WCER Working Paper No. 2006-7
October 2006

The Teacher Sense of Efficacy Scale: Validation Evidence and Behavioral Prediction

Herbert G. Heneman III
Consortium for Policy Research in Education
Wisconsin Center for Education Research
University of Wisconsin–Madison
hheneman@bus.wisc.edu

Steven Kimball
Consortium for Policy Research in Education
Wisconsin Center for Education Research
University of Wisconsin–Madison
skimball@education.wisc.edu

Anthony Milanowski
Consortium for Policy Research in Education
Wisconsin Center for Education Research
University of Wisconsin–Madison
amilanow@wisc.edu

WCER
Wisconsin Center for Education Research
School of Education • University of Wisconsin–Madison • http://www.wcer.wisc.edu/

Copyright © 2006 by Herbert G. Heneman III, Steven Kimball, and Anthony Milanowski
All rights reserved.

Readers may make verbatim copies of this document for noncommercial purposes by any means, provided that the above copyright notice appears on all copies.

WCER working papers are available on the Internet at http://www.wcer.wisc.edu/publications/workingPapers/index.php. Recommended citation:

> Heneman, H. G., III, Kimball, S., & Milanowski, A. (2006, October). *The teacher sense of efficacy scale: Validation evidence and behavioral prediction* (WCER Working Paper No. 2006-7). Madison: University of Wisconsin–Madison, Wisconsin Center for Education Research. Retrieved [e.g., October 15, 2006,] from http://www.wcer.wisc.edu/publications/workingPapers/papers.php

The research reported in this paper was supported in part by a grant from the U.S. Department of Education, Office of Educational Research and Improvement, National Institute on Educational Governance, Finance, Policy-Making and Management, to the Consortium for Policy Research in Education (CPRE) under Grant No. OERI-R308A60003 and by the Wisconsin Center for Education Research, School of Education, University of Wisconsin–Madison. Any opinions, findings, or conclusions expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies, WCER, or cooperating institutions.

The Teacher Sense of Efficacy Scale: Validation Evidence and Behavioral Prediction[1]

Herbert G. Heneman III, Steven Kimball, and Anthony Milanowski

Individual perceived self-efficacy (SE) is an enduring and widely researched construct. Bandura (1997) defined SE as "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (p. 3). As Bandura asserts, one's SE is specific to a particular task or role; it is not a generalized personality trait that guides behavioral choices in all situations.

Researchers interested in identifying the determinants of teacher classroom behavior and effectiveness adapted the concept of SE to the teaching context by exploring teacher self-efficacy (TSE). A number of studies have linked measures of TSE perceptions to important outcomes related to teacher performance and student achievement (see, for example, Ashton & Webb, 1986; Bandura, 1997; Tschannen-Moran, Hoy, & Hoy, 1998). While definitions of TSE have varied over the course of this line of research, the one used by Tschannen-Moran et al. (1998) fits best with Bandura's definition. According to Tschannen-Moran et al., TSE is "the teacher's belief in his or her capability to execute courses of action required to successfully accomplish a specific teaching task in a particular context" (p. 233). According to this definition, TSE is not a generalized trait of teachers, but rather is specific to the instructional task and context.

Researchers have engaged in intensive attempts to develop a valid measure of TSE. The history of those attempts, and a critique of them, are presented by Tschannen-Moran et al. (1998). According to this review, there has been a continuing and evolving interplay between conceptualizations of TSE and its measurement. Numerous measures of TSE have been developed over the years as an outgrowth of that interplay, including the frequently employed scale developed by Gibson and Dembo (1984). But close scrutiny of those measures and the accumulative research evidence on the construct validity of the measures has called into question their meaningfulness and usefulness for continued research (Tschannen-Moran & Hoy, 2001). Based on their historical review and critique of this research, Tschannen-Moran and Hoy (2001) concluded that "after nearly a quarter of a century of work on teacher efficacy, it seems apparent that a new measure of teacher efficacy that is both reliable and valid is needed" (p. 795).

A vexing problem in the measurement of TSE involves the level of specificity needed relating to the range of tasks and contexts within which the TSE belief is gauged. As pointed out by Tschannen-Moran et al. (1998), "in order to be useful and generalizable, measures of teacher efficacy need to tap teachers' assessments of their competencies across the wide range of activities and tasks they are expected to perform" (p. 219). Based on this premise, Tschannen-Moran et al. proposed a model in which TSE is a joint function of both the teacher's analysis of the teaching task and the teacher's assessment of personal teaching competence. This model provided the conceptual underpinnings for the development of a new measure of TSE referred to as the Teacher Sense of Efficacy Scale, or TSES, which was previously called the Ohio State Teacher Efficacy Scale (Tschannen-Moran & Hoy, 2001).

---

[1] An earlier version of this paper was presented at the annual meeting of the American Educational Research Association, April 15, Montreal, Quebec.

The TSES was carefully developed through a series of item development, item selection, and factor analysis-revision cycles, using numerous students and teachers to generate and critique the items. The final instrument taps teacher efficacy judgments in three contexts or domains: instructional strategies, classroom management, and student engagement. It uses a 9-point response scale, and the responses are anchored with the descriptors *1-nothing, 3-very little, 5-some influence, 7-quite a bit,* and *9-a great deal*. The TSES includes a long form (24 items) and a short form (12 items). The latter comprises the 4 items from each domain that have the highest factor loadings on the domain. The psychometric properties of the short form of the TSES are nearly identical to those of the long form (Tschannen-Moran & Hoy, 2001).

Construct validity analyses and results for the long and short forms of the TSES led Tschannen-Moran and Hoy (2001) to conclude that they "could be considered reasonably valid and reliable" (p. 801) and superior in content to the previously developed measures of TSE. In terms of content validity, the authors concluded that "the three dimensions of efficacy for instructional strategies, student engagement, and classroom management represent the richness of teachers' work lives and the requirements of good teaching" (p. 801). Despite these favorable conclusions, they were careful to call for additional testing and validation of the TSES. The research reported here is a response to that call.

The present study contributes to knowledge of the construct validity of the short form of the TSES (and by extension, given their similar content and psychometric properties, to the long form). As elaborated below, our research involves (a) examining the psychometric properties of the TSES on a large sample of elementary, middle, and high school teachers, and comparing our results to those reported by Tschannen-Moran and Hoy (2001), and (b) conducting a longitudinal analysis (predictive validity) of the TSES as a predictor of subsequent teacher classroom performance and student value-added learning, controlling for school characteristics and teacher experience, among elementary teachers. While the psychometric properties results provide important replication indications, the longitudinal analysis provides a much more compelling construct validity assessment within a broader nomological net of teacher sense of efficacy, teacher performance, student achievement, and teacher and school characteristics.

Before describing the methodology of our study, it is necessary to describe our conceptualization of teacher performance, because of its relevance to the task and context specificity of TSE. In our study, we conceptualized and measured teacher performance via a teacher performance competency model known as the Framework for Teaching (Danielson, 1996). The Framework for Teaching is based on a review and synthesis of the empirical and theoretical research on what teachers should know and be able to do in the classroom. Based on a constructivist approach to teaching, the Framework for Teaching assumes that instructional decisions by teachers are purposeful and that teaching activities and assignments are chosen because they serve instructional goals. The Framework for Teaching includes standards that focus on behavioral responsibilities and competencies rather than specific content or subject matter knowledge, though the conveyance of that knowledge is an important activity embedded within the Framework for Teaching.

Four domains of teaching constitute the overarching structure of the Framework for Teaching. These domains are (a) planning and preparation, (b) the classroom environment, (c) instruction, and (d) professional responsibilities. Twenty-two more specific components are

contained within the four domains. Each component, in turn, is broken down into more specific behavioral elements. Each element has four levels of performance: unsatisfactory, basic, proficient, and distinguished. Behavioral rubrics provide standards of performance for each of the four performance levels. To illustrate one of the rubrics, the *knowledge of content* element—one of the elements included in the component *demonstrating knowledge of content and pedagogy* within the *planning and preparation* domain—is presented as follows:

- Unsatisfactory: Teacher makes content errors or does not correct content errors students make.

- Basic: Teacher displays basic content knowledge but cannot articulate connections with other parts of the discipline or with other disciplines.

- Proficient: Teacher displays solid content knowledge and makes connections between content and other parts of the discipline and other disciplines.

- Distinguished: Teacher displays extensive content knowledge, with evidence of continuing pursuit of such knowledge (Danielson, 1996, p. 64).

It is important to note that the Framework for Teaching seeks to provide a comprehensive assessment of teaching practice, yet be general enough to apply to all subject areas and grade levels. It is a generalized depiction of teaching activities that must be assessed by considering the individual context of the teacher and classroom. The conception of teacher self-efficacy underlying the TSES and the conception of teacher performance underlying the Framework for Teaching are very similar. The similarity is particularly noticeable in the cases of instruction and classroom management. This led us to hypothesize that scores on the TSES would be significantly related to scores on a teacher performance measure that is based on the Framework for Teaching.

Numerous school districts throughout the country have developed and implemented teacher performance evaluation systems based on the Framework for Teaching. We studied those systems in multiple locations, including Cincinnati, Ohio, and Washoe County (Reno-Sparks), Nevada. These districts used the Framework for Teaching as the underlying competency model for their evaluation systems and then developed modifications to it to fit their local context. The result was an appraisal instrument in which teachers were rated on components and elements, using behavioral rubrics and multiple, authentic sources of teaching evidence (e.g., classroom observations, examples of lesson planning and student work). (The teacher evaluation system in Washoe County, which is the district of focus for this study, is described in the next section.)

In our research on these evaluation systems, we have examined teacher acceptance of the evaluation system and the validity of the evaluation scores as predictors of subsequent student achievement. We have found that teachers almost uniformly accepted the teacher evaluation instrument and the underlying Framework for Teaching (as modified in the district) as a clear and accurate depiction of the teaching role (Kimball, 2001; Milanowski & Heneman, 2001; Heneman & Milanowski, 2003). Such a result is important for the present study because it suggests our measure of teacher performance is a behaviorally sound one for use as a criterion against which TSES may be correlated. We have also found some preliminary indications that

the evaluation scores are significant predictors of student achievement, though those results vary according to grade level, content area (math or reading), school level, and teacher sample (Kimball, White, Milanowski, & Borman, 2004; Milanowski, 2004).

In short, sound theoretical and empirical reasons support linkages between TSE and teacher performance, and between teacher performance and student achievement. But what about the link between TSE and student achievement? Some research has reported empirical links between these two variables (see Ashton & Webb, 1986; Tschannen-Moran et al., 1998; Pajares, 1996; Muijis & Reynolds, 2002). That research, however, has not explicitly taken teacher performance into account and tested whether the presumed effect of teacher sense of efficacy on student achievement is direct or instead mediated by teacher performance. Given our previous discussion about the task and context specificity of TSE, in our study we hypothesized that the TSES would have a direct effect on teacher performance but not on student achievement. Instead, we anticipated that teacher performance would mediate the relationship between the TSES and student achievement. This expectation was buttressed by Pajares' (1996) research review, in which he stated that,

> When efficacy beliefs are globally assessed and/or do not correspond with their critical tasks with which they are compared, their predictive value is diminished or can even be nullified; and when efficacy assessments are tailored to the critical task, prediction can be enhanced (p. 557).

Based on the above theoretical and empirical considerations, we felt that a longitudinal analysis of the TSES as a predictor of teacher performance and student achievement was both called for and possible. We used nonrecursive structural equation modeling to conduct the analysis. Included in the model were TSE, teacher performance, student achievement, and several teacher and school characteristics (teacher experience, minority concentration, poverty concentration, and school performance level). We expected to find significant direct effects for TSE and teacher performance, and for teacher performance and student achievement. We also expected to find no significant direct effect between TSE and student achievement, since the conceptualization of TSE (and its measurement via the TSES) is performance-based, rather than student-achievement based.

Methodology

Research Setting

The Washoe County School District is the second largest in the state of Nevada and includes schools in urban, suburban, and rural areas. The communities of Reno and Sparks, the primary areas served by the district, have had a large and sustained population growth over the past decade. In addition, the communities have a sizable minority population, primarily represented by Latino families. The district has 88 schools with over 60,000 students and about 3,300 licensed staff. About 400 teachers are hired annually to serve the growing student population.

In 1997, the district started developing a new teacher evaluation system based on Danielson's (1996) Framework for Teaching. A design team that was made up of diverse district stakeholders completed work on the system in 1998 and then conducted a 2-year field test with

60 administrators. Full implementation occurred in 2000. The implemented system has four domains of practice: (a) planning and preparation, (b) classroom environment, (c) instruction, and (d) professional responsibilities. Each domain is represented by five or six components, and each component includes from two to five elements that illustrate performance relative to the component and domain. There are a total of 68 elements, and each element is assessed using a 4-point rubric that contains descriptions of teacher and student behaviors. The rubrics define the four performance levels as (0) unsatisfactory, (1) target for growth, (2) proficient, and (3) area of strength.

Teachers are evaluated by their building principal or assistant principal. The sources of evidence that evaluators may use include classroom and out-of-classroom observations (e.g, teacher collaborative activities, parent meetings), teacher self-assessments, lesson and unit plans, instructional materials (e.g., classroom assignments and student work with teacher comments), reflection sheets, and logs of professional development activities and parental contacts.

All teachers undergo an annual evaluation pursuant to state law. Under the district's system, teachers are evaluated annually on different domains and elements depending on their stage in the district evaluation cycle. The system has three evaluation stages: probationary, post-probationary major, and post-probationary minor. Probationary (non-tenured) teachers receive a comprehensive evaluation on all four performance domains. Evaluators are required to observe these teachers at least nine times over three time points during the year.

Post-probationary (tenured) teachers undergo a 3-year major-minor evaluation cycle. During the first year, they have a major evaluation on two performance domains, one of which they may chose and the other of which is selected by their evaluator. They are formally observed by the evaluator three times during the school year and receive one written evaluation at the end of the year. Over the next 2 years, the minor evaluations focus on one domain each year and involve at least one formal observation. Over this 3-year cycle, post-probationary teachers are assessed on each of the four performance domains. Because district school board members wanted teachers to be evaluated annually on their instruction, teachers who are not formally evaluated on the instruction domain are required to be evaluated on a subset of key instruction-related standards. We refer to this set of standards as the instruction composite, which is described below under teacher measures.

Sample

Respondents on TSES. A Web-based survey was administered in the late fall of 2002 to all licensed staff in the district. The short form of the TSES was embedded within the survey. As an incentive to respond, a small contribution was made on behalf of each responding teacher for literacy activities carried out by the Washoe County Education Foundation, which is a nonprofit entity that garners donations for the benefit of district students. Of 3,228 licensed staff invited to complete the survey, 1,075 classroom teachers provided complete results on the TSES, for a response rate of 33%. The respondents were 76% female and 12% probationary (non-tenured), with 55% holding a master's degree or above. The median amount of experience for the respondents was 11 years. There were no substantial differences between the sample of respondents and the population of licensed staff on these four characteristics.

Respondents for predicting performance and student achievement. Of those who completed the TSES, 180 elementary teachers had student achievement and performance evaluation data that could be compared to their responses on the TSES. These teachers constituted the sample used in the structural equation model analysis. Teachers at the high school and middle school level who had responses on the TSES and performance evaluation ratings could not be included because student achievement data was not available at those school levels.

Measures

Teacher sense of efficacy. The short form of TSES was contained in the teacher survey administered in the fall of 2002. As described above, the TSES short form consists of 12 items measuring three components: efficacy for instructional strategies, efficacy for classroom management, and efficacy for student engagement. For each item, the respondent rates (on a scale from 1 to 9) the extent to which he or she can demonstrate a particular capability relating to teaching. From the TSES items, three scale scores were produced, and a composite score representing the combined scores was also obtained. The items in the short form of the TSES are shown in Appendix A.

Teacher performance. To maximize the number of teachers who could be compared on the same teaching evaluation standards, we used scores from the district's instruction composite as the primary measure of teacher performance. The evaluation scores were obtained from the spring of 2003 (the end of the school year in which the TSES scores were obtained). The composite scores represent key elements from the planning and preparation and instruction domains. As with the individual element scores, teachers are rated on each of four composite scores as (0) unsatisfactory, (1) target for growth, (2) proficient, and (3) area of strength. The composite measure is made up of the following standards:

- The teaching displays solid content knowledge and uses a repertoire of current pedagogical practices for the discipline being taught. (Includes 10 items from two performance domains.)

- The teaching is designed coherently, using a logical sequence, matching materials and resources appropriately, and using a well-defined structure for connecting the individual activities to the entire unit. Instruction links student assessment data to instructional planning and implementation. (Includes nine items from two performance domains.)

- The teaching provides for adjustments in planned lessons to match the students' needs more specifically. The teacher is persistent in using alternative approaches and strategies for students who are not initially successful. (Includes three elements from one domain.)

- The teaching engages students cognitively in activities and assignments, groups are productive, and strategies are congruent to instructional objectives. (Includes three elements from one domain.)

We used the simple average of the four composite scores to obtain an overall measure of teacher performance. The average correlation among the composite scores was .72, and the coefficient alpha was .91 for the overall composite score.

Teacher experience. Experience was derived from a teacher's position on the district's salary schedule, referred to as the *step*. Unlike the date of hire, which was also available, this measure accounted for some or all of the experience of those who had teaching experience outside of the district. Teachers with more experience are given credit for their experience by being placed on a higher step in the salary schedule. Unfortunately, because teachers "top out" on the salary schedule at Step 20, this measure no longer reflects the relative experience of those teachers above the highest step.

Student measures. Student demographic data made available by the district were used to construct dummy variables for special education status, minority status, gender, and eligibility for free and reduced-price meals. Student achievement results in mathematics and reading were drawn from district and state criterion-referenced tests (CRTs) and the TerraNova norm-referenced test. The assessments are part of the Nevada testing regime that was designed to measure proficiency on the Nevada State Content and Performance Standards (Harcourt Assessment, 2004).

The Grades 3 and 5 state CRTs were developed by Harcourt Assessment, Incorporated, in collaboration with representatives from the Nevada Department of Education, state educators, and personnel from the WestEd Regional Education Laboratory (Harcourt Assessment, 2004). Test items and results from this assessment were reviewed by the Nevada Department of Education and Harcourt to demonstrate reliability, validity, and lack of bias. The Grades 4 and 6 district CRTs were also developed by Harcourt Assessment, in collaboration with district content experts. Finally, the norm-referenced Comprehensive Test of Basic Skills 5th Edition (CTBS/5) TerraNova exam, which was administered in the fourth grade, was also used in the analyses. The grade-specific analyses of pretest-to-posttest outcomes for Grades 4, 5, and 6 relied on the following assessments:

Grade 4	
Pretest:	Grade 3 state CRT, spring 2002
Posttest:	Grade 4 state CRT, spring 2003
Grade 5	
Pretest:	Grade 4 CTBS TerraNova, spring 2002
Posttest:	Grade 5 state CRT, spring 2003
Grade 6	
Pretest:	Grade 5 state CRT, spring 2002
Posttest:	Grade 6 district CRT, spring 2003

Using these student variables, we constructed a measure of value-added student achievement based on two-level hierarchical linear modeling, where student pretest and other student characteristics were used to predict achievement on posttests. From these models, empirical Bayes intercept residuals were derived, then standardized by subject and grade, and combined to provide a measure of student achievement. In addition to the individual student results, the district provided data on the following school characteristics: percentage of third graders scoring at the proficient level in reading on the state achievement test, percentage of non-White students, and percentage of students eligible for the free and reduced-price lunch program.

Analyses and Results

Psychometric Properties of the TSES

Our first concern was with the factor structure of the TSES and the extent to which it replicated the structure found by Tschannen-Moran and Hoy (2001). We conducted several confirmatory factor analyses with the TSES. First, three analyses were done with the entire $N = 1{,}075$ data set in order to compare the adequacy of three models: (a) treating the TSES as measuring three separate but correlated factors, (b) treating the TSES as measuring a single factor, and (c) treating the three subscales as distinct latent constructs reflecting a higher order factor. All models were estimated with item-specific variances assumed to be uncorrelated. All models were estimated using SAS PROC CALIS. Table 1 reports the various fit indices for these three models. Since the three models were not nested, it was not possible to compare fit using a statistical test. However, it would appear that the three-correlated-factor model was a substantially better fit than the other two models. The single-factor model was the worst fit, suggesting that the scales successfully represent the intention to reflect three separate subscales. However, since the three-correlated-factor model fit better than the hierarchical model, the subscales may be more appropriately considered as tapping three separate but related constructs. This is consistent with the conception of self-efficacy as relative to a task and context, in this case the teaching role.

Additional analyses suggested that while the three-factor model was a good fit for these data, fit could be improved slightly by allowing some items to load on factors other than those they were intended to measure, or by allowing some correlated uniqueness of items within subscales. Statistically significant improvement in fit might also be achieved by allowing correlated uniquenesses within subscales for two items in the management subscale and three items in the engagement subscale. There was also no obvious reason, such as wording commonalities, to allow these uniquenesses to correlate. Because it was not obvious from the modification indices which changes in the model items would most parsimoniously lead to better fit, we refrained from modifying and retesting the model.

The item factor loadings from the three-correlated-factor model are shown in Table 2. The data are shown for the total sample, as well as separately for elementary and for middle and high school teachers. It can be seen that all of the loadings are above .50, though two items have relatively smaller loadings than the others (*use of a variety of assessment strategies* and *assist families in helping their children do well in school*). The similarities between the loadings for elementary and middle and high school teachers are noteworthy and support the generalizability of the TSES to all three levels of teachers.

Since Tschannen-Moran and Hoy (2001) used exploratory factor analysis, rather than confirmatory factor analysis (which we used), to analyze the factor structure of the TSES, we also conducted exploratory factor analysis to facilitate a more direct comparison with their study. We used a principal components extraction with a varimax rotation. The results (not shown) were very similar to our own confirmatory factor analysis and the exploratory results reported by Tschannen-Moran and Hoy. Specifically, three factors were extracted, and after rotation each item had its highest loading on the same factor as was found in the exploratory analysis and by Tschannen-Moran and Hoy. Cross loadings were minimal. In short, our factor analysis results

provide strong replication of the Tschannen-Moran and Hoy results and extend those results to all three levels of teachers.

Statistical characteristics of the three TSES scales and a total score composite are shown in Table 3 for (a) the total sample, (b) the elementary teachers, and (c) the middle and high school teachers. It can be seen that the means on the scales are well above the midpoint (5) of the rating scale and the mean for the student engagement scale was the lowest of the three scales in all samples. Also, the TSES mean scores in our middle and high school samples were considerably lower than those in our elementary teacher sample. Tschannen-Moran and Hoy (2001) also reported relatively high means, though they did not break out their results by school level.

The standard deviations were all above 1.0 for the subscales in each sample, indicating some degree of variability in scores. The standard deviations for the total scores were just below 1.0. These data are very similar to those reported by Tschannen-Moran and Hoy (2001). Coefficient alphas for the subscales and the total scores were quite high, ranging from .75 to .90. Tschannen-Moran and Hoy reported similar coefficient alphas. Finally, there were moderate intercorrelations among the subscales, ranging from .44 to .66, which are somewhat lower than those reported by Tschannen-Moran and Hoy.

Longitudinal Analysis

In order to assess the hypothesized relationship among teacher efficacy, teacher performance, and student achievement, we estimated a structural equation model. In this analysis, we also controlled for school-level factors likely to be related to both student achievement and efficacy, including percentage of students receiving free or reduced-price lunch, percentage non-White, and percentage of students scoring at the proficient level on the third-grade reading test. We also included the effect of teacher experience on both efficacy and performance. Due to limitations in the district's testing program, student achievement data were not available for middle and high school teachers. Therefore, our analysis of the relationships among TSES scores, rated teacher performance, and student achievement was restricted to the elementary teacher sample. Further, student achievement data were available only for reading and mathematics, which reduced the number of teachers that could be included in the analysis to 180. Means, standard deviations, and intercorrelations for the variables in the model are shown in Table 4.

It can be seen that the subscales of the TSES have correlations with the four teacher performance component scores that range from .22 to .33, all of which are statistically significant at the .05 level. The correlations between the teacher performance components and the student achievement indicators (reading and math achievement, and the average of these) are small (.02–.11) and not significant. The correlations between the TSES subscales and the student achievement indicators are also small (-.06 to .11) and not significant.

Figure 1 shows the structural equation model we estimated. Even though we treated the three TSES subscales as indicators of a single factor rather than three separate, correlated factors, the fit of this model was considered good. The goodness-of-fit index was .944, and the chi-square value was 70.12 with 52 degrees of freedom, yielding a p-value of .045. The root mean

square error of approximation (RMSEA) was .04; the comparative fit index, .99; and the non-normed index, .98. As expected, school percentage receiving free and reduced-price lunch had a significant negative effect on teachers' sense of efficacy, and teacher experience had a significant positive effect. School percentage proficient in Grade 3 reading had a positive, but non-significant effect. Unexpectedly, the school percentage non-White had a positive, significant effect on efficacy, which is likely due to the very high correlation this indicator had with percentage receiving free and reduced-price lunch. Together, these variables explained 13% of the variance in teachers' sense of efficacy. As expected, teachers' sense of efficacy was positively related to teacher performance, controlling for teacher experience. Efficacy and experience accounted for 20% of the variance in teacher performance. Teacher performance was not significantly related to student achievement, nor was there a significant direct effect of teachers' sense of efficacy on student achievement. In this model, 22% of the variance in student achievement was explained. The strongest (and only significant) influence on student achievement in the model was the school percentage receiving free and reduced-price lunch.

Discussion

The findings demonstrate similarities to the results obtained by Tschannen-Moran and Hoy (2001) and also extend their findings in important ways. The results for the psychometric properties of the short form of the TSES are favorable and consistent with those reported by Tschannen-Moran and Hoy. Both confirmatory and exploratory factor analyses revealed the same three distinct (though moderately correlated) factors. Moreover, this factor structure held for teachers at each school level, which was not reported by Tschannen-Moran and Hoy. Other psychometric properties of the three scales were similar to those reported by Tschannen-Moran and Hoy, including scale reliabilities, intercorrelations, means, and standard deviations.

The results for the longitudinal analysis also provide some useful insight into the TSES. As we hypothesized, TSES scores had a significant direct effect on teachers' performance at the end of the school year, controlling for school characteristics and teacher experience. This is a powerful finding in two regards.

First, it lends credence to the conception and measurement of TSE as task- and context-specific, since scores on the TSES were significantly related to an independent measure of teacher performance in which teachers were rated on task performance within the context of identified domains of performance in the teacher role. To our knowledge, this is the first study to empirically investigate and demonstrate such a linkage.

Second, the longitudinal nature of the design ensures that rated teacher performance did not influence teachers' sense of efficacy, since the latter was measured prior to the former. This supports the interpretation that TSE levels near the beginning of the school year influenced teachers' choices and courses of action throughout the school year, resulting in differing levels of performance among the teachers at the end of the school year.

It is also important to note that the TSES did not have a significant direct effect on student achievement. This was in accordance with our expectations since we reasoned that any effect of TSE on student achievement would be mediated by its impact on teacher performance. The non-significant TSES-student achievement relationship requires further elaboration. From a

theoretical perspective, Tschannen-Moran and Hoy (2001) developed the TSES as task- and context-specific within the domain of actual teacher behavior, and not as a predictor of student achievement. The TSES was clearly intended to measure TSE within the teaching role, rather than TSE for influencing student achievement or reaching specific achievement goals. Our results support this conception. It is interesting to speculate on how, and why, other researchers have come to claim that there is (or should be) a link between TSE and student achievement. We have found accessing those individual studies difficult and believe a full-scale meta-analysis of the TSE–student achievement relationship is warranted. As part of that meta-analysis, it would be important to determine if those studies that have found a significant TSE–student achievement link controlled for teacher performance, in keeping with our notion that performance mediates the linkage.

We also note a methodological problem in our study that clouds our findings and interpretations, namely restriction of range in the TSES and teacher performance scores. While we and Tschannen-Moran and Hoy (2001) both found standard deviations of about 1.0 on our TSES scales, investigation of the score distribution reveals a troublesome degree of range restriction. We found that there were very small percentages of responses below the midpoints of the scales, and that with one exception 70% or more of the responses were an average of 6.5 or higher. In a similar vein, our performance rating data reveal that almost 80% of the average of the four composite scores are between 2.25 and 3.00 on the 0–3 rating scale. Such range restriction requires caution in concluding that the non-significant relationships we obtained would also be found in another sample.

Several implications for future research emerge from this study. Our results, coupled with those of Tschannen-Moran and Hoy (2001), suggest that the TSES should be the preferred measure of teachers' sense of efficacy in future research. Its replicable psychometric properties, behavioral richness in capturing the teacher role, and predictive capacity for explaining significant variance in teacher classroom performance all support this conclusion.

Research is needed on the skewed distribution of TSES scores. The tendency for a strong upward bias in teachers' reported self-efficacy may simply be due to the nature of the measuring instrument, or it may reflect a social desirability response bias on the part of teachers. Alternatively, the skew may reflect a true, strong TSE brought about by a combination of self-selection into teaching by those with a high TSE, attrition out of teaching by those with a low TSE, and favorable experiences while teaching that enhance TSE. The cross-sectional nature of our study does not allow us to address these alternative explanations. Exploration of these alternative explanations for high TSE will help us better understand the relative importance of TSE as being brought to the job, enhanced on the job, or lost on the job.

Since TSE is task- and context-specific, future research might profitably begin to explore the task and contextual factors that influence TSE. Bandura's (1997) suggestions that creative mastery experiences and experiences modeling the behavior of others enhance self-efficacy are certainly relevant here. Teacher induction, mentoring, and professional development programs could be designed to include self-efficacy-enhancing components, and their effects on subsequent teacher performance could be assessed. Leadership development programs could be constructed around methods of enhancing TSE (such as through coaching, feedback, and planned competency improvement), and evidence on impacts on teacher performance could be collected.

In conclusion, results of the present study are supportive of the conception and measurement of TSE as operationalized with the TSES. We found the psychometric properties of the TSES similar to those found by Tschannen-Moran and Hoy (2001). We extended their results in three ways. First, we found that the psychometric properties of the TSES hold for all levels of teachers (elementary, middle, and high school). Second, we found that scores on the TSES at the beginning of the school year for teachers were significantly predictive of teacher performance ratings at the end of the school year, with the performance measure based on a teacher performance competency model. Third, we found that the TSES did not have a significant direct effect on student achievement. The latter two findings are buttressed by the longitudinal design of the study and the statistical control for school characteristics and teacher experience. Restriction of range in the TSES and performance ratings prevents a definitive interpretation of their relationships to student achievement. Future research should investigate the distribution of TSE scores, along with task and context factors that potentially influence TSE.

References

Ashton, P. T., & Webb, R. B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement.* White Plains, NY: Longman.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman.

Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.

Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology, 76,* 569–582.

Harcourt Assessment. (2004, March). *Nevada technical report for the CRT and HSPE assessments, 2002–2003 cycle*. San Antonio, TX: Author.

Heneman, H. G., III, & Milanowski, A. (2003). Continuing assessment of teacher reactions to a standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education, 17*(2), 173–195.

Kimball, S. (2001). *Innovations in teacher evaluation: Case studies of two school districts with teacher evaluation systems based on the framework for teaching*. Ann Arbor, MI: UMI Dissertation Publishing Clearinghouse.

Kimball, S. M., White, B., Milanowski, A. T., & Borman, G. (2004). Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education, 79*(4), 54–78.

Milanowski, A. (2004). The criterion-related validity of the performance assessment system in Cincinnati. *Peabody Journal of Education, 79*(4), 33–53.

Milanowski, A., & Heneman, H. G., III. (2001). Assessment of teacher reactions to a standards-based evaluation system: A pilot study. *Journal of Personnel Evaluation in Education, 15*(3), 193–212.

Muijis, D., & Reynolds, D. (2002). Teacher beliefs and behaviors: What really matters? *Journal of Classroom Interaction, 37*(2), 3–15.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*, 533–587.

Tschannen-Moran, M., Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*(2), 202–248.

Tschannen-Moran, M., & Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*, 783–805.

Table 1

Confirmatory Factor Models for Efficacy Scales

Model	χ^2	df	GFI[a]	RMSEA[b]	CFI[c]	TLI/NNFI[d]
Three correlated factors	337.85	51	.95	.07	.96	.95
Hierarchical model	991.37	51	.88	.13	.86	.82
Single-factor model	1946.93	54	.71	.18	.73	.67

Note. All analyses were conducted on an *N* of 1,075.

[a]Goodness-of-fit index. [b]Root mean square error of approximation. [c]Comparative fit index. [d]Tucker-Lewis index, also referred to as the non-normed fit index.

Table 2
Factor Loadings From Three-Correlated-Factor Model

Item	All respondents	Elementary	Middle and high school
Use a variety of assessment strategies	.51	.52	.50
Provide alternative explanation or example when students are confused	.73	.77	.70
Craft good questions for your students	.76	.78	.74
Implement alternative strategies in your classroom	.74	.75	.72
Control disruptive behavior in the classroom	.74	.74	.73
Get children to follow classroom rules	.85	.85	.86
Calm a student who is disruptive or noisy	.89	.86	.90
Establish a classroom management system with each group of students	.85	.84	.85
Get students to believe they can do well in schoolwork	.82	.78	.83
Help your students value learning	.86	.81	.86
Motivate students who show low interest in schoolwork	.80	.77	.79
Assist families in helping their children do well in school	.57	.52	.54
N	1075	567	508
χ^2/df		195.95/51	188.57/51
RMSEA[a]		.07	.07
GFI[b]		.94	.94
CFI[c]		.96	.96
TLI/NNFI[d]		.95	.94

[a]Root mean square error of approximation. [b]Goodness-of-fit index. [c]Comparative fit index. [d]Tucker-Lewis index, also referred to as the non-normed fit index.

Table 3
Means, Standard Deviations, Correlations, and Coefficient Alphas for TSES

a. Full Sample (*n* = 1,075)

Scale	Mean	Standard deviation	1	2	3	4
1. Instruction	7.25	1.02	.77			
2. Management	7.58	1.12	.54	.90		
3. Engagement	6.60	1.25	.48	.55	.84	
4. Composite	7.14	0.93	.79	.84	.84	.89

b. Elementary Teachers (*n* = 567)

Scale	Mean	Standard deviation	1	2	3	4
1. Instruction	7.34	1.01	.78			
2. Management	7.73	1.04	.60	.89		
3. Engagement	7.03	1.05	.52	.59	.80	
4. Composite	7.37	0.87	.83	.86	.84	.90

c. Middle and High School Teachers (*n* = 508)

Scale	Mean	Standard deviation	1	2	3	4
1. Instruction	7.14	1.03	.75			
2. Management	7.42	1.18	.47	.90		
3. Engagement	6.12	1.28	.44	.49	.83	
4. Composite	6.89	0.94	.77	.82	.83	.88

Table 4
Means, Standard Deviations, and Intercorrelations for Variables in Structural Equation Model

N=180	Mean	S.D.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Instruction (TSES)	7.35	.98	-													
2. Management (TSES)	7.92	.93	.55	-												
3. Engagement (TSES)	6.95	.95	.50	.54	-											
4. Avg. composite (TSES)	7.40	.79	.83	.84	.81	-										
5. Perf. component 1	2.64	.45	.28	.26	.24	.31	-									
6. Perf. component 2	2.61	.45	.25	.33	.20	.31	.66	-								
7. Perf. component 3	2.63	.44	.28	.29	.23	.32	.87	.70	-							
8. Perf. component 4	2.62	.45	.29	.27	.25	.33	.94	.65	.80	-						
9. Perf. component avg	2.63	.41	.30	.32	.25	.35	.95	.83	.93	.93	-					
10. Reading achievement	-.08	1.01	.02	.11	.08	.08	.02	.10	.02	.06	.05	-				
11. Math achievement	-.03	.95	-.06	.08	.02	.01	.07	.10	.04	.08	.08	.53	-			
12. Avg. achievement	-.06	.86	-.02	.11	.06	.06	.05	.11	.03	.08	.08	.88	.87	-		
13. % proficient reading	.27	.07	.04	.00	.15	.08	.09	.02	.03	.07	.06	.19	.16	.20	-	
14. % free/reduced lunch	.40	.27	-.12	-.08	-.15	-.14	-.26	-.21	-.23	-.27	-.27	-.31	-.35	-.38	-.59	-
15. % non-White	.35	.24	-.06	.00	-.08	-.06	-.22	-.18	-.18	-.23	-.22	-.28	-.33	-.35	-.62	.92

Note. Correlations with an absolute value of .15 or greater are significant at the .05 level.

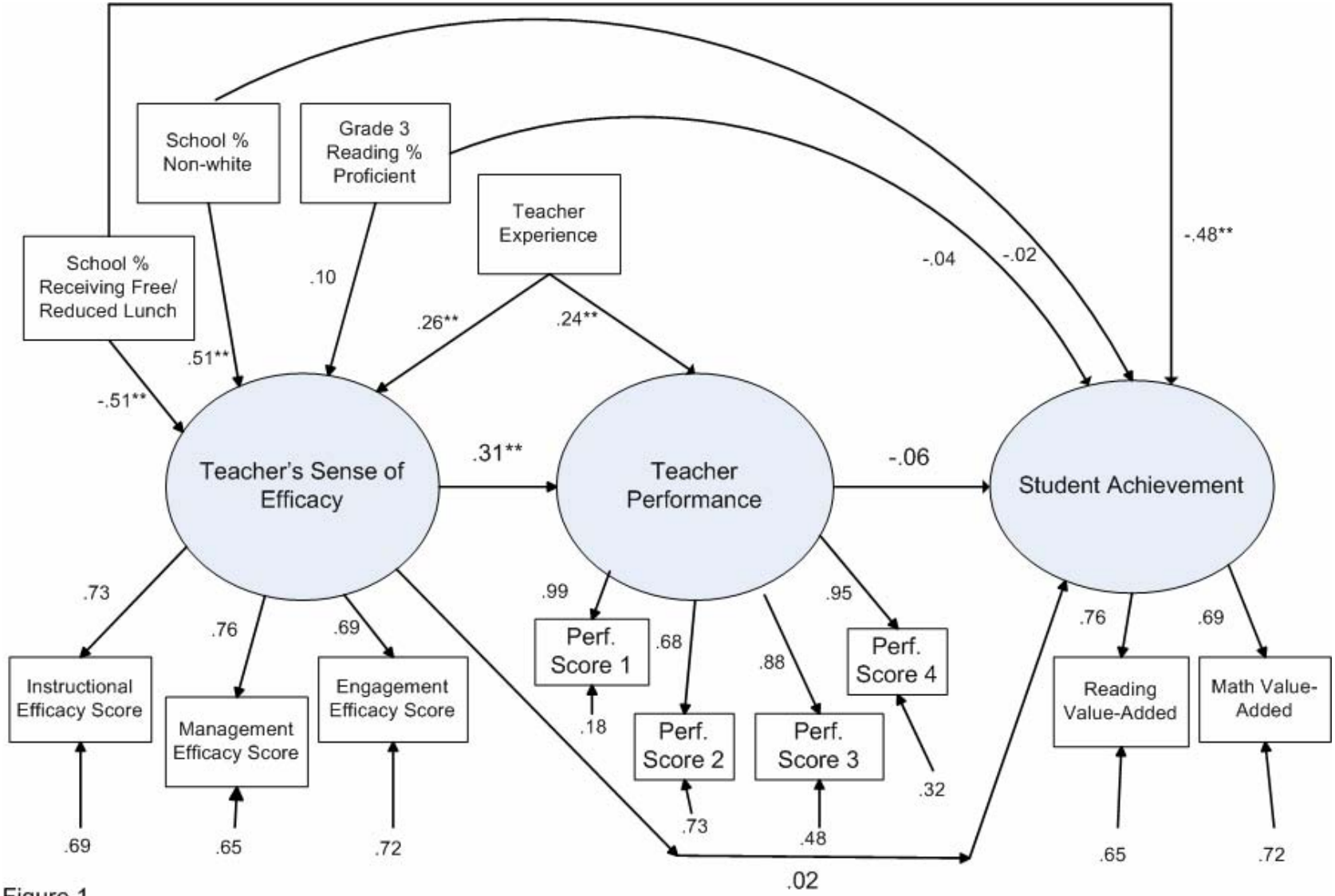

Figure 1
Structural Equation Model of Relationships Between Contextual Influences, Teacher Efficacy, Performance, and Student Achievement
Note. N=180 Elementary Teachers. ** p<.05.

Appendix A

Teacher Sense of Efficacy Short Form

Efficacy for Instructional Strategies

- To what extent can you use a variety of assessment strategies?
- To what extent can you provide an alternative explanation or example when students are confused?
- To what extent can you craft good questions for your students?
- How well can you implement alternative strategies in your classroom?

Efficacy for Classroom Management

- How much can you do to control disruptive behavior in the classroom?
- How much can you do to get children to follow classroom rules?
- How much can you do to calm a student who is disruptive or noisy?
- How well can you establish a classroom management system with each group of students?

Efficacy for Student Engagement

- How much can you do to get students to believe they can do well in schoolwork?
- How much can you help your students value learning?
- How much can you motivate students who show low interest in schoolwork?
- How much can you assist families in helping their children do well in school?

Note. Scales from Tschannen-Moran and Hoy (2001).