11-2013

Teaching Behavior and Well-Being in Students: Development and Concurrent Validity of an Instrument to Measure Student-Reported Teaching Behavior

Patrick Pössel
University of Louisville, patrick.possel@louisville.edu

Kathleen Moritz Rudasill
University of Nebraska-Lincoln, kmrudasill@vcu.edu

Jill L. Adelson
University of Louisville, jill.adelson@louisville.edu

Annie C. Bjerg
University of Louisville

Don T. Wooldridge
University of Louisville

See next page for additional authors

The International Journal of Emotional Education
ISSN: 2073-7629

Volume 5, Number 2, November 2013 pp 5-30
www.um.edu.mt/cres/ijee

Teaching Behavior and Well-Being in Students: Development and Concurrent Validity of an Instrument to Measure Student-Reported Teaching Behavior

Patrick Pössel,[1a] Kathleen Moritz Rudasill,[b] Jill L. Adelson[c], Annie C. Bjerg[c], Don T. Wooldridge[c], and Stephanie Winkeljohn Black[c]

[a]Department of Educational and Counseling Psychology, Counseling, and College Student Personnel, University of Louisville, KY, USA

[b]Department of Educational Psychology, University of Nebraska-Lincoln, NE, USA

[c]Department of Educational and Counseling Psychology, Counseling, and College Student Personnel, University of Louisville, KY, USA

Teaching behavior has important implications for students' emotional well-being. Multiple models suggest students' perceptions of teaching behaviors are more critical than other measures for predicting well-being, yet student-report instruments that measure concrete and specific teaching behavior are limited. The purpose of the present studies is to develop an instrument to assess students' perceptions of concrete and specific teaching behavior and to test which teaching behavior is associated students' well-being. Construct validity and internal consistency for the 37-item Teaching Behavior Questionnaire (TBQ-S), composed of instructional, negative teaching, socio-emotional, and organizational behavior were examined using data from two independent samples (Study 1: n = 703; Study 2: n = 822). The factor structure was stable across both samples and internal consistencies ranged from .77 to .97. Results indicated student-ratings of teaching behavior were associated with positive and negative affect in students.

Keywords: student-report, teaching behavior, students' well-being, psychometrics, factor analyses

First submission September 5th 2012; Accepted for publication October 22nd, 2013.

[1] Corresponding author. Email address: patrick.possel@louisville.edu

Introduction

Abundant research supports the notion that teacher support has clear implications for students' emotional well-being (hereafter called well-being). Consistent with previous research, we conceptualize well-being as comprising positive and negative affect (Huebner & Dew, 1996). Positive affect is the extent to which a person typically feels positive emotions (e.g., is enthusiastic, active, and alert). Negative affect encompasses frequent negative feelings (e.g., is distressed, angry, nervous). Well-being is not only of subjective importance for students; negative affect is associated with academic problems including reduced homework completion, less concentration in class, fewer interactions with peers, poorer class attendance, and lower rates of post-secondary degree attainment (Humensky, Kuwabara, Fogel, Wells, Goodwin, & Van Voorhees, 2010; Jonsson, Bohman, Hjern, von Knorring, Olsson, & von Knorring, 2010). To the contrary, positive affect in students towards school (e.g., school liking, a sense of belonging) tends to be associated with higher classroom engagement (Furrer & Skinner, 2003; Patrick, Ryan, & Kaplan, 2007; Ryan, Stiller, & Lynch, 1994) and academic achievement (Crosnoe, Johnson, & Elder, 2004; Davis, 2006; Niehaus, Rudasill, & Rakes, 2012; Roeser, Midgley, & Urdan, 1996).

Research has shown that teacher support plays an important role in students' overall well-being. For example, students who feel supported by their teachers are more likely to also feel safe and relaxed in class than their peers who reported feeling unsupported (Furrer & Skinner, 2003). Students who perceive their teachers as supportive also tend to report better psychological adjustment (Van Ryzin, Gravely, & Roseth, 2009), more positive affect and life satisfaction (Suldo, Shaffer, & Riley, 2008), and less self-consciousness while in school (Roeser et al., 1996). In contrast, students who do not feel supported by adults in school have lower self-esteem and less developed sense of identity (Ryan et al., 1994). Finally, two longitudinal studies showed that increases in students' perceptions of teacher support reliably predicted decreases in depressive symptomology over time (Pössel, Rudasill, Sawyer, Spence, & Bjerg, 2013; Reddy, Rhodes, & Mulhall, 2003). Collectively, results from these studies point to the importance of investigating what specific teaching behaviors are associated with well-being in students.

Multiple models of teaching behavior converge in the conceptualization of three components - instructional, socio-emotional, and organizational (Connor et al., 2009; Douglas, 2009; Pianta & Hamre, 2009) – each of which has been associated with students' academic and social success (e.g., Hamre & Pianta, 2005; Perry, Donohue, & Weinstein, 2007). Although classroom observation is the gold standard for measuring teaching behavior, this approach requires ample funding and time (Douglas, 2009). Teacher reports of their behavior, although cost effective, may not be accurate reflections of teaching behavior (Douglas, 2009), and some research suggests that students' perceptions of their teachers' behavior may be more valuable than third-party observer reports for understanding student outcomes (Eccles, Midgley, Buchanan, Wigfield, Reuman, & MacIver, 1993; Wubbels & Levy, 1991). However, there are few student-report measures of teaching behavior. Thus, the present studies have two purposes. The first is to develop a student-report assessment of teaching behavior that queries students' perceptions of teachers' concrete and specific behaviors within the areas of instructional, socio-emotional, and organizational support. The second is to test whether student-reported teaching behavior is associated with students' well-being.

Instructional, Organizational, and Socio-emotional Teaching Behavior

Mounting evidence links three broad components of teacher behavior (instructional, organizational, and socio-emotional) to students' academic and psychosocial adjustment. Instructional behavior is used during the delivery of instruction, with the intention of promoting concept or skill development and critical thinking (Croninger & Valli, 2009; Pianta, LaParo, & Hamre, 2008). Organizational behavior refers to structures established by the teacher to facilitate smooth transitions between activities, minimize disruptions, and efficiently use class time (Connor et al., 2009; Pianta et al., 2008). Teachers' socio-emotional behavior may be expressed at any time, during instruction or not, and is marked by warmth and responsiveness in interactions between students and teachers, and encourages students' feelings of belonging and acceptance in the classroom (Connor et al., 2009; Pianta et al., 2008).

Teaching behavior that is instructionally supportive (e.g., providing opportunities for students to respond, to choose, or to receive positive feedback) promotes academic achievement (Curby, Rudasill, Edwards, & Perez-Edgar, 2011; Hamre & Pianta, 2005). Organizational strategies designed to increase students' time on task and decrease disruptions have been linked to increased student engagement (Rimm-Kaufman, Curby, Grimm, Nathanson, & Brock, 2009) and more productive use of available instructional time across the school year (Cameron, Connor, & Morrison, 2005). Both instructional and organizational teaching behaviors promote academic achievement which, in turn, has been associated with less negative affect, but unassociated with positive affect (for a review see Soar & Soar, 1987). In addition, growing research shows that socio-emotional teaching behavior is positively linked to better academic performance (e.g., Hamre & Pianta, 2005; Rudasill, Gallagher, & White, 2010), a variety of psychosocial variables (e.g., more positive student-teacher relationships; Thijs, Koomen, & van der Leij, 2008), and mental health outcomes such as lower levels of anxiety and depression (Perry et al., 2007). Based on the tripartite model, both anxiety and depression are characterized by high negative affect and depression is also characterized by low positive affect (Clark & Watson, 1991). Thus, one can conclude that instructional, organizational, and socio-emotional teaching behaviors would be negatively associated with negative affect and socio-emotional teaching behavior would be positively associated with positive affect. However, as far as we know, no work has explored direct associations between teaching behavior and students' positive and negative affect.

Assessments of Teaching Behavior

While a multitude of studies has examined the influence of teachers on students (e.g., Crosnoe et al., 2004; Eisenhower, Baker, & Blacher, 2007), most studies assess teaching behavior via teacher report or classroom observations, each of which has its own strengths and weaknesses (Douglas, 2009). Although classroom observations by trained observers are the gold standard for measuring teaching behavior, these require ample investments of time and money (Douglas, 2009). In addition, classroom observations that measure quantity, rather than quality, of teaching behaviors (Pianta & Hamre, 2009) vary widely *within* teachers, suggesting that multiple observations may be necessary to capture typical behaviors (Croninger & Valli, 2009). Thus, instruments measuring teaching behavior using external observers are limited by practical considerations such as cost and time (for training, traveling to schools, observations; Achenbach,

McConaughy, & Howell, 1987). Teacher reports of teaching behavior are much less expensive than classroom observations (Douglas, 2009), yet these are plagued with problems of self-rating bias.

The weaknesses of classroom observations and teacher reports point to student report as a possible mode of assessing teaching behavior that maximizes observations of teaching behavior with minimal time and expense. Beyond that, multiple models propose that the use of student-ratings to measure teaching behavior is preferable because student perceptions and experiences are more critical for understanding and predicting student outcomes compared to other measures of teaching behavior, even when they are more objective (Eccles et al., 1993; Wubbels & Levy, 1991). That is, teaching behavior may be most meaningful when it is recorded as students perceive it. Indeed, students' perceptions of school and classroom processes (such as teacher support and school climate) are widely used in studies with adolescents (Reddy et al., 2003; Rueger, Malecki, & Demaray, 2010), and frequently associated with academic (Niehaus et al., 2012) and socio-emotional outcomes (Demaray, Malecki, Rueger, Brown, & Summers, 2009). Students' perceptions may provide insight into typical patterns of teaching behavior, as opposed to snapshot data gathered during discrete observations (Wubbels & Levy, 1991). In addition, student perceptions may allow the measurement of idiosyncratic teaching behavior, because students will be familiar with such behaviors while observers may not. However, to our knowledge, there are no studies testing the associations of students', teachers', and observers' ratings of teaching behavior with students' well-being.

Further, there are only a handful of assessments of teaching behavior that use student report. Among those, the behavior assessed is typically not specific (e.g., "He is strict," "He tries to make us look foolish;" Wubbels & Levy, 1991), or teaching behavior is combined with other constructs like classroom climate (Teddlie, Virgilio, & Oescher, 1990) or teacher expectations (Wubbels & Levy, 1991). This is problematic as these instruments might be influenced by student bias. As overt behaviors are more objective, behavior-oriented items should be more resistant to student bias (Keeley, Smith, & Buskist, 2006). In summary, the literature demonstrates a surprising lack of student-report instruments to measure concrete and specific teaching behavior. The aim of Study 1 is to develop the Teaching Behavior Questionnaire (TBQ-S), an instrument to assess students' perceptions of *concrete and specific* teaching behavior composed of instructional, socio-emotional, and organizational components. The aims of Study 2 are (a) to replicate the factor structure of the TBQ-S with a different sample and to provide further information about the internal consistencies of and intercorrelations between the TBQ-S scales in an independent sample and (b) to determine which TBQ scales are associated with students' well-being. It was predicted that all three TBQ-S scales would be negatively associated with negative affect and that socio-emotional behavior would be positively associated with positive affect.

Study 1

Method

Participants. Of the 1,429 students in grades 9 to 12 at a public high school in a medium-sized metropolitan area in the Southern United States, 1,200 volunteered to participate in this study (participation rate: 84%). Of the 1,122 students who reported their race/ethnicity, 50.8% were identified as Caucasian,

31.6% as African American, 8.5% as mixed race/ethnicity, and 9.1% as another race/ethnicity. Of the 1,138 students identifying their sex, 44% reported being male and 56% female. Of the 1,086 students who reported their grade level, 32% described themselves as being in the 9th grade, 28.3% in the 10th, 15.9% in the 11th and 23.8% in the 12th (mean age = 16.81, *SD* = 1.27). Of the students in this school, 51% were eligible for free or reduced lunch. There were no exclusion criteria, and students did not receive any incentive for their participation. This study was approved by the institutional review boards at the university and the public school system.

Measures.

Teaching Behavior Questionnaire (TBQ-S). The initial item pool of the TBQ-S consisted of items that had been used in past research to assess interpersonal teacher behavior (Questionnaire on Teacher Interaction; Wubbels & Levy, 1991), classroom management (instrument to assess teachers' classroom management styles; Bru, Stephens, & Torsheim, 2002), general behavior of college instructors (Teacher Behaviors Checklist; Keeley et al., 2006), and parental behavior (Alabama Parenting Questionnaire; Shelton, Frick, & Wootton, 1996). The instrument measuring parental behavior was included as children's relationships with parents and with teachers (Pianta, 1999) and the socialization of adolescents through parents and teachers is conceptually and empirically similar (e.g., Wentzel, 2002). To ensure the content validity of this initial item pool, a focus group composed of two teachers, one teacher educator, one school counselor educator, one principal, two parents, and two high school students were asked to (a) identify redundant or irrelevant items, (b) suggest corrections to items they viewed as difficult to understand, and (c) suggest additional items they believed are important to assess teaching behavior. A group of elementary school students was asked in one-on-one sessions to fill in a preliminary version of the questionnaire and to point out the items that they did not understand. Subsequently, these items were simplified by the authors with input from these elementary school students. This process resulted in a pool of 125 items where students rated the frequency of behaviors by a specific teacher towards themselves, with higher numbers indicating higher frequency of occurrence. Responses were set up with a four-point scale (1 = *never*, 4 = *always*).

Procedure. Upon approval by the Institutional Review Board for the protection of human subjects, the study was described to teachers during a faculty meeting, and consent forms from teachers were collected after the faculty meeting. Further, letters describing the study were sent to parents of all students in the participating school. Students who had parental consent were invited to participate. Student assent forms were collected at the beginning of the class period in which the questionnaires were administered. All students provided demographic information (such as age, gender, and ethnicity) and evaluated only *one* of their teachers using the TBQ-S. Random assignment was used to determine the specific teacher each individual student evaluated. Fifty-two of 72 teachers at the high school agreed to allow students to fill out questionnaires regarding their teaching behavior. Surveys from 703 students remained after accounting for student errors or obvious indifference to accurate completion of the questionnaire (i.e., prefabricated scantron forms with two extra answer bubbles than items in the questionnaire package were used; the excluded 497 students filled in all answer bubbles on the scantron form). No significant differences were found between

students that provided questionable data and the other students in terms of grade level (χ^2 (4) = 6.72, p = .151), race/ethnicity (χ^2 (6) = 8.32, p = .216), and sex (χ^2 (1) = 2.67, p = .102).

Statistical Analyses. Responses to the TBQ-S from 703 students were analyzed using SPSS 20 for Classical Test Theory (CTT) item difficulty, factor structure, part-whole item-scale correlation, and internal consistency. First, items with item means of less than 1.4 or greater than 3.6 ($p < .10$ or $p > .90$) were excluded (Schmid, 1992). Item difficulty reflects the agreement of all participants with each individual item. Thus, extreme item means (i.e., extreme item difficulty) suggests that a specific item does not differentiate between individual teachers. Next, exploratory factor analysis (EFA) with oblique (direct oblimin) rotation was conducted in order to identify unobservable latent factors of the TBQ-S (Preacher & MacCullum, 2003). Oblique rotation was selected as it allows the factors to correlate because the expected factors represent different components of teaching behavior (Connor et al., 2009; Douglas, 2009; Pianta & Hamre, 2009). To determine the number of factors to retain, scree test (Cattell, 1966), Kaiser-criterion (Kaiser, 1960), and parallel analyses (Hayton, Allen, & Scarpello, 2004; Horn, 1965) were used. In parallel analysis, eigenvalues are calculated with random data and then compared to those resulting from the factor analysis. Next, the random eigenvalues are compared to the eigenvalues of the factor analysis. When the eigenvalue generated by the factor analysis is higher than the eigenvalue generated by the parallel analysis, the factor accounts for more variance than what would be expected with random data, and the factor identified by the factor analysis is interpreted as meaningful. Nevertheless, it should be noted that parallel analysis results tend to indicate more factors than necessary (Hayton et al., 2004). Thus, the final factor solution was selected based on which factor structure theoretically made the most sense and which factor structure included the fewest items loading on multiple factors. Further, only items that loaded |0.4| or higher on one factor and that had loadings of |0.1| smaller on all other factors were retained (a criterion frequently used to determine importance of variables; Thompson, 2004). Finally, a part-whole item-scale correlation criteria of < |0.5| was used in order to determine which items to remove from each factor (Netemeyer, Bearden, & Sharma, 2003). In addition, the internal consistencies of (Barnette, 2005) and intercorrelations between the scores of the final TBQ-S scales were calculated.

Results

Based on item difficulties, 17 of 125 items were excluded. Next, the scree test of the EFA with the remaining 108 items suggested 2 factors and the Kaiser-criterion 1 factor. Results from the parallel analysis indicated that the eigenvalues of 14 factors generated by the factor analysis using data from student-reports were greater than in the model based on random data. Thus, EFAs with forced factor numbers were calculated with 2 to 14 factors. The six- and the eleven-factor solutions were the solutions with the lowest number of items that were excluded because they loaded on multiple factors or because they did not show a clear factor loading at all. Further, factors in the six-factor solution made theoretical sense. However, the separation of some of the factors in eleven-factor solution did not seem to make sense. For example, the items "... writes helpful comments on my returned work (for example. "Excellent spelling!" or "Let's practice some more subtraction.")" and "... helps me with my assignments for her class." loaded on separated factors

while other that seem not to fit together load on the same factor (e.g., "... talks with me before or after class." "The classroom rules are posted."). Thus, based on theoretical considerations and minimizing multiple factor loadings, a six-factor solution with 71 items clearly loading on one of the factors emerged while 29 items did not show a clear factor loading and were excluded[2]. Further, 19 items did not demonstrate a part-whole item-scale correlation of > |0.5| in the described high school student sample and were excluded. As a result, two factors were eliminated because one factor had no remaining items (F4) and one factor had only one remaining item (F6). Thus, 51 items and four factors were retained after the factor analysis. Finally, 14 items – all loading on factor 1 - were identified as insufficient in regard of measuring concrete and specific teaching behavior (e.g., "My teacher is excited about the subject area s/he teaches"). Therefore, these 14 items were eliminated from the final TBQ-S. This further reduction of the number of items in this factor kept the number of items manageable for participants.

The final TBQ-S includes four scales and a total of 37 items. The scale *Instructional Behavior* (13 items) indicates how a teacher manages the classroom and responds to individual students' needs (e.g. "My teacher uses examples that I understand" and "My teacher's grade requirements are clear"). *Negative Teaching Behavior* (9 items) indicates behaviors that students perceive as counter-productive or unpleasant (e.g. "My teacher threatens to punish me when I misbehave."). *Socio-Emotional Behavior* (10 items) indicates how well the teacher relates with students on a personal level (e.g. "My teacher talks with me about my interests"). *Organizational Behavior* (5 items) indicates how the teacher facilitates smooth transitions between activities and minimizes disruptions (e.g. "My teacher takes away a privilege if I abuse it"). Internal consistencies (Cronbach's α) for the four identified scales are presented in Table II, respectively.

The internal consistencies in the sample ranged from .78 (Organizational Behavior) to .97 (Instructional Behavior). The Skew Index of the TBQ-S items ranged from -.95 to 1.37. Because the Skew Indices of all items are below 3.0 (Kline, 2011), non-normality did not need to be addressed. The correlations between Instructional Behavior and Socio-Emotional Behavior ($r = .61$, $p < .01$) and Organizational Behavior ($r = .58$, $p < .01$) as well as the correlation between Socio-Emotional Behavior and Organizational Behavior ($r = .53$, $p < .01$) were high and positive. Correlations between Negative Teaching Behavior and Socio-Emotional Behavior ($r = .09$, $p < .05$) and Organizational Behavior ($r = .10$, $p < .05$) were also significant and positive, but much smaller in magnitude. The correlation between Negative Teaching Behavior and Instructional Behavior was the only significant and negative correlation ($r = -.23$, $p < .01$).

2 See endnote

Table I. Descriptive item statistics, factor loadings, and part-whole item-scale correlations for the 37 items of the final version of the TBQ-S

	Study 1										Study 2					
Item	**M**	**SD**	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**	**h^2**	**$r_{i(t-i)}$**	**M**	**SD**	**F1**	**F2**	**F3**	**F5**
Instructional Behavior																
1. My teacher makes sure I understand the material before moving to something new.	3.25	0.95	**.68**	-.17	-.31	-.37	.24	.31	.53	0.69	3.00	0.98	**.79**	-	-	-
2. My teacher answers my questions in a way that I understand.	3.46	0.83	**.74**	-.14	-.38	-.26	.19	.16	.58	0.70	3.07	0.96	**.84**	-	-	-
3. My teacher helps me with my assignments for her class.	3.25	0.95	**.62**	-.15	-.31	-.35	.38	.40	.48	0.65	3.02	0.94	**.82**	-	-	-
4. My teacher uses examples I understand.	3.34	0.84	**.70**	-.09	-.47	-.38	.21	.28	.57	0.68	3.06	0.93	**.84**	-	-	-
5. My teacher's grade requirements are clear.	3.49	0.84	**.66**	-.17	-.32	-.23	.28	.28	.44	0.66	3.21	0.93	**.83**	-	-	-
6. If I don't understand something, my teacher explains it another way.	3.27	0.91	**.68**	-.11	-.50	-.34	.35	.35	.54	0.71	3.06	0.94	**.69**	-	-	-
7. My teacher says that s/he appreciates participation, even if it is not always correct.	3.23	0.95	**.59**	-.15	-.43	-.28	.39	.36	.44	0.61	3.03	0.99	**.65**	-	-	-
8. My teacher returns my work quickly.	2.93	1.03	**.51**	-.05	-.39	-.36	.39	.36	.39	0.55	2.70	0.97	**.68**	-	-	-
9. My teacher starts class on time.	3.58	0.75	**.58**	-.27	-.27	-.04	.39	.25	.40	0.63	3.29	0.89	**.82**	-	-	-
10. My teacher stays on task.	3.46	0.75	**.66**	-.26	-.32	-.13	.46	.28	.51	0.65	3.15	0.90	**.74**	-	-	-
11. My teacher pays attention to me when I state my opinion.	3.44	0.86	**.71**	-.23	-.44	-.12	.38	.32	.55	0.72	3.13	0.97	**.77**	-	-	-
20. My teacher treats every student fairly.	3.51	0.80	**.67**	-.27	-.42	.03	.44	.27	.57	0.67	3.11	0.99	**.77**	-	-	-
23. My teacher's rules for class behavior are fair.	3.52	0.81	**.69**	-.26	-.37	.05	.51	.27	.62	0.68	3.21	0.92	**.74**	-	-	-
Negative Teaching Behavior																
26. Depending on her mood, my teacher may not follow through with consequences for misbehavior.	1.74	0.99	-.18	**.62**	-.01	-.08	-.02	-.08	.39	0.56	2.13	1.00	-	**.53**	-	-
27. Depending on her mood, my teacher threatens to punish me.	1.48	0.90	-.21	**.72**	-.08	-.13	.12	.01	.54	0.73	1.82	1.00	-	**.81**	-	-
28. My teacher threatens to punish me when I misbehave.	1.78	0.98	-.05	**.51**	-.15	-.17	.26	.13	.33	0.55	2.11	1.05	-	**.72**	-	-
29. My teacher threatens to punish me when I talk with my neighbor.	1.58	0.89	-.14	**.59**	-.10	-.17	.22	.10	.40	0.63	2.00	1.03	-	**.74**	-	-
30. My teacher's mood changes quickly.	1.72	0.95	-.28	**.61**	.03	-.08	.05	.06	.41	0.60	2.08	1.06	-	**.72**	-	-
31. My teacher does not know what to do next.	1.43	0.88	-.25	**.65**	.01	-.08	-.03	-.06	.43	0.64	1.64	1.00	-	**.70**	-	-

32. My teacher is easy to influence (e.g., it is easy to get my teacher to talk about something besides the class topic).	1.85	0.96	-.18	**.56**	-.09	-.02	-.02	-.06	.33	0.53	2.01	0.96	-	**.60**	-	-
34. My teacher is easy to provoke.	1.62	0.91	-.26	**.72**	-.04	-.14	.01	-.07	.53	0.70	1.92	0.97	-	**.67**	-	-
35. My teacher requires that I sit alone (e.g., in the hallway, in an empty room, with the librarian) if I am talking without permission during class time.	1.43	0.87	-.12	**.62**	-.06	-.20	.18	-.02	.43	0.62	1.70	0.99	-	**.67**	-	-
Socio-Emotional Behavior																
12. My teacher talks with me before or after class.	2.55	1.14	.39	.03	**-.59**	-.43	.29	.25	.44	0.62	2.47	1.06	-	-	**.76**	-
13. My teacher tells jokes or funny stories.	2.75	1.02	.40	.06	**-.69**	-.23	.14	.17	.49	0.61	2.58	1.02	-	-	**.76**	-
14. My teacher talks with me about school-related problems.	2.52	1.17	.38	.11	**-.73**	-.49	.38	.27	.62	0.73	2.35	1.03	-	-	**.77**	-
15. My teacher talks with me about non-school-related problems (e.g., at home, with other kids).	2.08	1.15	.26	.31	**-.68**	-.35	.22	.09	.54	0.66	1.98	1.05	-	-	**.68**	-
16. My teacher shows her emotions (e.g., changes tone of voice, uses facial expressions).	2.85	1.04	.22	.03	**-.47**	-.07	.26	.19	.25	0.51	2.69	1.02	-	-	**.51**	-
17. My teacher smiles at me.	3.07	1.00	.44	-.03	**-.59**	-.28	.46	.36	.47	0.62	2.81	0.98	-	-	**.71**	-
18. My teacher greets me.	3.07	1.03	.49	-.05	**-.61**	-.31	.45	.39	.51	0.62	2.88	1.02	-	-	**.70**	-
19. My teacher talks with me about my interests.	2.42	1.17	.35	.10	**-.75**	-.46	.34	.28	.64	0.74	2.31	1.08	-	-	**.78**	-
21. My teacher uses sarcasm in a funny way.	2.66	1.06	.38	.08	**-.64**	-.22	.10	.16	.44	0.55	2.67	1.07	-	-	**.71**	-
33. If I had a problem with school or life, this is my teacher I would ask for help	2.29	1.14	.37	.15	**-.61**	-.44	.33	.21	.46	0.59	2.24	1.06	-	-	**.68**	-
Organizational Behavior																
22. My teacher makes sure I understand the classroom rules.	3.38	0.91	.56	-.09	-.38	-.16	**.67**	.31	.57	0.56	3.20	0.89	-	-	-	**.75**
24. My teacher corrects me when I misbehave.	3.14	1.05	.38	.01	-.40	-.23	**.66**	.34	.49	0.66	3.13	0.97	-	-	-	**.77**
25. When I misbehave, my teacher explains to me why my behavior was wrong.	2.80	1.19	.41	.01	-.47	-.37	**.57**	.27	.47	0.57	2.81	1.05	-	-	-	**.75**
36. My teacher takes away a privilege if I abuse it.	2.38	1.02	.21	.23	-.30	-.32	**.54**	.15	.38	0.55	2.34	1.03	-	-	-	**.39**
37. My teacher does not let me use equipment if I use it improperly or dangerously.	2.83	1.16	.27	.04	-.23	-.26	**.57**	.13	.36	0.59	2.68	1.14	-	-	-	**.42**
Eigenvalue	-	-	30.48	11.15	3.54	3.13	2.66	2.15	-	-	-	-	-	-	-	-
Explained Variance	-	-	25.61	9.37	2.97	2.63	2.24	1.81	-	-	-	-	-	-	-	-

Note. M = mean, SD = standard deviation, F1 = Instructional Behavior, F2 = Negative Teaching Behavior, F3 = Socio-Emotional Behavior, F4 = factor 4, F5 = Organizational Behavior, F6 = factor 6, = h^2 = communalities, $r_{i(t-i)}$ = part-whole item-scale correlations. The item responses were set up with a four-point scale (1 = *never*, 4 = *always*) allowing the item scores to range from 1 to 4.

Table II. Descriptives and Internal Consistencies in Study 1 (N ≥ 558) and 2 (TBQ-S: N ≤ 822; TBQ-T & TBQ-O: N ≤ 72)

	Study 1			Study 2		
	α (CI_{95})	M	SD	α/*ICC* (CI_{95})	M	SD
TBQ-S						
Instructional Behavior	.97*** (.967-.973)	3.34	0.62	.95*** (.947-.955)	3.07	0.75
Negative Teaching Behavior	.88*** (.866-.893)	1.69	0.71	.89*** (.881-.900)	1.94	0.73
Socio-Emotional Behavior	.86*** (.844-.875)	2.62	0.77	.91*** (.906-.921)	2.49	0.77
Organizational Behavior	.78*** (.753-.805)	2.86	0.81	.77*** (.745-.788)	2.83	0.73
TBQ-T						
Instructional Behavior	-	-	-	.75*** (.655-.833)	3.51	0.32
Negative Teaching Behavior	-	-	-	.69*** (.550-.793)	1.53	0.39
Socio-Emotional Behavior	-	-	-	.88*** (.829-.919)	2.80	0.60
Organizational Behavior	-	-	-	.59*** (.398-.733)	3.12	0.54
TBQ-O						
Instructional Behavior	-	-	-	.92*** (.826-.989)	1.41	0.23
Negative Teaching Behavior	-	-	-	.66*** (.423-.852)	1.02	0.05
Socio-Emotional Behavior	-	-	-	.72*** (.544-.851)	1.43	0.26
Organizational Behavior	-	-	-	.68*** (.439-.820)	1.09	0.13
PANAS-C						
Positive Affect	-	-	-	.75*** (.655-.833)	40.68	14.43
Negative Affect	-	-	-	.69*** (.550-.793)	24.91	11.82

Note. TBQ-S = student-rated TBQ, TBQ-T = teacher-rated TBQ, TBQ-O = observer-rated TBQ, M = mean, SD = standard deviation, α = Cronbach's alpha, *ICC* = interrater reliability. *** $p < .001$.

Study 2

Method

Participants. Students and teachers of one public high school in a medium-sized metropolitan area in the Southern United States were invited to participate in this study. Of the 1,026 students at the school, 822 volunteered to participate in this study (participation rate: 80.1%). Of the participating students, 44.6% identified as Caucasian, 32.5% as African American, 7.3% as mixed race/ethnicity, 6.7% as Hispanic, 4.4% as another race/ethnicity, 2.9% as Asian/Pacific Islander, and 0.7% as Native American; 54.6% reported being male and 45.4% female. Further, 28.8% of the students described themselves as being in the 9th grade, 30.2% in the 10th, 20.8% in the 11th, and 19.9% in the 12th (mean age = 15.69, *SD* = 1.24). Of the students in this school 48% were eligible for free or reduced lunch. Of the 79 teachers at the high school 36 agreed to participate (participation rate: 45.6%). Of the participating teachers, 29.4% were male and 70.6% were female. Further, 22.2% of the participating teachers taught English, 13.9% Mathematics, 11.1% social science and technical classes (including machine tools, welding) each, 8.3% foreign languages, music (including band, chorus, orchestra), and science each, and 2.8% Business, Health, and Special Education each. There were no exclusion criteria, and neither students nor teachers received any incentive for their participation. This study was approved by the institutional review boards at the university and the public school system.

Measures

Teaching Behavior Questionnaire-Students (TBQ-S). The TBQ-S administered in Study 2 is identical with the final 37-item version of the TBQ-S administered in Study 1. Item descriptives and factor loadings for the TBQ-S in Study 2 are presented in Table I; scale descriptives and internal consistencies are presented in Table II. As in Study 1, responses were set up with a four-point scale (1 = *never*, 4 = *always*) and scores for the four TBQ-S scales were calculated by averaging the item scores.

Teaching Behavior Questionnaire-Teachers (TBQ-T). This version of the TBQ is identical to the TBQ-S, except that the TBQ-T asks teachers to rate their own behavior towards specific students. Scale descriptives and internal consistencies are presented in Table II. As with the TBQ-S, responses were set up with a four-point scale (1 = *never*, 4 = *always*) and the scores of the four TBQ-S scales were calculated by averaging the item scores.

Teacher Behavior Questionnaire-Observer (TBQ-O). The TBQ-O consists of the same items as the TBQ-S. External observers rated teaching behavior during two 18-minute periods of class time. Two observers rated teachers' behavior toward a particular student within a four-minute interval, then recorded their observations for two minutes. This procedure was repeated three times within each observation period. Later, each observer coded the observations using the TBQ-O. Thus, the possible range of each TBQ-O item ranged from never observed (= 1) to always observed (= 4) per observation period and observer, which is parallel to the possible range of the TBQ-S and TBQ-T items. The scores of the four TBQ-O scales were calculated by averaging the item scores across both observation periods (excluding items that measure teaching behavior which is only observable at the beginning or end of a class) and both observers for each

scale. Because some teaching behaviors are only observable at the beginning or end of a class (e.g., "... talks with me before or after class."), one of the observations from each observer dyad was always scheduled for the beginning of a class and the other observation was scheduled for the last 18 minutes of a class. The observers were one PhD-level psychologist and six graduate students in a counseling psychology program. Observers were trained in a different public high school until exact interrater reliability of 90% for each of the 37 items was reached. Scale descriptives and interrater reliabilities, based on double coding of all observation dyads at the scale level in the Study 2 sample, are presented in Table I.

Positive Affect and Negative Affect Scale for Children (PANAS-C). The PANAS-C (Laurent et al., 1999) is a 30-item self-report instrument including 15 items measuring positive affect (e.g., happy) and 15 items measuring negative affect (e.g., sad). Students responded based on their feelings during the past few weeks on a 5-point Likert scale (1 = *very slightly or not at all* to 5 = *extremely*). The stability of the PANAS-C ratings by the same students in different classrooms are $r = .57$ for the Positive Affect scale and $r = .61$ for the Negative Affect scale for all participating students ($N = 822$). While we are not aware of studies reporting short-term test-retest reliability of the PANAS-C, previous studies reporting test-retest reliabilities of the PANAS scales over only a few weeks found correlations similar to the ones reported in this study (Krohne, Egloff, Kohlmann, & Tausch, 1996; Watson, Clark, & Tellegen, 1988). Scale descriptives and internal consistencies are presented in Table II.

Procedures

After approval by the Institutional Review Board for the protection of human subjects, the study was described to teachers during a faculty meeting, and consent forms from teachers were collected after the faculty meeting. Further, letters describing the study were sent to parents of all students in the participating school. Students who had parental consent were invited to participate, and student assent forms were collected during the following two weeks of school by participating teachers.

Before the data collection began, two specific students of each of the 36 participating teachers were randomly selected by the researchers using a two-step procedure, resulting in a total selection of 72 students. First, of all classes a participating teacher taught, two classes were randomly selected. Next, of all students in these two classes, one student from each class was randomly selected. Both randomizations for all students took place before the data collection started using a random number generator (http://www.random.org). Inspection of the assignments revealed that no student was selected for more than one class; thus, each student in this smaller sample only reported on one teacher.

The first data collected were observer-rated teaching behavior of each of the participating teachers (n = 36) towards two specific students. To minimize the likelihood that the observations would impact the interaction between the selected teacher-student dyads, neither teachers nor students were told which students were selected. After completion of all observations, student- and teacher-ratings were collected. The TBQ-S was part of a questionnaire package administered to all participating 822 students. Demographic information (i.e., age, sex, and race/ethnicity) and the PANAS-C were also collected from these students. Students evaluated the behavior of multiple teachers towards themselves and also completed the PANAS-C during the

classes for the teachers they rated (mean number of ratings: 2, SD: 0.69, range: 1-4). Thus, a total of 1,255 student-ratings about mood and teaching behavior were collected. Teachers simultaneously completed the TBQ-T to rate their behavior towards the same two specific students who were part of the 72 teacher-student dyads rated using the TBQ-O. Thus, student-, teacher-, and observer-ratings of teaching behavior in the same 72 teacher-student dyads were collected. In these dyads the teacher behavior toward two students was rated by the teachers and external observers while each student rated only the behavior of one teacher towards them.

Statistical Analyses.

In order to test how well the four-factor model found in Study 1 applied to the sample of 822 students, confirmatory factor analysis (CFA) was performed on the TBQ-S with the maximum likelihood method using IBM Amos 20.0. In addition to the model with the four correlated factors found in Study 1, a model consistent with the originally predicted three-factor model was tested to ensure that the best fitting factor structure for multiple samples was used.

The χ^2 statistic, Comparative Fit Index (CFI; Bentler, 1990), Normed-Fit Index (NFI; Bentler & Bonnet, 1980), and root mean squared error of approximation (RMSEA; Steiger & Lind, 1980) were used to evaluate the goodness of fit of the models. Statistically non-significant values of χ^2 (Kline, 2011; Ullman, 1996) indicate a good fit of the model to the data. CFI and NFI values of $\geq .95$ indicate a good model fit and values of $\geq .90$ are regarded as acceptable (Hu & Bentler, 1999). An RMSEA value of 0 indicates a perfect model fit, values of .05 are considered a good model fit, and values of .08 are regarded as acceptable (Hu & Bentler, 1999). To compare models, ΔCFI was calculated by subtracting the CFI values of the models from each other. When ΔCFI is $> .002$ the model with higher CFI fits the data significantly better. However, when ΔCFI is $\leq .002$ both models fit equally well from a statistical point of view and the more parsimonious model should be accepted (Meade, Johnson, & Braddy, 2008). In addition, nested models were compared by subtracting the χ^2 values as well as the *df*s of the models from each other (χ^2 difference tests). When $\Delta\chi^2$ is significant for Δdf, the models are seen as significantly different from each other.

To examine the reliability of scores of the TBQ with different samples of students, internal consistencies (Cronbach´s α; Barnette, 2005) for each scale of all three versions of the TBQ and intercorrelations between all those TBQ scales were calculated with student-, teacher-, and observer-reports using SPSS 20.0.

Finally, to test which teaching behaviors are significantly associated with students' positive and negative affect, two-level hierarchical linear model (HLM) analyses were calculated using HLM version 6.07 (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2004). Independent of whether student-, teacher-, or observer-report of teaching behavior was used as the predictor, in all these analyses students were nested in teachers. To deal with collinearity (condition number $\kappa = 226.05$; Montgomery & Peck, 1992), the TBQ-S, TBQ-T, and TBQ-O scale scores served as predictors in three separate HLM analyses for each of the two dependent variables of interest, the PANAS-C Positive Affect scale and the Negative Affect scale scores.

Results

The CFAs using the data from all 822 students with the TBQ-S showed that RMSEA for the four-factor model [χ^2 (623, N = 1255) = 4634.32, $p < .001$, RMSEA (.072), CFI (.860), NFI (.842)] was within the acceptable range. The CFAs showed further that none of the indices of goodness of fit for the three-factor model [χ^2 (626, N = 1255) = 8898.16, $p < .001$, RMSEA (.103), CFI (.711), NFI (.696)] were within the acceptable range. Comparing both models with the ΔCFI test revealed that the correlated four-factor model fit the data of the TBQ-S in the sample of Study 2 better than the three-factor model. This finding was confirmed comparing both models ($\Delta\chi^2$ (3) = 4263.84, $p < .001$) using χ^2 difference tests. However, although the RMSEA of the four-factor model was acceptable, the χ^2, the CFI and NFI were not. Nevertheless, the results from CFAs provide some support for the four-factor model in the TBQ-S and, as such, that Negative Teaching Behavior can be seen by students as distinct from the other three components of teaching behavior that are already well established in the literature.

Internal consistencies of the scales of the three TBQ versions and intercorrelations between all four scales of all three TBQ versions are presented in Table II and III, respectively. Similar to the findings in Study 1, the internal consistencies of the TBQ-S in the sample ranged from .77 to .95, with Organizational Behavior the lowest and Instructional Behavior the highest. Most intercorrelations between the TBQ-S scales were positive, only the correlation between Instructional Behavior and Negative Teaching Behavior was negative. The pattern of internal consistencies and intercorrelations of the TBQ-S mirror the findings in Study 1 (Table II). The Skew Index of the TBQ-S items in Study 2 ranged from -.95 to 1.37. Because the Skew Indices of all items are below 3.0 (Kline, 2011), non-normality did not need to be addressed.

The internal consistencies of the TBQ-T ranged from .59 to .88, with Organizational Behavior the lowest and Socio-Emotional Behavior the highest. The pattern of the intercorrelations between the TBQ-T scales mirrors the intercorrelations between the TBQ-S scales. The only difference is that the direction of the non-significant intercorrelation between the Socio-Emotional and the Negative Teaching Behavior scales changes from positive to negative. The internal consistencies of the TBQ-O ranged from .66 to .92, with Negative Teaching Behavior the lowest and Instructional Behavior the highest. The most important difference between the intercorrelations between the TBQ-O scales and the intercorrelations between the scales of the other two TBQ versions is that the intercorrelations between most of the TBQ-O scales are very low.

Table III. Intercorrelations between the TBQ Scales in Study 2 (N ≤ 72)

	TBQ-S				TBQ-T				TBQ-O		
	Instr	**Neg**	**Socio**	**Org**	**Instr**	**Neg**	**Socio**	**Org**	**Instr**	**Neg**	**Socio**
TBQ-S											
Neg	-.10										
Socio	.77**	.08									
Org	.67**	.25	.62**								
TBQ-T											
Instr	*.28*	-.01	.11	.10							
Neg	-.09	*.00*	-.09	-.04	-.19						
Socio	.42**	-.06	*.41***	.23	.50**	-.04					
Org	.19	-.19	.00	*.16*	.30*	.06	.33				
TBQ-O											
Instr	*.23*	-.05	.27	.07	*.26**	-.27*	.33*	.19			
Neg	-.25	*-.33**	-.17	-.49**	.03	*.01*	.18	-.10	-.02		
Socio	.28	-.10	*.30**	.15	.20	-.12	*.32**	.17	.27*	-.01	
Org	-.08	-.19	-.07	*-.12*	-.14	.08	.11	*.11*	.02	.29*	-.03

Note. Intercorrelations in italics are correlations among the different reporters (students, teachers, and observers), TBQ-S = student-rated TBQ, TBQ-T = teacher-rated TBQ, TBQ-O = observer-rated TBQ, Instr = Instructional Behavior, Neg = Negative Teaching Behavior, Socio = Socio-Emotional Behavior, Org = Organizational Behavior. *** $p < .001$, ** $p < .01$, * $p < .05$.

Before calculating the correlations between the different versions of the TBQ using only data from the 72 student-teacher dyads, we tested whether the 72 students were significantly different from the other 752 students in terms of grade level ($\chi^2(4) = 0.62$, $p = .975$), race/ethnicity ($\chi^2(8) = 4.69$, $p = .790$), sex ($\chi^2(1) = 0.27$, $p = .602$), ratings on the TBQ-S Instructional Behavior ($t(789) = 0.68$, $p = .498$), Negative Teaching Behavior ($t(785) = 0.83$, $p = .408$), Socio-Emotional Behavior ($t(788) = 0.50$, $p = .614$), and Organizational Behavior ($t(787) = 0.41$, $p = .684$) scales, or the PANAS-C Positive Affect ($t(757) = 0.59$, $p = .554$) and Negative Affect ($t(756) = 0.77$, $p = .444$) scales. No significant differences between the subsample and the total sample were found regarding any of the variables.

Some of the intercorrelations presented in Table III are correlations among the different reporters (students, teachers, and observers). The correlations among the different reporters (i.e., students, teachers, and observers) of the Socio-Emotional Behavior scores were the highest and were all significant. Next, the correlations among the different reporters of the Instructional Behavior scores of the three TBQ versions are

consistently below .30 although only the correlations between TBQ-T and TBQ-O were statistically significant. The correlations of teacher-rating with student- and observer-rating of the Negative and Organizational Behavior scores are non-significant and close to zero. However, there was a significant negative correlation between the student- and observer-rating of negative behavior scores. The pattern of correlations between students, teachers, and observers understandably follows the internal consistencies. To be more precise, the correlations among the different reporters that include only scale scores with high internal consistencies have the highest correlations among the different reporters, while correlations among the different reporters that include one or two scale scores with low internal consistencies have low correlations among the different reporters. To test whether internal consistencies are the cause for these patterns, attenuation corrected correlations should be calculated.

The remaining intercorrelations are between scales of one version of the TBQ and the other scales of another version of the TBQ. As one could expect based on the fact that these intercorrelations are between different scales rated by different persons, they are mostly low and non-significant.

Using the data from all participating students and student-teacher dyads, six HLMs were conducted to regress Positive and Negative Affect scores on student-rated teaching behavior (TBQ-S; n = 822), teacher-rated teaching behavior (TBQ-T; n = 72), and observer-rated teaching behavior (TBQ-O; n = 72). Results showed that the Negative Teaching Behavior scale of the TBQ-S was negatively related to students' Positive Affect, and the Socio-Emotional Behavior scale of the TBQ-O was positively related to students' Positive Affect (Table IV). In addition, the Negative Teaching Behavior and the Socio-Emotional scales of the TBQ-S were positively related to students' Negative Affect, and the Instructional Behavior and the Organizational Behavior scales of the TBQ-S and the TBQ-O were negatively related to students' Negative Affect (Table V).

Discussion

Despite the fact that teaching behavior is associated with important psychosocial outcomes in students (e.g., Perry et al., 2007; Thijs et al., 2008) there is a surprising dearth of evaluated student-report instruments to measure concrete and specific teaching behaviors. To overcome this shortcoming, the purpose of these studies is to describe the development of an instrument to assess students' perceptions of *concrete and specific* teaching behavior, to report the psychometrics of this instrument, and to use that instrument to explore associations between teaching behavior and students' well-being.

Development of the TBQ-S

Data from two independent high school student samples informed the development of the 37-item TBQ-S, a questionnaire measuring instructional, negative teaching, socio-emotional, and organizational behaviors. These four scales reflect components of teaching behavior found in previous studies (Connor et al., 2009; Douglas, 2009; Pianta & Hamre, 2009). However, items indicative of negative teaching behavior were not expected to form a separate scale but to load onto each of the other scales. However, the factor structure with four correlated scales was somewhat supported in both samples, even when the confirmatory factor analyses revealed that only the RMSEA but not the CFI and the NFI of this model was acceptable.

Table IV. Estimated Fixed Effects of the TBQ-S, TBQ-T, and TBQ-O Scales on the PANAS-C Scale Positive Affect from three Separate HLMs

Parameter	Parameter estimate	SE
TBQ-S (*n* = 822)		
Instructional Behavior (γ_{10})	1.610	0.891
Negative Teaching Behavior (γ_{20})	-1.820**	0.629
Socio-Emotional Behavior (γ_{30})	4.506***	0.792
Organizational Behavior (γ_{40})	1.103	0.755
TBQ-T (*n* = 72)		
Instructional Behavior (γ_{10})	7.148	8.385
Negative Teaching Behavior (γ_{20})	-0.191	5.915
Socio-Emotional Behavior (γ_{30})	-0.159	4.501
Organizational Behavior (γ_{40})	2.274	4.351
TBQ-O (*n* = 72)		
Instructional Behavior (γ_{10})	-7.453	9.871
Negative Teaching Behavior (γ_{20})	-14.871	76.424
Socio-Emotional Behavior (γ_{30})	17.992	9.423
Organizational Behavior (γ_{40})	3.914	16.768

Note. TBQ-S = student-rated TBQ, TBQ-T = teacher-rated TBQ, TBQ-O = observer-rated TBQ.

When interpreting the level of the goodness-of-fit indices, Marsh, Hau, and Wen (2004) work demonstrating that the goodness-of-fit indices are very good at distinguishing between more and less misspecified models, but not as helpful for providing absolute guidelines about the acceptability of a particular model should be considered. Instead, Marsh et al. (2004) recommend considering fit indices together with expectations based on theory and sample characteristics. Nevertheless, to strengthen the confidence in the correlated four-factor model, future studies should test the factor structure in diverse samples. So far, the samples were limited to public high schools. Thus, as the setting and the age of students might impact teaching behavior and student ratings of teaching behavior, private schools and elementary and middle-school samples should be included in these replication studies. Further, cultural factors might impact teaching behavior and student-ratings of teaching behavior as well. For example, in East Asia teachers are highly valued and respected, which leads to a hierarchical interaction between students and teachers (Chen & Wei, 2011). Thus, samples from countries other than the US, different cultures within the US, or in different settings could contribute important information about the factor structure of the TBQ-S.

Table V. Estimated Fixed Effects of the TBQ-S, TBQ-T, and TBQ-O Scales on the PANAS-C Scale Negative Affect from three Separate HLMs.

Parameter	Parameter estimate	SE
TBQ-S (*n* = 822)		
Instructional Behavior (γ_{10})	-1.873*	0.752
Negative Teaching Behavior (γ_{20})	2.264***	0.532
Socio-Emotional Behavior (γ_{30})	2.312***	0.668
Organizational Behavior (γ_{40})	-1.9135**	0.638
TBQ-T (*n* = 72)		
Instructional Behavior (γ_{10})	7.168	4.873
Negative Teaching Behavior (γ_{20})	5.249	3.473
Socio-Emotional Behavior (γ_{30})	-0.377	2.589
Organizational Behavior (γ_{40})	-3.411	2.485
TBQ-O (*n* = 72)		
Instructional Behavior (γ_{10})	-15.252*	6.414
Negative Teaching Behavior (γ_{20})	15.537	49.370
Socio-Emotional Behavior (γ_{30})	7.557	6.150
Organizational Behavior (γ_{40})	-24.497*	10.973

Note. TBQ-S = student-rated TBQ, TBQ-T = teacher-rated TBQ, TBQ-O = observer-rated TBQ. * $p < .05$; ** $p < .01$; *** $p < .001$.

Further, it is surprising that items indicative of negative teaching behaviors formed a separate factor instead of loading on the other three factors already well-established in the literature (Connor et al., 2009; Douglas, 2009; Pianta & Hamre, 2009). Although unexpected, it makes sense that students, rather than outside observers as mainly used in the previous literature, might group behaviors indicative of poor teaching together, given their personal investment in what happens in the classroom. Whereas an outside observer is removed from the consequences of negative teaching behaviors, students directly experience the repercussions of such behavior, perhaps resulting in a reaction to it that is, first and foremost, negative. Further, the distribution of all items loading on the Negative Teaching Behavior TBQ-S scale and the TBQ-S scale scores are positively skewed in both studies. This reflects what one would hope to see as this likely means that students experience only a low frequency of negative teaching behavior in classrooms.

Not surprisingly, the pattern of correlations among the different reporters (students, teachers, and observers) follows the internal consistencies. In other words, the correlations among the different reporters

that include only scale scores with high internal consistencies have the highest correlations among the different reporters, while those that include one or two scale scores with low internal consistencies have low correlations among the different reporters. Similarly, the intercorrelations of the different scales within the same version of the TBQ seem generally consistent across the three TBQ versions and both studies. Intercorrelations between the Instructional Behavior, Socio-Emotional Behavior, and Organizational Behavior scales are positive and tend to be significant. However, intercorrelations of the TBQ-O Organizational Behavior scale with the Instructional and the Socio-Emotional Behavior scales are non-significant. Thus, it seems observers see organizational behavior as more different from the other two scales than students and teachers do. Indeed, the intercorrelations of the TBQ-O suggests that observers' perceptions of teacher behavior are more orthogonal than students' or teachers' ratings, suggesting that the history of the student-teacher interactions, as well as personal characteristics and perhaps bias inherent in assessment of proximal experiences, informs students' and teachers' global view of teacher behavior.

Next, the intercorrelations of the Negative Teaching Behavior scales with the other scales are negative but tend to be non-significant across the three TBQ versions in Study 2. However, the intercorrelations between the TBQ-S Negative Teaching Behavior scale and the other TBQ-S scales in Study 1 are also negative, but significant. The likely reason for this difference is that the sample size in Study 1 was much largely than the sample size in Study 2.

The negative associations in both samples between the TBQ-S negative teaching behavior and instructional behavior scales seem to make sense as one would expect that behavior students perceive as counter-productive or unpleasant (negative teaching behavior) and behavior that is demonstrated in response to students' needs (instructional behavior) are contrary. However, the positive associations of negative teaching behavior with socio-emotional and organizational behavior are not so clear. One possible explanation for the positive association between the TBQ-S Negative Teaching Behavior and Socio-Emotional Behavior scales is that some of the teaching behaviors that are commonly perceived as negative may actually make a teacher appear more responsive to students [e.g., "My teacher is easy to influence (e.g., it is easy to get my teacher to talk about something besides the class topic)."]. Similarly, the Negative Teaching Behavior and Organizational Behavior scales contain items that describe negative consequences of misbehavior (e.g., for negative teaching behavior: "My teacher threatens to punish me when I misbehave;" and for organizational behavior: "My teacher takes away a privilege if I abuse it."). Thus, it is understandable that students may perceive both types of behaviors as equally punishing and negative (Weinstein, 2003).

Teaching Behavior and Student Well-Being

The majority of relations between teacher behavior and student well-being were in line with hypotheses. For example, higher levels of student-reported negative teaching behavior were associated with less positive affect, whereas higher levels of socio-emotional teaching behavior were associated with more positive affect. Further, higher levels of negative teaching behavior were associated with more negative affect, whereas higher levels of both instructional and organizational teaching behavior were associated with less negative affect. However, the finding that student perceptions of more social-emotional behavior by

teachers were associated with more negative affect was not expected. This finding might be caused by the cross-sectional design of Study 2. It is very likely that students in a negative emotional state seek socio-emotional support from their teachers, and may therefore be more attuned to it. However, previous research has shown iatrogenic effects of emotional support from teachers of students with low levels of stress in a 5-year longitudinal study with high school students (Pössel et al., 2013). Pössel et al. (2013) hypothesize that the iatrogenic effect of teacher support can be explained by the *deviancy training hypothesis*, such that students are role models for each other, learning to attend to or describe negative affect. Regarding our finding in Study 2, this could mean that socio-emotional teaching behavior provides a safe space for such learning as it allows students to express their negative affect in the classroom. Nevertheless, the finding that socio-emotional teaching behavior might have iatrogenic effects on students' negative affect needs to be interpreted cautiously until explored further.

Although the focus of these studies was the development of a student-rated questionnaire on teaching behavior (TBQ-S), it is important to point out that scores based on observations of teaching behavior were good predictors of students' well-being as well. Indeed, observations of teaching behavior indicated that instructional and organizational behaviors were negatively associated with students' negative affect. Instructional and organizational teaching behaviors are features of classrooms where interactions and activities are intentional, and students know what to expect (e.g., Pianta & Hamre, 2009). It is likely that students in classrooms where instructional and organizational teaching behavior are high are more engaged in learning (see Fredericks, Blumenfeld, & Paris, 2004 for a review; Ryan & Patrick, 2001), and academic engagement has been consistently linked to positive outcomes for students, such as higher achievement (Connell, Spencer, & Aber, 1994; Fredericks et al., 2004; Marks, 2000), and lower likelihood of dropping out (Fredericks et al., 2004). A lack of engagement, on the other hand, is linked to disaffection, such as worry, sadness, and withdrawal (Skinner, Furrer, Marchand, & Kindermann, 2008). The fact that these associations emerged only from the TBQ-O ratings suggest that a third-party view provides a glimpse of important aspects of teaching behavior that is untainted by bias inherent in teacher or student ratings.

Limitations

Several limitations should be noted. First of all, the overall model fit was not particularly strong. Thus, the four-factor structure of the TBQ-S needs to be replicated. Another limitation of the TBQ-S is that it is unclear whether the instrument can be used reliably in all classroom settings. For example, Keeley et al. (2006) point out that it might be unrealistic to expect a teacher in a very large class to know all students' names; however typical secondary classrooms have an average of 24.3 students (Organization for Economic Cooperation and Development, 2009). In addition, it is unknown whether a physical education teacher displays all the same (expected) behaviors as teachers of other subjects (e.g. English, mathematics). While data from classroom settings with large classes were collected in Study 2 (e.g., orchestra, band), because of the small number of teachers with these classes, it was not possible to test for the effect of class size. The large percentage of questionable data in Study 1 can be seen as another limitation. However, the lack of differences on demographic variables between the students providing acceptable and questionable data

suggests that the acceptable data may still be representative of the student body, at least in terms of key demographic variables. Another limitation is that no interviews with students were conducted to explore how students were interpreting items and response choices. Further, the low internal consistency of some of the TBQ-T and the TBQ-O scales (i.e., Negative Teaching Behavior, Organizational Behavior) and the fact that no factor analyses were calculated in Study 2 with the TBQ-T and TBQ-O are additional limitations. Thus, it is possible that these two versions of the TBQ have another factor structure than the TBQ-S. Another limitation is that some students did report on multiple teachers in Study 2, but we were unable to account for that in our analyses. Finally, no evidence for the temporal stability of the observed teaching behavior or about the longitudinal relationships between TBQ-S scales and students' well-being were collected. Thus, longitudinal studies with large numbers of participating teachers are needed.

To summarize, based on data from two independent studies, the 37-item TBQ-S, a questionnaire measuring student-reports of teachers' instructional, negative teaching, socio-emotional, and organizational behavior was developed. Further, associations of the individual teaching behaviors with positive and negative affect in students were measured. To be more precise, higher levels of negative teaching behavior measured with the TBQ-S was associated with less positive and more negative affect in students. Further, more student-reported social-emotional teaching behavior was associated with more positive and more negative affect in students. Finally, higher levels of both instructional and organizational teaching behavior were not significantly associated with positive affect but inversely associated with negative affect in students. Because the negative teaching behaviors measured are concrete and specific, it is possible to train teachers to use more adaptive behaviors that are empirically linked to positive student outcomes (Hamre et al., 2012). Thus, if the association between teaching behavior and well-being are replicated in longitudinal studies, attempts to improve well-being in students by changing teaching behavior seem possible.

Acknowledgements

We would like to thank the students and teachers involved in this project for making this project possible. The described publication was made possible by two grants both to the first two authors from the College of Education and Human Development, University of Louisville.

References

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. Psychological Bulletin, 101(2), 213-232.

Barnette, J. J. (2005). ScoreRel CI: An EXCEL program for computing confidence intervals for commonly used score reliability coefficients. Educational and Psychological Measurement, 65(6), 980-983.

Bentler, P. M. (1990). Comparative fit indexes in structural models. Psychological Bulletin, 107(2), 238-246.

Bentler, P. M., & Bonnet, D. C. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88(3), 588-606.

Bru, E., Stephens, P., & Torsheim, T. R. (2002). Students' perceptions of class management and reports of their own misbehavior. Journal of School Psychology, 40(4), 287-307.

Cameron, C. E., Connor, C. M., & Morrison, F. J. (2005). Effects of variation in teacher organization on classroom functioning. Journal of School Psychology, 43(1), 61-85.

Cattell, R. B. (1966). The scree test for the number of factors. Multivariate Behavioral Research, 1(2), 245-276.

Chen, J. K., & Wei, H. S. (2011).The impact of school violence on self-esteem and depression among Taiwanese junior high school students. Social Indicator Research, 100(3), 479-498.

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. Journal of Abnormal Psychology, 100(3), 316-336.

Connell, J. P., Spencer, M. B., & Aber, J. L. (1994). Educational risk and resilience in African_American youth: Context, self, action, and outcomes in school. Child Development, 65(2), 493-506.

Connor, C. M., Morrison, F. J., Fishman, B. J., Ponitz, C. C., Glasney, S., Underwood, P. S., ... Schatschneider, C. (2009). The ISI Classroom Observation System: Examining the literacy instruction provided to individual students. Educational Researcher, 38(2), 85-99.

Croninger, R. G., & Valli, L. (2009). “Where is the action?” Challenges to studying the teaching of reading in elementary classrooms. Educational Researcher, 38(2), 100-108.

Crosnoe, R., Johnson, M. K., & Elder, G. H., Jr. (2004). Intergenerational Bonding in School: The Behavioral and Contextual Correlates of Student-Teacher Relationships. Sociology of Education, 77(1), 60-81.

Curby, T. W., Rudasill, K. M., Edwards, T., & Perez-Edgar, K. (2011). The role of classroom quality in ameliorating the academic and social risks associated with difficult temperament. School Psychology Quarterly, 26(2), 175-188.

Davis, H. (2006). Exploring the contexts of relationship quality between middle school students and teachers. The Elementary School Journal, 106(3), 193-223.

Demaray, M. K., Malecki, C. K., Rueger, S. Y., Brown, S. E., & Summers, K. H. (2009). Importance of socially supportive behaviors in the relationship between social support and self-concept. Journal of Youth and Adolescence, 38(1), 13-28.

Douglas, K. (2009). Sharpening our focus in measuring classroom instruction. Educational Researcher, 38(7), 518-521.

Eccles, J. S., Midgley, C., Buchanan, C. M., Wigfield, A., Reuman, D., & MacIver, D. (1993). Development during adolescence: The impact of stage/environment fit. American Psychologist, 48(2), 90-101.

Eisenhower, A. S., Baker, B. L., & Blacher, J. (2007). Early student-teacher relationships of children with and without intellectual disability: Contributions of behavioral, social, and self-regulatory competence. Journal of School Psychology, 45(4), 363-383.

Fredericks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. Review of Educational Research, 74(1), 59-109.

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.

Hamre, B. K., & Pianta, R. C. (2005). Can Instructional and Emotional Support in the First-Grade Classroom Make a Difference for Children at Risk of School Failure? Child Development, 76(5), 949-967.

Hamre, B. K., Pianta, R. C., Burchinal, M., Field, S., LoCasale-Crouch, J., Downer, J. T., . . . Scott-Little, C. (2012). A course on effective teacher-child interactions: Effects on teacher beliefs, knowledge, and observed practice. American Educational Research Journal, 49(1), 88-123. http://dx.doi.org/10.3102%2F0002831211434596

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. Organizational Research Methods, 7(2), 191-205.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. Psychometrika, 30(2), 179-185.

Hu, L., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6(1), 1-55.

Huebner, E. S., & Dew, T. (1996). The interrelationships of positive affect, negative affect and life satisfaction in an adolescent sample. Social Indicators Research, 38(2), 129-137.

Humensky, J., Kuwabara, S. A., Fogel, J., Wells, C., Goodwin, B., & Van Voorhees, B. W. (2010). Adolescents with depressive symptoms and their challenges with learning in school. The Journal of School Nursing, 26(5), 377-392.

Jonsson, U., Bohman, H., Hjern, A., von Knorring, L., Olsson, G., & von Knorring, A-L. (2010). Subsequent higher education after adolescent depression: A 15-year follow-up register study. European Psychiatry, 25(7), 396-401.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20(1), 141-151.

Keeley, J., Smith, D., & Buskist, W. (2006). The Teacher Behaviors Checklist: Factor Analysis of Its Utility for Evaluating Teaching. Teaching of Psychology, 33(2), 84-91.

Kline, R. B. (2011). Principals and practice of structural equation modeling (3rd ed.). New York, NY: Guilford.

Krohne, H. E., Egloff, B., Kohlmann, C-W., & Tausch, A. (1996). Untersuchungen mit einer deutschen Version der 'Positive and Negative Affect Schedule' (PANAS). [Investigations with a German version of the Positive and Negative Affect Schedule' (PANAS).] Diagnostica, 42(2), 139-156.

Laurent, J., Catanzaro, S. J., Joiner, T. E. Jr., Rudolph, K. D., Potter, K. I., Lambert, S., … Gathright, T. (1999). A measure of positive and negative affect for children: Scale development and preliminary validation. Psychological Assessment, 11(3), 326-338.

Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. American Educational Research Journal, 37(1), 153-184.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indixes and dangers in overgeneralizing Hu and Bentler's (1999) findings. Structural Equation Modeling, 11(3), 320 -341.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. Journal of Applied Psychology, 93(3), 568-592.

Montgomery, D. C., & Peck, E. A. (1992). Introduction to Linear Regression Analysis. John Willy: New York, NY.

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). Scaling procedures, issues and applications. Sage: Thousand Oaks, CA.

Niehaus, S. K., Rudasill, K. M., & Rakes, C. (2012). A longitudinal study of school connectedness and academic outcomes across sixth grade. Journal of School Psychology, 50(4), 443-460.

Organization for Economic Cooperation and Development. (2009). Education at a glance 2009. Paris: Author.

Patrick, H., Ryan, A., & Kaplan, A. (2007). Early adolescents' perceptions of the classroom social environment, motivational beliefs, and engagement. Journal of Educational Psychology, 99(1), 83-98.

Perry, K. E., Donohue, K. M., & Weinstein, R. S. (2007). Teaching practices and the promotion of achievement and adjustment in first grade. Journal of School Psychology, 45(3), 269-292.

Pianta, R. (1999). Enhancing relationships between children and teachers. Washington, DC: American Psychological Association.

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.

Pianta, R. C., LaParo, K. M., & Hamre, B. K. (2008). Classroom Assessment Scoring System. Baltimore, MD: Brookes.

Pössel, P., Rudasill, K. M., Sawyer M. G., Spence, S. H., & Bjerg, A. C. (2013). Associations between teacher support and depression in Australian adolescents. A 5-year longitudinal study. Developmental Psychology, 49(11), 2135-2146.

Preacher, K. J., & MacCullum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. Understanding Statistics, 2(1), 13-43.

Raudenbush, S., Bryk, A., Cheong, Y. F., Congdon, R., & du Toit, M. (2004). HLM6: Hierarchical linear and non-linear modeling. Lincolnwood, IL: Scientific Software International.

Reddy, R., Rhodes, J., & Mulhall, P. (2003). The influence of teacher support on student adjustment in the middle school years: A latent growth curve study. Development and Psychopathology, 15(1), 119-138.

Rimm-Kaufman, S. E., Curby, T. W., Grimm, K., Nathanson, L., & Brock, L. L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. Developmental Psychology, 45(4), 958-972.

Roeser, R., Midgley, C., & Urdan, T. (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.

Rudasill, K. M., Gallagher, K., & White, J. M. (2010). Temperamental attention and activity, classroom emotional support, and academic achievement in third grade. Journal of School Psychology, 48(8), 113-134.

Rueger, S. Y., Malecki, C. K., & Demaray, M. K. (2010). Relationship between multiple sources of perceived social support and psychological and academic adjustment in early adolescence: Comparisons across gender. Journal of Youth and Adolescence, 39(1), 47-61.

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.

Ryan, R., Stiller, J., & Lynch, J. (1994). Representations of relationships to teachers, parents, and friends as predictors of academic motivation and self-esteem. Journal of Early Adolescence, 14(2), 226-249.

Schmid, H. (1992). Psychologische Tests: Theorie und Konstruktion [Psychological tests: theory and construction]. Universitätsverlag Fribourg, Switzerland.

Shelton, K. K., Frick, P. J., & Wootton, J. (1996). Assessment of parenting practices in families of elementary school-age children. Journal of Clinical Child Psychology, 25(3), 317-329.

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.

Soar, R. S., & Soar, R. M. (1987). Classroom management and affect expression. Professional School Psychology, 2(1), 3-14.

Steiger, J. H., & Lind, J. M. (1980, May). Statistically based tests for the number of common factors. Paper presented at the Psychometrika Society meeting, Iowa City, Iowa.

Suldo, S. M., Shaffer, E. J., & Riley, K. N. (2008). A social-cognitive-behavioral model of academic predictors of adolescents' life satisfaction. School Psychology Quarterly, 23(1), 56-69.

Teddlie, C., Virgilio, I., & Oescher, J. (1990). Development and validation of the Virgilio Teacher Behavior instrument. Educational and Psychological Measurement, 50(2), 421-430.

Thijs, J. T., Koomen, H. M. Y., & van der Leij, A. (2008). Teacher-child relationships and pedagogical practices: Considering the teacher's perspective. School Psychology Review, 37(2), 244-260.

Thompson, B. (2004). Exploratory and confirmatory factor analysis: Understanding concepts and applications. Washington, DC: American Psychological Association.

Ullman, J. B. (1996). Structural equation modeling. In B. Tabachnick & L. Fidell (Eds.), Using multivariate statistics (3rd ed., pp. 709-812). New York, NY: HarperCollins.

Van Ryzin, M., Gravely, A., & Roseth, C. (2009). Autonomy, belongingness, and engagement in school as contributors to adolescent psychological well-being. Journal of Youth & Adolescence, 38(1), 1-12.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal of Personality and Social Psychology, 54(6), 1063-1070.

Weinstein, C. S. (2003). Secondary classroom management: Lessons from research and practice. Boston: McGraw Hill.

Wentzel, K. R. (2002). Are effective teachers like good parents? Interpersonal predictors of school adjustment in early adolescence. Child Development, 73, 287-301.

Wubbels, T., & Levy, J. (1991). A comparison of interpersonal behavior of Dutch and American teachers. International Journal of Intercultural Relations, 15(1), 1-18.

Endnote

2 factor solution: 13 items loaded on multiple factors, 18 items did not show a clear factor loading

3 factor solution: 20 items loaded on multiple factors, 14 items did not show a clear factor loading

4 factor solution: 20 items loaded on multiple factors, 17 items did not show a clear factor loading

5 factor solution: 19 items loaded on multiple factors, 14 items did not show a clear factor loading

6 factor solution: 12 items loaded on multiple factors, 17 items did not show a clear factor loading

7 factor solution: 18 items loaded on multiple factors, 14 items did not show a clear factor loading

8 factor solution: 16 items loaded on multiple factors, 16 items did not show a clear factor loading

9 factor solution: 17 items loaded on multiple factors, 17 items did not show a clear factor loading

10 factor solution: 19 items loaded on multiple factors, 13 items did not show a clear factor loading

11 factor solution: 16 items loaded on multiple factors, 13 items did not show a clear factor loading

12 factor solution: 18 items loaded on multiple factors, 15 items did not show a clear factor loading

13 factor solution: 22 items loaded on multiple factors, 15 items did not show a clear factor loading

14 factor solution: 23 items loaded on multiple factors, 14 items did not show a clear factor loading