.18.4.04x0023w
- Erickson, F., Bagrodia, R., Cook-Sather, A., Espinoza, M., Jurow, S., Shultz, J. J., & Spencer, J. (2007). Students' experience of school curriculum: The everyday circumstances of granting and withholding assent to learn. In F. M. Connelly, M. F. He, & J. Phillion (Eds.), *The Sage Handbook of Curriculum and Instruction* (pp. 198–218). Thousand Oaks, CA: Sage.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction, 29*, 1–9. doi:10.1016/j.learninstruc.2013.07.001
- Ferguson, R. (2010, October 14). Student perceptions of teaching effectiveness. Discussion brief from the National Center for Teacher Effectiveness and the Achievement Gap Initiative, Harvard University, Cambridge, MA.
- Ferguson, R. F., & Danielson, C. (2014). How framework for teaching and Tripod 7Cs evidence distinguish key components of effective teaching. In T. Kane, K. Kerr, & R. Pianta (Eds.), *Designing teacher evaluation systems: New guidance from the Measures of Effective Teaching project* (pp. 98–143). San Francisco: Jossey-Bass.
- Friedrich, A., Flunger, B., Nagengast, B., Jonkmann, K., & Trautwein, U. (2015). Pygmalion effects in the classroom: Teacher expectancy effects on students' math achievement. *Contemporary Educational Psychology, 41*, 1–12. doi:10.1016/j.cedpsych.2014.10.006
- Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology, 95*(1), 148–162. doi:10.1037/0022-0663.95.1.148
- Gay, G. (2010). *Culturally responsive teaching: Theory, research, and practice*. New York: Teachers College Press.
- Gest, S. D., & Rodkin, P. C. (2011). Teaching practices and elementary classroom peer ecologies. *Journal of Applied Developmental Psychology, 32*(5), 288–296. doi:10.1016/j.appdev.2011.02.004
- Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika, 57*, 423–436.
- Good, T. L. (1987). Two decades of research on teacher expectations: Findings and future directions. *Journal of Teacher Education, 38*(4), 32–47. doi:10.1177/002248718703800406
- Goodenow, C. (1993). The psychological sense of school membership among adolescents: Scale development and educational correlates. *Psychology in the Schools, 30*(1), 79–90. doi:10.1002/1520-6807(199301)30:1<79::AID-PITS2310300113>3.0.CO;2-X

- Greeno, J. (1998). The situativity of knowing, learning, and research. *American Psychologist*, 53(1), 5–26. doi:10.1037/0003-066X.53.1.5
- Gregory, A., & Ripski, M. B. (2008). Adolescent trust in teachers: Implications for behavior in the high school classroom. *School Psychology Review*, 37(3), 337.
- Hamre, B., Hatfield, B., Pianta, R., & Jamil, F. (2014). Evidence for general and domain-specific elements of teacher-child interactions: Associations with pre-school children's development. *Child Development*, 85(3), 1257–1274. doi:10.1111/cdev.12184
- 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. doi:10.3102/0034654308325693
- Jöreskog, K. G., & Moustaki, I. (2006). *Factor analysis of ordinal variables with full information maximum likelihood*. Unpublished manuscript.
- Kane, T. J., & Cantrell, S. (2010). *Learning about teaching: Initial findings from the measures of effective teaching project*. Seattle: The Bill & Melinda Gates Foundation.
- Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle: The Bill & Melinda Gates Foundation.
- Kleinfeld, J. (1975). Effective teachers of Eskimo and Indian students. *The School Review*, 83, 301–344.
- Kline, R. B. (2010). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.
- Kulik, J. A., & Kulik, C. L. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, 58(1), 79–97. doi:10.3102/00346543058001079
- Kunter, M., & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, 9(3), 231–251. doi:10.1007/s10984-006-9015-7
- Kunter, M., Tsai, Y.-M., Klusmann, U., Brunner, M., Krauss, S., & Baumert, J. (2008). Students' and mathematics teachers' perceptions of teacher enthusiasm and instruction. *Learning and Instruction*, 18(5), 468–482. doi:10.1016/j.learninstruc.2008.06.008
- Lee, C. D. (1992). Literacy, cultural diversity, and instruction. *Education and Urban Society*, 24(2), 279–291. doi:10.1177/0013124592024002008
- Lee, C. D. (2001). Is October Brown Chinese? A cultural modeling activity system for underachieving students. *American Educational Research Journal*, 38(1), 97–141. doi:10.3102/00028312038001097
- Lee, V. E., & Smith, J. B. (1999). Social support and achievement for young adolescents in Chicago: The role of school academic press. *American Educational Research Journal*, 36(4), 907–945. doi:10.3102/00028312036004907
- Lee, V. E., Smith, J. B., Perry, T. E., & Smylie, M. A. (1999). Social support, academic press, and student achievement: A view from the middle grades in Chicago. Improving Chicago's schools. Chicago: Chicago Annenberg Research Project.
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1, 130–149.
- Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, 52(11), 1187. doi:10.1037/0003-066X.52.11.1187
- 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. doi:10.3102/0034654308325583
- Maydeu-Olivares, A., & Joe, H. (2005). Limited-and full-information estimation and goodness-of-fit testing in 2 n contingency tables: a unified framework. *Journal of the American Statistical Association*, 100(471), 1009–1020. doi:10.2139/ssrn.1019782
- Mayer, R. E. (2002). Rote versus meaningful learning. *Theory Into Practice*, 41(4), 226–233. doi:10.1207/s15430421tip4104_4
- McCarthy, C. J., Lineback, S., & Reiser, J. (2014). Teacher stress, emotion, and classroom management. In E. Emmer & E. J. Sabornie (Eds.), *Handbook of classroom management* (2nd ed.). New York: Routledge.
- McCaslin, M. (2009). Co-regulation of student motivation and emergent identity. *Educational Psychologist*, 44(2), 137–146. doi:10.1080/00461520902832384
- McCaslin, M., & Hickey, D. T. (2001). Self-regulated learning and academic achievement: A Vygotskian view. In B. J. Zimmerman & D. Schunk (Eds.), *Self-regulated learning and academic achievement: Theory, research, and practice* (2nd ed., pp. 227–252). Mahwah, NJ: Erlbaum.
- McGee, E. O., & Martin, D. B. (2011). “You would not believe what I have to go through to prove my intellectual value!” Stereotype management among academically successful black mathematics and engineering students. *American Educational Research Journal*, 48(6), 1347–1389. 10.3102/0002831211423972
- Milner, H. R., & Tenore, F. B. (2010). Classroom management in diverse classrooms. *Urban Education*, 45(5), 560–603. doi: 10.1177/0042085910377290
- Nakkula, M. J., & Toshalis, E. (2006). *Understanding youth: Adolescent development for educators*. Cambridge: Harvard Education Publishing Group.
- Nelson, R. M., & DeBacker, T. K. (2008). Achievement motivation in adolescents: The role of peer climate and best friends. *Journal of Experimental Education*, 76(2), 170–189. doi:10.3200/JEXE.76.2.170–190
- Noguera, P. A. (2003). Schools, prisons, and social implications of punishment: Rethinking disciplinary practices. *Theory Into Practice*, 42, 341–350. doi:10.1207/s15430421tip4204_12
- Patrick, H., Anderman, L. H., & Ryan, A. M. (2004). Social motivation and the classroom environment. In C. Midgley (Ed.), *Goals, goal structures, and patterns of adaptive learning* (pp. 85–108). Mahwah, NJ: Lawrence Erlbaum Associates.
- Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109–119. doi:10.3102/0013189X09332374
- Pianta, R. C., La Paro, K., & Hamre, B. K. (2008). *Classroom assessment scoring system*. Baltimore: Brookes.
- Raudenbush, S. W., & Jean, M. (2014). To what extent do student perceptions of classroom quality predict teacher value added? In T. Kane, K. Kerr, & R. Pianta (Eds.), *Designing teacher evaluation systems: New guidance from the Measures of Effective Teaching project* (pp. 170–202). San Francisco: Jossey-Bass.
- Reeve, J., & Jang, H. (2006). What teachers say and do to support students’ autonomy during a learning activity. *Journal of Educational Psychology*, 98(1), 209–218. doi:10.1037/0022-0663.98.1.209
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47(5), 667–696. doi:10.1080/00273171.2012.715555
- Rijmen, F. (2013). Hierarchical factor item response theory models for PIRLS: Capturing clustering effects at multiple levels. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, 4, 59–74.

- Roeser, R. W., Skinner, E., Beers, J., & Jennings, P. A. (2012). Mindfulness training and teachers' professional development: An emerging area of research and practice. *Child Development Perspectives*, 6(2), 167–173. doi:10.1111/j.1750-8606.2012.00238.x
- Ruzek, E. A., Hafen, C. A., Allen, J. P., Gregory, A., Mikami, A. Y., & Pianta, R. C. (2016). How teacher emotional support motivates students: The mediating roles of perceived peer relatedness, autonomy support, and competence. *Learning and Instruction*, 42, 95–103. doi:10.1016/j.learninstruc.2016.01.004
- Ryan, A. M., & Patrick, H. (2001). The classroom social environment and changes in adolescents' motivation and engagement during middle school. *American Educational Research Journal*, 38(2), 437–460. doi:10.3102/00028312038002437
- Schweig, J. (2014). Cross-level measurement invariance in school and classroom environment surveys: Implications for policy and practice. *Educational Evaluation and Policy Analysis*, 36(3), 259–280. doi:10.3102/0162373713509880
- Shermoff, D. J., Ruzek, E. A., & Sinha, S. (2016). The influence of the high school classroom environment on learning as mediated by student engagement. *School Psychology International*. doi:10.1177/0143034316666413
- Sliter, K. A., & Zickar, M. J. (2014). An IRT examination of the psychometric functioning of negatively worded personality items. *Educational and Psychological Measurement*, 74(2), 214–226. doi:10.1177/0013164413504584
- Spearman, J., & Watt, H. M. G. (2013). Perception shapes experience: The influence of actual and perceived classroom environment dimensions on girls' motivations for science. *Learning Environments Research*, 16(2), 217–238. doi:10.1007/s10984-013-9129-7
- Thompson, J., Windschitl, M., & Braaten, M. (2013). Developing a theory of ambitious early-career teacher practice. *American Educational Research Journal*, 50(3), 574–615. doi:10.3102/0002831213476334
- Toshalis, E. (2015). *Make me: Understanding and engaging student resistance in school*. Cambridge: Harvard Education Publishing Group.
- Turner, J. C., & Meyer, D. K. (2000). Studying and understanding the instructional contexts of classrooms: Using our past to forge our future. *Educational Psychologist*, 35(2), 69–85. doi:10.1207/S15326985EP3502_2
- Wagner, W., Göllner, R., Helmke, A., Trautwein, U., & Lüdtke, O. (2013). Construct validity of student perceptions of instructional quality is high, but not perfect: Dimensionality and generalizability of domain-independent assessments. *Learning and Instruction*, 28, 1–11. doi:10.1016/j.learninstruc.2013.03.003
- Wallace, T. L., & Chhuon, V. (2014). Proximal processes in urban classrooms: Engagement and disaffection in urban youth of color. *American Educational Research Journal*, 51(5), 937–973. doi:10.3102/0002831214531324
- Wallace, T. L., Sung, H. C., & Williams, J. D. (2014). The defining features of teacher talk within autonomy-supportive classroom management. *Teaching and Teacher Education*, (42), 34–46. doi:10.1016/j.tate.2014.04.005
- Wanless, S.B. (2016). The role of psychological safety in human development. *Research in Human Development*, 13(1), 6–14. doi: 10.1080/15427609.2016.1141283
- Ware, F. (2006). Warm demander pedagogy culturally responsive teaching that supports a culture of achievement for African American students. *Urban Education*, 41(4), 427–456. doi:10.1177/0042085906289710
- Weinstein, C. S., Tomlinson-Clarke, S., & Curran, M. (2004). Toward a conception of culturally responsive classroom management. *Journal of Teacher Education*, 55(1), 25–38. doi:10.1177/0022487103259812

- Wentzel, K. R. (2005). Peer relationships, motivation, and academic performance at school. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 279–295). New York: Guilford Press.
- 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. doi:10.1016/j.cedpsych.2010.03.002
- Williams, J. D., Woodson, A. N., & Wallace, T. L. (2016). “Can we say the N-word?” Exploring psychological safety during race talk. *Research in Human Development, 13*(1), 15–31. doi:10.1080/15427609.2016.1141279
- White, M., & Rowan, B. (2014). *User guide to the Measures of Effective Teaching Longitudinal Database (MET LDB)*. Ann Arbor: Inter-University Consortium for Political and Social Research, University of Michigan.

Manuscript received October 7, 2015

Final revision received July 1, 2016

Accepted September 1, 2016