PERGAMON

Teaching and Teacher Education 17 (2001) 783–805

TEACHING AND TEACHER EDUCATION

www.elsevier.com/locate/tate

Teacher efficacy: capturing an elusive construct

Megan Tschannen-Moran[a,*], Anita Woolfolk Hoy[b]

[a] *School of Education, College of William and Mary, P.O. Box 8795, Williamsburg, VA 23187 8795, USA*
[b] *College of Education, The Ohio State University, Columbus, OH 43210, USA*

Received 20 June 2000; received in revised form 18 December 2000; accepted 13 February 2001

Abstract

Teacher efficacy has proved to be powerfully related to many meaningful educational outcomes such as teachers' persistence, enthusiasm, commitment and instructional behavior, as well as student outcomes such as achievement, motivation, and self-efficacy beliefs. However, persistent measurement problems have plagued those who have sought to study teacher efficacy. We review many of the major measures that have been used to capture the construct, noting problems that have arisen with each. We then propose a promising new measure of teacher efficacy along with validity and reliability data from three separate studies. Finally, new directions for research made possible by this instrument are explored. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Teacher efficacy; Teacher beliefs; Self-efficacy; Measurement

Teacher efficacy is a simple idea with significant implications. A teacher's efficacy belief is a judgment of his or her capabilities to bring about desired outcomes of student engagement and learning, even among those students who may be difficult or unmotivated (Armor et al., 1976; Bandura, 1977). This judgment has powerful effects.

Teachers' sense of efficacy has been related to student outcomes such as achievement (Armor et al., 1976; Ashton & Webb, 1986; Moore & Esselman, 1992; Ross, 1992), motivation (Midgley, Feldlaufer, & Eccles, 1989), and students' own sense of efficacy (Anderson, Greene, & Loewen, 1988). In addition, teachers' efficacy beliefs also relate to their behavior in the classroom. Efficacy affects the effort they invest in teaching, the goals they set, and their level of aspiration. Teachers with a strong sense of efficacy tend to exhibit greater levels of planning and organization (Allinder, 1994). They also are more open to new ideas and are more willing to experiment with new methods to better meet the needs of their students (Berman, McLaughlin, Bass, Pauly, & Zellman, 1977; Guskey, 1988; Stein & Wang, 1988). Efficacy beliefs influence teachers' persistence when things do not go smoothly and their resilience in the face of setbacks. Greater efficacy enables teachers to be less critical of students when they make errors (Ashton & Webb, 1986), to work longer with a student

*Corresponding author. Tel.: +1-757-221-2187; fax: +1-757-221-2988.
E-mail address: mxtsch@wm.edu (M. Tschannen-Moran).

PII: S0742-051X(01)00036-1

who is struggling (Gibson & Dembo, 1984), and to be less inclined to refer a difficult student to special education (Meijer & Foster, 1988; Podell & Soodak, 1993; Soodak & Podell, 1993). Teachers with a higher sense of efficacy exhibit greater enthusiasm for teaching (Allinder, 1994; Guskey, 1984; Hall, Burley, Villeme, & Brockmeier, 1992), have greater commitment to teaching (Coladarci, 1992; Evans & Tribble, 1986; Trentham, Silvern, & Brogdon, 1985) and are more likely to stay in teaching (Burley, Hall, Villeme, & Brockmeier, 1991; Glickman & Tamashiro, 1982). Clearly the study of this construct has borne much fruit in the field of education. And yet researchers have had difficulty developing a measurement tool to capture it.

There are a variety of problems with existing measures of teacher efficacy. Researchers have questioned the validity and reliability of existing measures. In addition, many measures reveal a two-factor structure when subjected to factor analysis, and there is confusion and debate about the meaning of these two factors. These and other unresolved issues continue to perplex scholars working to improve the measurement of teacher efficacy. For example, there has been disagreement over the conceptualization of teacher efficacy that has contributed to a lack of clarity in measuring the construct. There are questions about the extent to which teacher efficacy is specific to given contexts and to what extent efficacy beliefs are transferable across contexts. In addition, the appropriate level of specificity in the measure of teacher efficacy has been difficult to discern.

The purpose of this paper is to explore issues related to the measurement of teacher efficacy and to propose a new measure. First, we examine various instruments that have been used to assess teacher efficacy as well as the problems that have arisen with each. Next, we introduce a new measure of teacher efficacy based on a model of teacher efficacy suggested by Tschannen-Moran, Woolfolk Hoy, and Hoy (1998), along with reliability and validity data from three studies. Finally, we propose new directions for research in light of the new measure.

1. A first attempt at measurement: Rotter's locus of control

The search for ways to measure teacher efficacy has not suffered from a lack of effort. In the attempt to capture the meaning of this apparently powerful construct, researchers have tried both long, detailed measures and short, general ones. The first measures were grounded in Rotter's social learning theory.

1.1. The Rand measure

The simple idea that teachers' perceptions of their own capabilities are important, began with a simple measure—just two items. These two items were buried in an otherwise extensive questionnaire, and yet they turned out to be among the most powerful factors examined by Rand researchers in their study of teacher characteristics and student learning (Armor et al., 1976). With the work of Rotter (1966) as a theoretical base, the Rand researchers conceived teacher efficacy as the extent to which teachers believed that they could control the reinforcement of their actions, that is, whether control of reinforcement lay within them or in the environment. Teachers who concur that the influence of the environment overwhelms a teacher's ability to have an impact on a student's learning exhibit a belief that reinforcement of their teaching efforts lies outside their control or is *external* to them. Teachers who express confidence in their ability to teach difficult or unmotivated students evidence a belief that reinforcement of teaching activities lies within the teacher's control or is *internal*.

To measure efficacy, teachers were asked to indicate their level of agreement with the two statements below. The sum of the two items was called teacher efficacy (TE), a construct that purported to reveal the extent to which a teacher believed that the consequences of teaching—student motivation and learning—were in the hands of the teacher, that is, internally controlled.

Rand item 1. When it comes right down to it, a teacher really can't do much because most of a student's motivation and performance depends on his or her home environment." A teacher who

expresses strong agreement with this statement indicates that environmental factors overwhelm any power that teachers can exert in schools. This assessment extends beyond the individual capabilities of the particular teacher to teachers in general. Factors such as the value placed on education at home; the conflict, violence, or substance abuse in the home or community; the social and economic realities concerning class, race, and gender; and the physiological, emotional and cognitive needs of a particular child all have a very real impact on a student's motivation and performance in school. Teachers' beliefs about the power of these external factors compared to the influence of teachers and schools have since been labeled general teaching efficacy (GTE) (Ashton, Olejnik, Crocker, & McAuliffe, 1982).

Rand item 2. If I really try hard, I can get through to even the most difficult or unmotivated students." Teachers who agree with this statement indicate confidence in their abilities as teachers to overcome factors that could make learning difficult for a student. The teachers are making a statement about the efficacy of their own teaching, reflecting confidence that they have adequate training or experience to develop strategies for overcoming obstacles to student learning. These teachers may well have experienced past success in boosting students' achievement. This aspect of efficacy has been labeled personal teaching efficacy (PTE); it is more specific and individual than a belief about what teachers in general can accomplish.

This appealing notion, that teachers' beliefs in their own capabilities somehow matter, proved to be significantly related to teachers' success in teaching reading to minority students in an urban context (Armor et al., 1976). In a second study, Rand researchers found teacher efficacy to be a strong predictor of the continuation of federally funded innovations after the end of funding (Berman et al., 1977). Teachers' sense of efficacy had a strong positive link not only to student performance but to the percent of project goals achieved, to the amount of teacher change, and to the continued use of project methods and materials after the project ended.

Spurred on by the success of the Rand studies, several researchers sought to expand and refine the notion of teacher efficacy, developing measures they hoped would capture more of this powerful construct. Researchers were concerned about the reliability of the two-item scale and attempted to develop longer, more comprehensive measures. Three such instruments are reviewed below. Each of these builds on the foundation laid by Rotter, conceptualizing teacher efficacy as teachers' beliefs that factors under their control ultimately have greater impact on the results of teaching than factors in the environment or in the student—factors beyond the influence of teachers.

1.2. Responsibility for student achievement

Shortly after the first Rand study was published, Guskey developed a 30-item instrument measuring *responsibility for student achievement* (Guskey, 1981). For each item, participants were asked to distribute 100 percentage points between two alternatives, one stating that the event was caused by the teacher and the other stating that the event occurred because of factors outside the teacher's immediate control. Consistent with explanations from attribution theory (Weiner, 1979, 1992), four types of causes were offered for success or failure: specific teaching abilities, the effort put into teaching, the task difficulty, and luck. (See Fig. 1 for sample items.) Scores on the *responsibility for student achievement* (*RSA*) yielded a measure of how much the teacher assumed responsibility for student outcomes in general, as well as two subscale scores indicating responsibility for student success (R+) and for student failure (R−). The 100-point scale proved cumbersome and in subsequent uses the scale was reduced to 10 points for the teacher to divide between the alternative explanations.

When Guskey (1982, 1988) compared scores from the RSA with teacher efficacy (TE) as measured by the sum of the two Rand items, he found significant positive correlations between teacher efficacy and responsibility for both student success (R+) and student failure (R−). He reported strong intercorrelations ranging from 0.72 to 0.81 between overall responsibility and responsibility for student success and student failure, while the subscales for student success

Format:	**Sample Items:**
Participants are asked to give a weight or percent to each of the two choices. **Scoring:** A global measure of responsibility, with two subscales: Responsibility for student success (R+) & Responsibility for student failure (R-)	If a student does well in your class, would it probably be a. because that student had the natural ability to do well, or b. because of the encouragement you offered? When your students seem to have difficulty learning something, is it usually a. because you are not willing to really work at it, or b. because you weren't able to make it interesting for them?

Fig. 1. Responsibility for student achievement (Guskey, 1981).

Format:	**Sample Items:**
28 items with a forced-choice format. **Scoring:** Half of the items describe situations of student success (I+) and half describe student failure (I-).	Suppose you are teaching a student a particular concept in arithmetic or math and the student has trouble learning it. Would this happen a. because the student wasn't able to understand it, or b. because you couldn't explain it very well? If the students in your class perform better than they usually do on a test, would this happen a. because the students studied a lot for the test, or b. because you did a good job of teaching the subject area?

Fig. 2. Teacher locus of control (Rose & Medway, 1981).

and student failure were only weakly related (0.20) or not at all (Guskey, 1981, 1988). Guskey asserted that positive and negative performance outcomes represent separate dimensions, not opposite ends of a single continuum, and that these dimensions operate independently in their influence on perceptions of efficacy (Guskey, 1987). In general, teachers assumed greater responsibility for positive results than for negative results, that is, they were more confident in their ability to influence positive outcomes than to prevent negative ones. Greater efficacy was related to a high level of confidence in teaching abilities on a measure of teaching self-concept (Guskey, 1984). In our extensive review of the research on teacher efficacy, no published studies were found in which other researchers had adopted this measure.

1.3. Teacher locus of control

At the same time that Guskey developed the *RSA*, Rose and Medway (1981) proposed a 28-item measure called the *teacher locus of control* (*TLC*) in which teachers were asked to assign responsibility for student successes or failures by choosing between two competing explanations for the situations described. Half the items on the TLC described situations of student success and the other half described student failure. For each success situation, one explanation attributed the positive outcome internally to the teacher (I+) and the other assigned responsibility outside the teacher, usually to the students. Similarly, for each failure situation, one explanation gave an internal teacher attribution (I−) and the other blamed external factors. (See Fig. 2 for sample items.)

Scores on the TLC have been weakly but significantly related to the individual Rand items (GTE and PTE) as well as to the sum of the two Rand items (TE) with correlations generally ranging from 0.11 to 0.41 (Coladarci, 1992; Parkay, Greenwood, Olejnik, & Proller, 1988). Rose and Medway (1981) found that the TLC was a better predictor of teacher behaviors than Rotter's internal–external (I–E) scale, probably because it was more specific to a teaching context.

For example, the TLC predicted teachers' willingness to implement new instructional techniques, whereas Rotter's I–E scale did not. To further examine the TLC and the two Rand items, Greenwood, Olejnik, and Parkay (1990) dichotomized teachers' scores on the two Rand questions and cross-partitioned them into four efficacy patterns. They found that teachers with high efficacy on both measures (I can, teachers can) had more internally oriented scores on the TLC for both student success and student failure than teachers who scored low on both (I can't, teachers can't). This measure never received wide acceptance and has all but disappeared from the literature in the past two decades.

1.4. The Webb scale

At about the same time as the RSA and the TLC were being developed, a third group of researchers sought to expand the Rand efficacy questions to increase their reliability. *The Webb scale* (Ashton et al., 1982) was an attempt to extend the measure of teacher efficacy while maintaining a narrow conceptualization of the construct. To reduce the problem of social desirability bias, a forced-choice format with items matched for social desirability was used. (See Fig. 3 for sample items.) Webb and his colleagues found that teachers who scored higher on the *Webb efficacy scale* evidenced fewer angry or impatient interactions (less negative affect) in their teaching (Ashton et al., 1982). This measure, however, never met with wide acceptance and we found no published work beyond the original study in which the scale was used.

2. A second conceptual strand: Bandura's social cognitive theory

While one strand of research grounded in Rotter's social learning theory developed, a second strand emerged, growing out of Bandura's social cognitive theory and his construct of self-efficacy, as initially described in his 1977 article, "Self-efficacy: Toward a unifying theory of behavioral change". Bandura (1997) defined perceived self-efficacy as "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (p. 3). Self-efficacy is a future-oriented belief about the level of competence a person expects he or she will display in a given situation. Self-efficacy beliefs influence thought patterns and emotions that enable actions in which people expend substantial effort in pursuit of goals, persist in the face of adversity, rebound from temporary setbacks, and exercise some control over events that affect their lives (Bandura, 1986, 1993, 1997).

Social cognitive theory proposes a second kind of expectation, outcome expectancy, which is distinct from efficacy expectations. An efficacy expectation is the individual's conviction that he or she can orchestrate the necessary actions to perform a given task, while outcome expectancy is the individual's estimate of the likely consequences of performing that task at the expected level of competence (Bandura, 1986). Bandura asserted that because they stem from the projected level of competence a person expects to bring to a given situation, outcome expectancies add little to the predictive power of efficacy measures. However, outcome expectancies, in the form of physical or

Format:	**Sample Items:**
7 items, forced choice. Participants must determine if they agree most strongly with the first or the second statement.	A. A teacher should not be expected to reach every child; some students are not going to make academic progress. B. Every child is reachable. It is a teacher's obligation to see to it that every child makes academic progress. A. My skills are best suited for dealing with students who have low motivation and who have a history of misbehavior in school. B. My skills are best suited for dealing with students who are academically motivated and generally well behaved.

Fig. 3. Webb efficacy scale (Ashton et al., 1982).

social rewards, recognitions, punishments, criticisms, or self-evaluations can provide incentives and disincentives for a given behavior (Bandura, 1986, 1997). Several researchers attempted to draw on both Rotter and Bandura, reconciling the two conceptualizations or simply ignoring the distinctions.

2.1. The Ashton vignettes

In order to address the assumption that teacher efficacy is context specific, Ashton and her colleagues (Ashton, Buhr, & Crocker, 1984) developed a series of vignettes describing situations a teacher might encounter and asking teachers to make judgments as to their effectiveness in handling the situation. The researchers tested two frames of reference for judgments. The first asked teachers to judge how they would perform in the described situation on a scale from "extremely ineffective" to "extremely effective". The second version asked teachers to make a comparison to other teachers, from "much less effective than most teachers" to "much more effective than most teachers". The norm-referenced vignettes in which teachers compared themselves to other teachers were significantly correlated with the Rand items but the self-referenced vignettes, rating effectiveness or ineffectiveness, were not (Ashton et al., 1984; Ashton & Webb, 1986). (See Fig. 4 for sample items.) Teachers also were asked to indicate the level of stress in each of the situations but, with correlations between efficacy and stress ranging from −0.05 to −0.82 with an average of −0.39, it was concluded that stress could not be used as a proxy for efficacy. This measure has not received wide acceptance. We located only one study that used this scale since it was developed for the original study.

2.2. Gibson and Dembo's teacher efficacy scale (TES)

The early 1980s was a fertile time for attempts to measure teacher efficacy. It was in those years that Gibson and Dembo developed the TES, building on the formulations of the Rand studies, but bringing to bear the conceptual underpinnings of Bandura as well. Beginning with teacher interviews and analyses of previous studies of teachers reported to have a strong sense of efficacy, Gibson and Dembo (1984) developed a 30-item measure of teacher efficacy. (See Fig. 5 for sample items.) Perplexed when factor analysis of the items yielded a two-factor structure, Gibson and Dembo assumed that the two factors reflected the two expectancies of Bandura's social cognitive theory: self-efficacy and outcome expectancy. Consequently, Gibson and Dembo called the first factor *personal teaching efficacy* (PTE, $\alpha = 0.75$) assuming that it reflected self-efficacy, and the second they called *teaching efficacy* (GTE, $\alpha = 0.79$) assuming that it captured outcome expectancy. They wrote

> If we apply Bandura's theory to the construct of teacher efficacy, outcome expectancy would

Format:	**Sample Items:**
50 items describing problem situations concerning various dimensions of teaching, including motivation, discipline, academic instruction, planning, evaluation, and work with parents. **Self-referenced:** "extremely ineffective" to "extremely effective." **Norm-referenced:** "much less effective than most teachers" to "much more effective than other teachers."	Your school district has adopted a self-paced instructional program for remedial students in your area. How effective would you be in keeping a group of remedial students on task and engaged in meaningful learning while using these materials? A small group of students is constantly whispering, passing notes and ignoring class activities. Their academic performance on tests and homework is adequate and sometimes even good. Their classroom performance, however, is irritating and disruptive. How effective would you be in eliminating their disruptive behavior?

Fig. 4. Ashton vignettes (Ashton et al., 1984).

Format:	Sample Items:
30 items on a 6-point Likert scale from strongly disagree to strongly agree. **Scoring:** A global measure of teacher efficacy derived from the sum of all items. Two subscales emerge from factor analysis: personal teaching efficacy and general teaching efficacy.	When a student gets a better grade than he usually gets, it is usually because I found better ways of teaching. The hours in my class have little influence on students compared to the influence of their home environment. If a student masters a new math concept quickly, this might be because I knew the necessary steps in teaching that concept.

Fig. 5. Teacher efficacy scale (Gibson & Dembo, 1984).

essentially reflect the degree to which teachers believed that environment could be controlled, that is, the extent to which students can be taught given such factors as family background, IQ, and school conditions. Self-efficacy beliefs would be teachers' evaluation of their abilities to bring about positive student change (Gibson & Dembo, 1984, p. 570).

As will be discussed later in this paper, questions have been raised regarding this interpretation of outcome expectancy (Tschannen-Moran et al., 1998; Woolfolk & Hoy, 1990).

Using the Gibson and Dembo items, other researchers have confirmed the existence of two factors (Anderson et al., 1988; Burley et al., 1991; Hoy & Woolfolk, 1993; Moore & Esselman, 1992; Saklofske, Michaluk, & Randhawa, 1988; Soodak & Podell, 1993) with α ranging from 0.75 to 0.81 for PTE and 0.64 to 0.77 for GTE. When the Rand items were included in the factor analysis with the Gibson and Dembo measure, Rand 1 *(when it comes right down to it, a teacher really can't do much because most of a student's motivation and performance depends on his or her home environment)* usually loaded on the GTE factor and Rand 2 *(if I really try hard, I can get through to even the most difficult or unmotivated students)* usually loaded on the PTE factor (Coladarci, 1992; Woolfolk & Hoy, 1990). Studies of both preservice and inservice teachers have found that from 18% to 30% of the variance between teachers is explained by these two factors. In general, researchers have found the two factors to be only moderately related, with correlations ranging from 0.15 to 0.20.

Continued research with the Gibson and Dembo items began to identify inconsistencies. Factor analysis of the 30-item instrument indicated that several items loaded on both factors, consequently some researchers have used a shortened version, selecting only the 16 items that loaded uniquely on one factor or the other (Soodak & Podell, 1993; Woolfolk & Hoy, 1990). Even so, problems have arisen around particular items. Using the 16-item version of the Gibson and Dembo instrument Soodak and Podell (1993) found that, contrary to expectations, one GTE item loaded on the PTE factor and that another item did not have a strong enough loading on either factor to be included. In light of these findings, Hoy and Woolfolk (1993) used an even more abbreviated form with just 10 items: five personal and five general teaching efficacy items. They found reliabilities for both subtests within the range found for the longer versions (α is 0.77 for PTE and 0.72 for GTE). They urged researchers to conduct factor analysis on their own data, because of frequent inconsistencies across studies.

Although the Gibson and Dembo measure has been the most popular of the teacher efficacy instruments to date, problems remain both conceptually and statistically. The lack of clarity about the meaning of the two factors and the instability of the factor structure make this instrument problematic for researchers. A new, clearer measure is needed.

2.3. Subject-matter specific modifications of Gibson and Dembo's instrument

One of the unresolved issues in the measurement of teacher efficacy is determining the optimal level

of specificity. Teacher efficacy has been defined as both context and subject-matter specific. A teacher may feel very competent in one area of study or when working with one kind of student and feel less able in other subjects or with different students. Although researchers and theorists agree that teacher efficacy is situation specific, it is less clear what is the appropriate level of specificity for its measure. For example, is efficacy specific to teaching mathematics, or more specific to teaching algebra, or even more specific to teaching quadratic equations? This situation is not unlike issues faced by researchers studying self-efficacy for school achievement. Pintrich and Schunk (1996) have noted that the level of specificity is one of the most difficult issues to be resolved for cognitive or motivational theories that propose domain specificity (p. 79). In general, attempts to limit the scope of the efficacy beliefs have been fruitful in terms of finding significant results. But whether these measures have greater predictive value and generalizability than more global measures has yet to be determined. Recognizing that many standard efficacy instruments overlook the specific teaching context, some researchers have modified the Gibson and Dembo instrument to explore teachers' sense of efficacy within particular curriculum areas.

Science teaching. Science educators have conducted extensive research on the effects of efficacy on science teaching and learning. Riggs and Enochs (1990) developed an instrument, based on the Gibson and Dembo approach, to measure efficacy of teaching science—the *Science Teaching Efficacy Belief Instrument* (STEBI). (See Fig. 6 for sample items.) Consistent with Gibson and Dembo they have found two separate factors, one they called *personal science teaching efficacy* (PSTE) and a second factor they labeled *science teaching outcome expectancy* (STOE). The two factors were uncorrelated. Exploring an even greater level of specificity, Rubeck and Enochs (1991) distinguished chemistry teaching efficacy from science teaching efficacy. They found that among middle-school science teachers, personal science teaching efficacy (PTE for teaching science) was correlated with preference to teach science, and that chemistry teaching self-efficacy (PTE for teaching chemistry) was related to preference to teach chemistry. Chemistry teaching self-efficacy was related to science teaching self-efficacy, and science teaching self-efficacy was significantly higher than chemistry teaching self-efficacy. Science teaching self-efficacy was related to the teacher's experiences taking science courses with laboratory experiences and to experience teaching science, while chemistry self-efficacy was related to chemistry course work involving lab experiences and chemistry teaching experience. This instrument has enjoyed popularity across several studies (see Enochs, Posnanski, & Hagedorn, 1999).

Classroom management. In an attempt to extend the TES to better reflect the domain of classroom management, Emmer (1990) adapted the Gibson and Dembo instrument, yielding a 36-item measure with three efficacy subscales: efficacy for classroom management and discipline, external influences, and personal teaching efficacy. Among a sample of preservice teachers, the efficacy subscales were correlated with preferences for using positive strategies for classroom management, that is, strategies aimed at increasing or encouraging desirable student responses through praise, encouragement, attention, and rewards (Emmer, 1990; Emmer & Hickman, 1990).

Special education. To explore efficacy in the context of special education, Coladarci and Breton (1997) used a 30-item instrument modified from Gibson and Dembo (1984) and reworded to apply specifically to special education. In order to study

Format:	**Sample Items:**
25 items, 5-point Likert scale from strongly agree to strongly disagree.	I understand science concepts well enough to be effective in teaching elementary science. Effectiveness in science teaching has little influence on the achievement of students with low motivation.

Fig. 6. Science teaching efficacy belief instrument (Riggs & Enochs, 1990).

the likelihood of referral to special education in the Netherlands, Meijer and Foster (1988) developed the *Dutch teacher self-efficacy scales*, an 11-item instrument probing personal teaching efficacy beliefs. Teachers were asked to respond along a 4-point Likert scale to questions such as "I become truly discouraged when I see a pupil returning to problem behavior" or "I can handle virtually any learning problem well". The researchers found that high efficacy teachers were more likely to feel that a problem student was appropriately placed in the regular classroom.

2.4. Brief eclectic measures

Some researchers, dissatisfied with any of the existing measures, have used a combination of items from several instruments. Midgley et al. (1989) created a 5-item personal teaching efficacy measure consisting of the Rand personal efficacy item, two items of academic futility (Brookover et al., 1978), one item from the Webb Scale, and one original item, and then summed across the five items ($\alpha = 0.65$). Several researchers who made use of the High School and Beyond database used a 2-item measure of self-efficacy and two items indicating satisfaction; however, because these measures were so highly correlated they combined them into a single measure (Lee, Dedick, & Smith, 1991; Newmann, Rutter, & Smith, 1989). This seems an unfortunate approach because even though efficacy and satisfaction were correlated, they remain conceptually distinct constructs. Raudenbush, Rowen, and Cheong (1992) decided to use a very brief measure of efficacy. They asked teachers to respond to the single question, "To what extent do you feel successful in providing the kind of education you would like to provide for this class?" with responses along a 4-point Likert scale.

2.5. Bandura's teacher self-efficacy scale

In the midst of the confusion about how to best measure teacher efficacy, an unpublished measure used by Bandura (undated) in his work on teacher efficacy has begun quietly circulating among researchers. Bandura (1997) pointed out that teachers' sense of efficacy is not necessarily uniform across the many different types of tasks teachers are asked to perform, nor across different subject matter. In response, he constructed a 30-item instrument with seven subscales: efficacy to influence decision making, efficacy to influence school resources, instructional efficacy, disciplinary efficacy, efficacy to enlist parental involvement, efficacy to enlist community involvement, and efficacy to create a positive school climate. Each item is measured on a 9-point scale anchored with the notations: "nothing, very little, some influence, quite a bit, a great deal". (See Fig. 7 for sample items.) This measure attempted to provide a multifaceted picture of teachers' efficacy beliefs without becoming too narrow or specific. Unfortunately, reliability and validity information about the measure have not been available.

Format:	**Sample Items:**
30 items on a 9-point scale anchored at nothing, very little, some influence, quite a bit, a great deal.	How much can you influence the decisions that are made in your school?
7 subscales:	How much can you do to overcome the influence of adverse community conditions on student learning?
Influence on decision making, influence on school resources, instructional efficacy, disciplinary efficacy, enlisting parental involvement, enlisting community involvement, and creating a positive school climate.	How much can you do to get children to follow classroom rules?
	How much can you assist parents in helping their children do well in school?
	How much can you do to get local colleges and universities involved in working with your school?
	How much can you do to make students enjoy coming to school?
	How much can you do to get students to believe they can do well in schoolwork?

Fig. 7. Bandura's teacher efficacy scale.

The conceptual confusion around the concept of teacher efficacy has made developing appropriate measures of efficacy difficult. Researchers have tried very simple, general measures as well as long complex vignettes. None of the measures currently in use seems to have found the proper balance between specificity and generality. In addition, there are conceptual problems in the interpretation of the factor structure and the poor correlation between the factors where two or more have been found.

3. Challenges in the measure of teacher efficacy

Studies of teacher efficacy have frequently found two separate dimensions or factors, although considerable confusion and debate have arisen over their meaning. While there is general agreement that the first factor, commonly called personal teaching efficacy, has to do with one's own feelings of competence as a teacher, the meaning of the second factor has been in question. Although it is often called general teaching efficacy, some have argued for other labels. Emmer and Hickman (1990) called the second factor "external influences" which is reminiscent of Rotter's construct of external control. Riggs and Enochs in the development of the Science Teaching Efficacy Belief Instrument (Riggs and Enochs, 1990) have labeled Factor 2 as an outcome expectancy, the second component of Bandura's social cognitive theory in which a person assesses the likely consequences of the performance level he or she expects to achieve. Riggs and Enochs, (along with Ashton et al. (1982), Gibson and Dembo (1984) and Soodak and Podell (1996)), reasoned that what teachers in general could be expected to accomplish was the outcome an individual teacher could expect from his or her own teaching.

Bandura (1986) argued that an outcome expectancy is a judgment of the likely consequences of a specific action, given an individual's anticipated level of performance. Bandura pointed out that outcome expectancy adds little to the explanation of motivation because the outcome a person expects stems from that person's assessment of his or her own capabilities and expected level of performance, not from what it would be possible for others to accomplish under similar circumstances. Therefore the items used to measure the second factor of teacher efficacy about the potential impact of teachers in general in the face of external impediments (GTE) cannot be considered an outcome expectancy (Tschannen-Moran et al., 1998; Woolfolk & Hoy, 1990). To capture the contingency relationship between means and ends (Skinner, 1996), items would have to refer to outcomes the individual teacher could expect, given certain actions or means he or she felt capable of delivering. Emmer and Hickman's label "external influences" strikes closer to the mark of what the current GTE items capture.

3.1. Relationships among existing measures

Coladarci and Fink (1995) undertook an examination of the major measures of teacher efficacy and their relationships to one another. In a sample of elementary and secondary public school teachers, they found a correlation between the Rand measure and the TES (Gibson & Dembo, 1984) of 0.64. The TES correlated with the *teacher locus of control scale* (Rose & Medway, 1981) at 0.47 and with *responsibility for student achievement questionnaire* (Guskey, 1981) at 0.57 (see Table 1). These moderate correlations suggest that these measures are describing related constructs, but the overlap is not perfect. How much of what each scale measures accurately captures teacher efficacy and how much is something else?

A closer examination of the relationships between the subscales adds more intriguing information (see Table 2). Intercorrelations between the general teaching efficacy of the TES and Rand 1 (general) measure of 0.53, and between the personal teaching efficacy and Rand 2 (personal) of 0.41 are not as strong as might have been expected. Previous studies have generally found that when the Rand items were included in the TES, the factor structure remained intact (Coladarci, 1992; Woolfolk & Hoy, 1990). Both the *teacher locus of control* subscale for student success (I+) and the *responsibility for student success* (R+) correlated

Table 1
Intercorrelations among efficacy/non-efficacy measures ($N = 333$)[a]

	TLC	RSA	TES	EV	Webb	AfT	TSC
Rand items	0.47	0.50	0.64	0.27	0.39	0.45	0.48
Teacher locus of control scale		0.68	0.47	0.18	0.28	0.34	0.33
Responsibility for student achievement questionnaire			0.57	0.22	0.41	0.39	0.46
Teacher efficacy scale				0.39	0.42	0.50	0.54
Efficacy vignettes					0.34	0.36	0.38
Webb efficacy scale						0.32	0.40
Affect for teaching							0.72
Teaching self-concept							

[a]Source: Coladarci and Fink (1995); used with permission.

Table 2
Subscale intercorrelations among efficacy/non-efficacy measures ($N = 333$)[a]

	1	2	3	4	5	6	7	8	9	10
General teaching efficacy										
1. Rand 1		0.53	0.45	0.35	0.41	0.39	0.37	0.35	0.35	0.37
2. Teacher efficacy scale			0.42	0.25	0.28	0.40	0.35	0.39	0.39	0.39
Personal teaching efficacy										
3. Rand 2				0.41	0.34	0.36	0.27	0.34	0.42	0.46
4. Teacher efficacy scale					0.47	0.47	0.21	0.28	0.39	0.47
Efficacy for classroom success/positive student outcomes										
5. Teacher locus of control: I+						0.53	0.54	0.49	0.33	0.35
6. Responsibility for student Ach: R+							0.30	0.41	0.43	0.49
Efficacy for classroom failure/negative student achievement										
7. Teacher locus of control: I-								0.65	0.26	0.22
8. Responsibility for student Ach: R-									0.26	0.31
Non-efficacy measures										
9. Affect for teaching										0.72
10. Teaching self-concept										

[a]Source: Coladarci and Fink (1995); used with permission.

most strongly with personal teaching efficacy (0.47), but their relationship to Rand 1 (general or external) was almost as high (0.41 and 0.39, respectively). The external measures, the TLC for student failure (I−) and RSA subscale for student failure (R−), were related to general teaching efficacy (0.35 and 0.39, respectively); however, a stronger relationship was found between these two external measures and the internal scale of the TLC (I+: 0.54 and 0.49, respectively) and between the two subscales of the RSA (0.41) (see Table 1). When the various measures of efficacy (Rand, Gibson & Dembo, TLC, and RSA) were compared to two measures thought to be distinct from teacher efficacy (Affect for teaching (Guskey, 1987) and Teaching self-concept (Guskey, 1987)), the relationships found were in the same range as those between various instruments attempting to measure teacher efficacy (0.22–0.54).

3.2. Guskey and Passaro's challenge

Guskey and Passaro (1994) attempted to add clarity to the meaning of these two factors of Gibson and Dembo's TES by modifying the wording of the items. They noted that all of the 11 items on the Gibson and Dembo TES that loaded on the personal teaching efficacy factor were worded positively and thus geared to an internal orientation ("I can"), whereas the items that loaded on the second factor, labeled general teaching efficacy, were negatively worded, consistently reflecting an external orientation, ("teachers can't"). When Guskey and Passaro reworded the personal efficacy items so that half reflected an internal and half an external orientation, and did the same with the general teaching efficacy items, the results conformed to an internal/external dichotomy rather than the personal/general dimensions. (See Table 3 for an example of the rewording.) The finding that the internal and external factors were only moderately correlated ($r - 0.24$) suggests that the internal and external dimensions are separate dimensions, not opposite ends of the same continuum. Thus, as Guskey and Passaro noted, these factors are not identical to the internal/external distinction made in locus-of-control or attribution theories of motivation. Guskey and Passaro concluded that

> The internal and external distinction identified in this study more accurately represents teachers' perceptions of the strength of different and independent factors. The internal factor appears to represent perceptions of *personal* influence, power, and impact in teaching and learning situations... . The external factor, on the other hand, relates to perceptions of the influence, power, and impact of elements that lie *outside the classroom* and, hence, may be beyond the direct control of individual teachers (p. 639).

This challenge provokes further reflection on the meaning of the two factors that have often been found in measures of teacher efficacy. These findings invite us to question once again the nature of teacher efficacy and how it can best be measured.

4. The development of a new measure of teacher efficacy

Deciding how to measure teacher efficacy presents thorny issues. Bandura (1997, 2001) recommended including various levels of task demands, allowing respondents to indicate the strength of their efficacy beliefs in light of a variety of impediments or obstacles and providing a broad range of response options. But perhaps the greatest challenge has to do with finding the optimal level of specificity for measurement. Although Bandura would applaud efforts to expand measures of teacher efficacy beyond single-item measures, which often are unreliable and cannot capture multifaceted dimensions of the construct, he nonetheless finds most currently available measures of teachers' sense of efficacy to be too general. Pajares (1996) complained that, in relation to student self-efficacy, global measures obscure what is being measured,

Table 3
Guskey and Passaro's rewording of teacher efficacy items[a]

Example of alternative forms of an item Item: Gibson and Dembo (1984) Number 15, and Woolfolk and Hoy (1990) Number 8	
Personal-internal (P-I) (original Item)	When I really try, I can get through to most difficult students
Personal-external (P-E)	Even when I really try, it is hard to get through to the difficult students
Teaching-internal (T-I)	When teachers really try, they can get through to most difficult students
Teaching-external (T-E)	Even when they really try, it is hard for teachers to get through to the difficult students

[a] Source: Guskey and Passaro (1994); used with permission.

> Omnibus tests that aim to assess general self-efficacy provide global scores that decontextualize the self-efficacy-behavior correspondence and transform self-efficacy beliefs into a generalized personality trait rather than the context-specific judgment Bandura suggests they are... . The problem with such assessments is that students must generate judgments about their academic capabilities without a clear activity or task in mind. As a result, they generate the judgments by in some fashion mentally aggregating to related perceptions that they hope will be related to imagined tasks (p. 547).

On the other hand, Pajares noted that, "specificity and precision are often purchased at the expense of external validity and practical relevance" (p. 561). There is a danger of developing measures that are so specific they lose their predictive power for anything beyond the specific skills and contexts being measured (e.g., I am confident I can teach simple subtraction to middle-income second graders in a rural setting who do not have specific learning disabilities, as long as my class is smaller than 22 students and good manipulatives are available). Discerning what is the most useful level of specificity depends on the purposes of the research, but either extreme of highly general or highly specific may pose problems for researchers.

In order to be useful and generalizable, measures of teacher efficacy need to tap teachers' assessments of their competence across the wide range of activities and tasks they are asked to perform. The Tschannen-Moran et al. (1998) model of teacher efficacy suggests that a valid measure of teacher efficacy must assess both personal competence and an analysis of the task in terms of the resources and constraints in particular teaching contexts. Most existing measures of teacher efficacy do not include both dimensions of efficacy. For example, the first Rand item and other measures of general teaching efficacy tend to assess just the external constraints faced by teachers and not the resources, while the second Rand item and other measures of personal teaching efficacy assess teaching strengths but not personal challenges. Studies need to test the relative predictive power of assessments of personal competence and of the analysis of the task. Certainly some context is inferred in assessments of personal competence (presumably those the person has had experience with) but a more careful and fine-grained assessment of those factors that both facilitate and impede teaching in a particular teaching context is likely to produce more powerful instruments.

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 (Henson, Bennett, Sienty, & Chambers, 2000; Tschannen-Moran et al., 1998). Although Bandura's instrument addresses many of the issues of measurement he has raised, problems remain. For example, both teachers and teacher educators who examined Bandura's instrument (in a seminar described below) were concerned that the distribution of items within the seven subscales did not accurately reflect the kinds of tasks that typically make up a teacher's work life.

4.1. Instrument development

Work on a new measure of efficacy was undertaken by participants in a seminar on *self-efficacy in teaching and learning* in the College of Education at The Ohio State University. The seminar included two researchers and eight graduate students.[1] The graduate students included two teacher educators, two full time doctoral students, and four practicing teachers. All eight had teaching experience, ranging from 5 to 28 years, with a mean of 11.9.

Several possible formats for a new efficacy measure were explored, including a Likert-type scale similar to the Gibson and Dembo instrument and the expanded scale advocated by Bandura. In the end, the group decided on a measure based on Bandura's scale, but with an expanded list of

[1]The authors are grateful to the seminar participants, Candace Fox, Pam Gaskill, Angela Lee, Matthew Maurer, Leena Patel, Hilary Raymond, Georgene Risko, and Hsiu-Han Yu for their work in generating items for the proposed measure.

teacher capabilities. Each seminar member independently selected items from the Bandura scale that she or he believed represented important tasks or elements of teaching. In addition, each member generated 8–10 new items to reflect areas of teaching not represented on the Bandura Scale, such as assessment. This process produced over 100 items, though there were many overlaps and similarities among items.

All of the items were pooled and the group then discussed each item as a possible candidate for the final scale, attempting to reach consensus on each item or revise items to capture important and frequently nominated areas of teaching. Using this nomination, discussion, and revision approach, 52 items were generated to assess the full range of teaching tasks and capabilities. From Bandura's 30-item scale, 23 items were retained and 7 were discarded as not being representative of frequent activities within a teachers' work life. Items not included were

- How much can you influence the class sizes in your school?
- How much can you do to get community groups involved in working with the school?
- How much can you do to get churches involved in working with the school?
- How much can you do to get businesses involved in working with the school?
- How much can you do to get local colleges and universities involved in working with the school?
- How much can you help other teachers with their teaching skills?
- How much can you do to enhance collaboration between teachers and the administration to make the school run effectively?

The remaining 19 items generated by the group described significant tasks of teaching not represented on the Bandura scale, such as assessment, adjusting the lesson to individual student needs, dealing with learning difficulties, repairing student misconceptions, and motivating student engagement and interest. A 9-point scale was used for each item, with anchors at 1—nothing, 3—very little, 5—some influence, 7—quite a bit, and 9—a great deal.

Sample items include

- How much can you do to motivate students who show low interest in schoolwork?
- How much can you assist parents in helping their children do well in school?
- How much can you gauge student comprehension of what you have taught?
- To what extent are you able to tailor your lessons to the academic level of your students?

5. Testing the instrument

The new measure, named the *Ohio State teacher efficacy scale* (*OSTES*), was examined in three separate studies. In the first study, the original 52 items were reduced to 32 and in the second, the scale was further reduced to 18 items made up of three subscales. In the third study, 18 additional items were developed and tested. The resulting instrument had two forms, a long form with 24 items and a short form with 12 items. Finally, the factor structure, reliability, and validity of the new measure was examined, as well as the appropriateness of the new scale for both preservice and inservice teacher populations.

5.1. Study 1

In the first study, the 52-item scale was refined and reduced to 32 items. The sample, importance ratings for each item, and results of a factor analysis are described below.

The sample. The instrument was tested on a sample of 224 participants, including 146 preservice teachers (124 females and 22 males) and 78 inservice teachers (43 females and 35 males). All were taking classes at The Ohio State University. The preservice teachers ranged in age from 18 to 47 years (mean = 23.9, SD = 5.5). The inservice teachers ranged in age from 20 to 56 years (mean = 31.6, SD = 7.2). The sample included 184 European Americans, 23 African Americans, 4 Latinos/Latinas, 3 Asian Americans, and 10 who self-identified as other.

Importance ratings and factor analysis. In addition to responding to each of the 52 items using the

9-point scale described above, respondents were asked to rate the importance of each item for effective teaching on a 4-point scale (*not at all, somewhat, important, or critical*). There was very little variability in the importance ratings of the 52 items. All tasks were considered "important" to "critical" for effective teaching. Thus no items were eliminated based on importance ratings.

The 52 items were submitted to principal-axis factoring with varimax rotation. Ten factors emerged with eigenvalues greater than one, accounting for 57.2% of the variance in the respondents' scores. Rotation failed to converge after 25 iterations, so the unrotated factor matrix was examined. The first factor had an eigenvalue of 20.7 and accounted for 39.9% of the variance in respondents' scores. Examining the first factor, we set as a criterion loadings higher than 0.60 to select items for further analysis. This yielded 31 items with loading ranging from 0.62 to 0.78. One item with a loading of 0.595 was included as well because it pertained to the important area of motivation, a topic that the seminar group of teachers believed was a critical task of teaching and not well represented in the 31 items chosen. Thus we selected 32 of the original 52 items for further testing.

5.2. Study 2

Another group of inservice and preservice teachers participated in the second study. Based on the procedures described below, the 32-item scale was further reduced to 18 items with three subscales.

The sample. The sample of 217 participants in the second study included 70 preservice teachers (49 female, 20 male, 1 no indication) and 147 inservice teachers (94 female, 53 male) and 3 respondents who failed to indicate their teaching status. The participants were students at three universities (Ohio State, William and Mary, and Southern Mississippi). The inservice teachers had from 1 to 26 years of experience with a mean of 8.5 (SD = 6.3). The preservice teachers ranged in age from 20 to 46 years (mean = 27.5, SD = 6.9), while the inservice teachers ranged in age from 22 to 62 years (mean = 33.5, SD = 8.5). The sample included 172 European Americans, 22 African Americans, 6 Latinos/Latinas, 6 Asian Americans, and 8 who self-identified as other.

Factor analysis and reliabilities. Principal-axis factoring with varimax rotation of the 32-item scale yielded eight factors with eigenvalues greater than one, accounting for 63% of the variance in the respondents' scores. A scree test suggested two or three factors could be extracted, thus we examined each of these solutions. In the two-factor solution, items related to classroom management loaded across both factors almost equally, but the loadings were low. In the three-factor solution, management emerged as a separate factor and the other two factors were more clearly specified. Because classroom management is an important element in effective teaching (Brophy & Good, 1986) and of concern to beginning teachers (Veenman, 1984) we believed that the three-factor solution better represented the tasks of teaching. This three-factor solution appeared both parsimonious and interpretable, thus we examined this solution to identify possible items to eliminate and further reduce the scale.

The scale was further reduced to 18 items by removing items that had the lowest loadings within each of the three factors, items that loaded clearly on more than one factor, and items that seemed redundant. For example, "How much can you do to adjust your lessons to the proper level for individual students?" and "To what extent are you able to tailor your lessons to the academic level of your students?" both loaded on the same factor and were moderately correlated ($r = 0.54$). Therefore we eliminated the second item because it had a lower loading on the factor.

The three factors, accounting for 51% of the variance, emerged from the varimax rotation of the 18 items in the respondents' scores. We labeled these factors, *efficacy for student engagement* (8 items), *efficacy for instructional strategies* (7 items), and *efficacy for classroom management* (3 items). An efficacy subscale score was computed for each factor by calculating the mean of the responses to the items retained within each factor. α reliabilities for the subscales were 0.82 for *engagement*, 0.81 for *instruction*, and 0.72 for *management*.

Second-order factor analysis of combined data. Using the responses from both Study 1 and Study 2, a principal-axis factoring of the three teacher efficacy subscales (*engagement, instruction,* and *management*) revealed one strong factor with factor loadings ranging from 0.74 to 0.84. The emergence of this second-order factor and the moderate positive correlations of the three subscales suggested that the 18 items could be considered to measure the underlying construct of efficacy and that a total score as well as three subscale scores could be calculated based on the 18 items. To further examine the appropriateness of calculating a total score for the 18 items, we conducted a principal-axis factor analysis specifying one factor. All 18 items loaded on this factor, with loadings ranging from 0.48 to.70. The reliability for this 18-item scale was 0.95.

Construct validity. In order to further test the validity of the OSTES, we examined construct validity by assessing the correlation of this new measure with other existing measures (Kerlinger, 1986). Participants in Study 2 were asked to respond not only to the OSTES, but also to the Rand Items, the Hoy and Woolfolk (1993) 10-item adaptation of the Gibson and Dembo TES, the pupil control ideology form (Willower, Eidell, & Hoy, 1967), and the work alienation scale (Forsyth & Hoy, 1978). As expected, total scores on the OSTES were positively related to both the Rand items ($r = 0.35$ and 0.28, $p<0.01$) as well as to both the personal teaching efficacy (PTE) factor of the Gibson and Dembo measure ($r = 0.48$, $p<0.01$) and the general teacher efficacy (GTE) factor ($r = 0.30$, $p<0.01$).

Discriminant validity for teacher efficacy was measured using a survey of work alienation because alienation was presumed to be conceptually distinct and negatively related to teacher efficacy. "Work alienation is defined in terms of the extent to which individuals fail to experience intrinsic pride or meaning in their work" (Forsyth & Hoy, 1978, p. 85). Results indicated that teacher efficacy was significantly negatively related to work alienation ($r - 0.31$, $p<0.01$). Pupil Control Ideology is the extent to which a teacher takes a custodial rather than a humanistic stance toward students (Willower et al., 1967) and has been related to teachers' sense of efficacy as measured by the Gibson and Dembo instrument (Woolfolk & Hoy, 1990; Woolfolk, Rosoff, & Hoy, 1990). As expected, teacher efficacy, as measured by the OSTES, was found to be negatively related to pupil control ideology, that is, teachers with a greater sense of their own efficacy tended to be less custodial in their attitudes towards students ($r - 0.25$, $p<0.01$). To insure that these correlations were not skewed by the inclusion of preservice teachers, the correlations were run again using only the responses of inservice teachers with very similar results.

The findings of Study 2 were encouraging. The 18-item instrument had good validity and the factors were conceptually sound representations of the various tasks of teaching. The weakness of the management factor as well as the strength of the instructional strategies and student engagement factors, however, led us to design a third study that would bolster the weaknesses and enhance the strengths of the nascent instrument.

5.3. Study 3

The purpose of the third study was to further refine the OSTES. Roberts and Henson (2001) echoed our concerns about the 18-item instrument. In a confirmatory factor analysis with a sample of 183 inservice teachers, they also found that the classroom management factor of the 18-item scale was weak and recommended its elimination. But our experience with both preservice and inservice teachers convinced us that classroom management is an important element of teaching. In addition, we suspected that the weakness of the management factor might be the consequence of the brevity of the 3-item scale. So, rather than eliminate this scale, we decided to write more items to capture this potentially important dimension of teacher efficacy. To develop new management items, we consulted Emmer's (1990) teacher efficacy for classroom management scale. We also took the opportunity to include items that assessed additional aspects of teaching that have been neglected in the measurement of teacher efficacy. Because the original Rand items focused on difficult and unmotivated students, most measures since that

time have continued in this vein. Thus, these instruments have overlooked the instructional challenges of responding to the needs of capable students as well as using a variety of instructional strategies to promote student thinking.

The resulting measure was field-tested in a class at the Ohio State University, *psychological perspectives on teachers, teaching, and teacher education*. Of the 19 people in the class, 17 were teachers and 2 were teacher educators. Based on their feedback, several questions were added and the wording of other items was modified. The final instrument included 36 items.

The sample. A sample of 410 participants in the third study included 103 preservice teachers (84 female, 15 male) and 255 inservice teachers (170 female, 84 male, 1 no indication), and 38 respondents who failed to indicate their teaching experience. The participants were students at three universities (Ohio State, William and Mary and Cincinnati) as well as teacher volunteers from two elementary, one middle, and one high school. The inservice teachers had from 1 to 29 years of experience with a mean of 8.2 (SD = 6.8). The preservice teachers ranged in age from 18 to 52 years (mean = 24.5, SD = 5.7); the inservice teachers ranged in age from 21 to 57 years (mean = 34.8, SD = 9.8). The sample included 332 European Americans, 38 African Americans, 3 Latinos/Latinas, 7 Asian Americans/Pacific Islanders, and 10 who self-identified as other. Of those who indicated the grade level at which they taught, 29% taught high school, 29% taught middle school, 37% taught elementary grades and 5% taught preschool.

Factor analysis and reliabilities. Principal-axis factoring with varimax rotation of the 36-items yielded four factors with eigenvalues greater than one, accounting for 58% of the variance in the respondents' scores. A scree test suggested three factors could be extracted. This solution replicated the three factors identified in Study 2—*efficacy for instructional strategies* (15 items), *efficacy for classroom management* (9 items), and *efficacy for student engagement* (12 items). We reduced the scale by selecting the eight items with the highest loadings on each factor. Using these 24 items, principal-axis factoring with varimax rotation yielded the same three factors with loadings ranging from 0.50 to 0.78. See Table 4 for factor loadings and eigenvalues for the 24-item scale. An efficacy subscale score was computed for each factor by calculating the mean of the eight responses to the items loading highest on that factor. Reliabilities for the teacher efficacy subscales were 0.91 for *instruction*, 0.90 for *management*, and 0.87 for *engagement*. Intercorrelations between the subscales of *instruction*, *management*, and *engagement* were 0.60, 0.70, and 0.58, respectively ($p < 0.001$). Means for the three subscales, ranging from 6.71 to 7.27 in the Study 3 sample, are displayed in Table 5.

Based on the high reliabilities of the three scales we explored the possibility that an even more parsimonious scale would be viable. When we selected the four items with the highest loadings on each scale, the factor structure remained intact (see Table 4) and the reliabilities continued to be high: 0.86 for *instruction*, 0.86 for *management*, and 0.81 for *engagement*. Furthermore, the intercorrelations between the short and long forms for the total scale and the three subscales were high, ranging from 0.95 to 0.98. Consequently, we tested both the long (24 items) and short form (12 items) in further analyses.

5.4. Factor structures for preservice and inservice teachers

Next, both the 24-item and the 12-item forms were subjected to two separate factor analyses, one using the responses of preservice teachers ($N = 111$), and the other using the responses of inservice teachers ($N = 255$). Principal-axis factoring with varimax rotation revealed three strong factors for the inservice teachers, the same three-factor structure as was found in Study 2. These three actors accounted for 54% (long form) and 65% (short form) of the variance in the inservice teachers' responses. The factor structure for preservice teachers was less distinct, therefore it appeared that the best solution for preservice teachers was a single factor. When the principal-axis factoring was conducted calling for only one factor to be extracted from the preservice teachers' responses, on both the 24- and 12-item scales all

Table 4
Factor loadings for the OSTES (study 3)

Ohio State teacher efficacy scale (OSTES)	24 items	12 items
Factor 1: *Efficacy for instructional strategies*		
1. To what extent can you use a variety of assessment strategies?	0.72	0.73
2. To what extent can you provide an alternative explanation or example when students are confused?	0.70	0.75
3. To what extent can you craft good questions for your students?	0.68	0.63
4. How well can you implement alternative strategies in your classroom?	0.66	0.73
5. How well can you respond to difficult questions from your students?	0.66	
6. How much can you do to adjust your lessons to the proper level for individual students?	0.59	
7. To what extent can you gauge student comprehension of what you have taught?	0.57	
8. How well can you provide appropriate challenges for very capable students?	0.55	
Factor 2: *Efficacy for classroom management*		
9. How much can you do to control disruptive behavior in the classroom?	0.78	0.83
10. How much can you do to get children to follow classroom rules?	0.69	0.66
11. How much can you do to calm a student who is disruptive or noisy?	0.66	0.63
12. How well can you establish a classroom management system with each group of students?	0.66	0.61
13. How well can you keep a few problem students from ruining an entire lesson?	0.62	
14. How well can you respond to defiant students?	0.61	
15. To what extent can you make your expectation clear about student behavior?	0.53	
16. How well can you establish routines to keep activities running smoothly?	0.50	
Factor 3: *Efficacy for student engagement*		
17. How much can you do to get students to believe they can do well in schoolwork?	0.75	0.75
18. How much can you do to help your students value learning?	0.70	0.69
19. How much can you do to motivate students who show low interest in schoolwork?	0.66	0.64
20. How much can you assist families in helping their children do well in school?	0.63	0.62
21. How much can you do to improve the understanding of a student who is failing?	0.57	
22. How much can you do to help your students think critically?	0.56	
23. How much can you do to foster student creativity?	0.50	
24. How much can you do to get through to the most difficult students?	0.47	

	Long form		Short form	
	Eigenvalue	Cum %	Eigenvalue	Cum %
Factor1	10.38	43.25	5.68	47.30
Factor 2	2.03	51.72	1.51	59.89
Factor 3	1.62	58.47	1.11	69.10

Table 5
Means for OSTES subscales and total score for long and short forms

	Long form			Short form		
	Mean	SD	α	Mean	SD	α
OSTES	7.1	0.94	0.94	7.1	0.98	0.90
Instruction	7.3	1.1	0.91	7.3	1.2	0.86
Management	6.7	1.1	0.90	6.7	1.2	0.86
Engagement	7.3	1.1	0.87	7.2	1.2	0.81

items loaded on this factor with factor loadings ranging from 0.60 to 0.85 and accounted for 57% and 61% of the variance, respectively.

Second-order factor analysis. Using data from the entire sample in Study 3, principal-axis factoring of the three teacher efficacy subscales (*instruction, management* and *engagement*) from the 24-item instrument revealed one strong factor accounting for 75% of the variance; and with the 12-item instrument again one factor emerged, accounting for 68% of the variance. The emergence of this second-order factor and the moderate positive correlations of the three subscales (see Table 6) suggested that both the 24 and 12-item scales could be considered to measure the underlying construct of efficacy and that a total score as well as three subscale scores could be calculated. To further examine the appropriateness of calculating a total score for the 24 and 12 items, we conducted a principal-axis factor analysis specifying one factor. All items loaded on this factor, with loadings ranging from 0.49 to 0.76 for the long scale and from 0.49 to 0.75 for the short form. The reliability for the 24-item scale was 0.94 and for the 12-item scale was 0.90. (See Table 5.) Thus both the subscale scores and the total score for both forms can be used to assess efficacy. However, as noted above, for preservice teachers, the total score seems to be the most appropriate gauge of efficacy. Subscale scores may have little meaning for prospective teachers who have yet to assume real teaching responsibilities.

Construct validity. We examined the construct validity of the short and long forms of the OSTES by assessing the correlation of this new measure and other existing measures of teacher efficacy (Kerlinger, 1986). Participants in Study 3 responded not only to the OSTES, but also to the Rand Items and the Hoy and Woolfolk (1993) 10-item adaptation of the Gibson and Dembo TES (See Table 6.) As expected, total scores on the OSTES (24-item long form) were positively related to both the Rand items ($r = 0.18$ and 0.53, $p < 0.01$) as well as to both the personal teaching efficacy (PTE) factor of the Gibson and Dembo measure ($r = 0.64$, $p < 0.01$) and the general teacher efficacy (GTE) factor ($r = 0.16$, $p < 0.01$). For the short form, the results proved to be similar and are reported in the bottom half of Table 6. Clearly, the strongest correlations between the OSTES and other measures are with scales that assess personal teaching efficacy. Once again, the lower correlations between GTE and other measures of efficacy suggest that this scale is the least successful in capturing the essence of efficacy.

The results of these analyses indicate that the OSTES could be considered reasonably valid and reliable. With either 24 or 12 items, it is of reasonable length and should prove to be a useful tool for researchers interested in exploring the construct of teacher efficacy.[2] Positive correlations with other measures of personal teaching efficacy provide evidence for construct validity. But the OSTES moves beyond previous measures to capture a wider range of teaching tasks. Both the Rand and Gibson and Dembo instruments focused on coping with student difficulties and disruptions as well overcoming the impediments posed by an unsupportive environment. Lacking were assessments of teaching in support of student thinking, effectiveness with capable students, creativity in teaching, and the flexible application of alternative assessment and teaching strategies. The OSTES addresses some of these limitations by including items that assess a broader range of teaching tasks. 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. The OSTES as well as other scales mentioned in this article are available at: http://www.coe.ohio-state.edu/ahoy/research-instruments.htm

6. Implications and directions for future research

The development of the OSTES is a step forward in capturing what has been an elusive construct. It is superior to previous measures of teacher efficacy in that it has a unified and stable factor structure and assesses a broad range of

[2]The instrument is copyrighted by the authors, however, there are no copyright restrictions on the instrument for use in scholarly research and for non-profit educational purposes.

Table 6
Validity correlations[a]

	OSTES	Instruct	Manage	Engage	Rand 1	Rand 2	GTE	PTE
OSTES		0.89**	0.84**	0.87**	0.18**	0.53**	0.16**	0.64**
Instructional strategies	0.84**		0.60**	0.70**	0.07	0.45**	0.06	0.62**
Classroom management	0.79**	0.46**		0.58**	0.29**	0.46**	0.30**	0.45**
Student engagement	0.85**	0.61**	0.50**		0.11*	0.47**	0.06	0.58**
Rand 1	0.18**	0.08	0.26**	0.11*		0.23**	0.65**	0.12*
Rand 2	0.52**	0.45**	0.39**	0.45**	0.23**		0.13*	0.65**
General teaching efficacy	0.16**	0.08	0.26**	0.06	0.65**	0.13*		0.07
Personal teacher efficacy	0.61**	0.60**	0.37**	0.56**	0.12*	0.65**	0.07	

[a] Above diagonal, long form (24 items); below diagonal, short form (12 items); ** $p<0.01$ (2-tailed); * $p<0.05$ (2-tailed).

capabilities that teachers consider important to good teaching, without being so specific as to render it useless for comparisons of teachers across contexts, levels, and subjects. Clearly this new scale needs further testing and validation. Clarification of the meaning of teacher efficacy and the relative weight of teachers' assessments of their skills and liabilities in light of the resources and constraints they face in particular teaching contexts promises to aid both those who would study and those who train teachers.

Even in its nascent form, however, this instrument opens new possibilities for research. Because efficacy beliefs are presumed to be relatively stable once set, more information is needed as to the factors that contribute to efficacy judgments and how efficacy beliefs are established (Hoy & Woolfolk, 1990). Longitudinal studies following preservice teachers through their training and first years in the field would be instructive. How do the efficacy beliefs of the supervising teacher in the practicum and subsequent mentors impact the sense of efficacy of the novice? Many schools have initiated mentoring programs for teachers in their induction year. What features of mentoring have the greatest impact on efficacy beliefs? What are the effects of the teaching environment and context? What structural features and supports make a difference in the formation of efficacy beliefs? For example, do teachers who start in middle schools with a team structure have higher efficacy than teachers who begin their careers in a traditional departmental structure? What leadership behaviors on the part of the principal make a difference?

There is also work to be done to understand efficacy beliefs among inservice teachers. For example, what kinds of challenges or changes are strong enough to provoke a reexamination of established efficacy beliefs? Theoretically, a teacher's efficacy beliefs will transfer to the extent that he or she perceives similarity in the task resources and constraints from one teaching situation to another (Bandura, 1997; Tschannen-Moran et al., 1998). To what extent would a change in grade level or curriculum generate such a reexamination? How much does a change in context, such as a move from an urban to a suburban or rural context, arouse a reassessment? Could the efficacy beliefs of teachers change in response to differing principal efficacy beliefs when there is a change of leadership at the school? Finally, how could a greater understanding of teachers' efficacy beliefs contribute to fostering greater equity in schools? Evidence suggests that the collective efficacy of a faculty can be a stronger predictor of student achievement than the socioeconomic level of the students (Bandura, 1993; Goddard, Hoy, & Woolfolk Hoy, 2000). What supports could build strong efficacy beliefs among teachers working with students of low socioeconomic status, students in racially diverse settings or urban contexts?

Exciting possibilities lay ahead as we learn more about this simple yet powerful idea. If the significant effects of teachers' beliefs in their capabilities were taken seriously, it could provoke significant changes in the way teachers were prepared and supported in their early years in the profession. Teacher preparation programs

could come to look more like apprenticeships, with a gradual shift from the vicarious experience and verbal persuasion of a university classroom to more mastery teaching experiences throughout the program, with steadily increasing levels of complexity and responsibility. There would be a gradual withdrawing of scaffolding and supports rather than the sink-or-swim practicum experiences many novice teachers now experience.

Taking seriously the potency of efficacy beliefs to impact teacher motivation and persistence over the course of a career could also well lead to a rethinking of the induction-year experiences of novice teachers, allowing for greater protection and support. The norms of teaching have often treated class assignments as rewards for status and seniority. In this kind of system, new teachers are given the least-desirable and most challenging teaching assignments to signify their position at the bottom of the status hierarchy. However, from an efficacy perspective this is a dysfunctional practice as it can lead novice teachers to adopt a low sense of efficacy that, unless reassessed by a significant change later, could result in decreased effort and enthusiasm for teaching for the length of one's teaching career. And finally, the professional development of teachers would be structured as powerful mastery experiences with an eye toward helping teachers garner evidence of improved learning on the part of their students in order to reap the efficacy pay-off that would result. In these days of hard-nosed accountability, teachers' sense of efficacy is an idea that neither researchers nor practitioners can afford to ignore. The OSTES is a promising tool for capturing this powerful construct and putting it to constructive use.

References

Allinder, R. M. (1994). The relationship between efficacy and the instructional practices of special education teachers and consultants. *Teacher Education and Special Education, 17*, 86–95.

Anderson, R., Greene, M., & Loewen, P. (1988). Relationships among teachers' and students' thinking skills, sense of efficacy, and student achievement. *Alberta Journal of Educational Research, 34*(2), 148–165.

Armor, D., Conroy-Oseguera, P., Cox, M., King, N., McDonnell, L., Pascal, A., Pauly, E., & Zellman, G. (1976). *Analysis of the school preferred reading programs in selected Los Angeles minority schools, REPORT NO. R-2007-LAUSD.* Santa Monica, CA: Rand Corporation (ERIC Document Reproduction Service No. 130 243).

Ashton, P. T., Buhr, D., & Crocker, L. (1984). Teachers' sense of efficacy: A self- or norm-referenced construct? *Florida Journal of Educational Research, 26*(1), 29–41.

Ashton, P. T., Olejnik, S., Crocker, L., & McAuliffe, M. (1982). Measurement problems in the study of teachers' sense of efficacy. *Paper presented at the annual meeting of the American Educational Research Association*, New York.

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

Bandura, A. (undated). *Teacher self-efficacy scale.* Available on-line at: http://www.coe.ohio-state.edu/ahoy/researchinstruments.htm#Ban

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191–215.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28*(2), 117–148.

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

Bandura, A. (2001). *Guide for constructing self-efficacy scales (Monograph).* Stanford, CA: Stanford University.

Berman, P., McLaughlin, M., Bass, G., Pauly, E., & Zellman, G. (1977). *Federal programs supporting educational change. Vol. VII: Factors affecting implementation and continuation (Report No. R-1589/7-HEW).* Santa Monica, CA: The Rand Corporation (ERIC Document Reproduction Service No. 140 432).

Brookover, W., Schweitzer, J., Schneider, C., Beady, C., Flood, P., & Wisenbaker, J. (1978). Elementary school social climate and student achievement. *American Educational Research Journal, 15*, 301–318.

Brophy, J. E., & Good, T. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.) (pp. 328–375). New York: Macmillan.

Burley, W. W., Hall, B. W., Villeme, M. G., & Brockmeier, L. L. (1991). A path analysis of the mediating role of efficacy in first-year teachers' experiences, reactions, and plans. *Paper presented at the annual meeting of the American Educational Research Association*, Chicago.

Coladarci, T. (1992). Teachers' sense of efficacy and commitment to teaching. *Journal of Experimental Education, 60*, 323–337.

Coladarci, T., & Breton, W. (1997). Teacher efficacy, supervision, and the special education resource-room teacher. *Journal of Educational Research, 90*, 230–239.

Coladarci, T., & Fink, D. R. (1995). Correlations among measures of teacher efficacy: Are they measuring the same thing? *Paper presented at the annual meeting of the American Educational Research Association*, San Francisco.

Emmer, E. (1990). A scale for measuring teacher efficacy in classroom management and discipline. *Paper presented at the annual meeting of the American Educational Research Association*, Boston, MA (Revised, June, 1990).

Emmer, E., & Hickman, J. (1990). Teacher decision making as a function of efficacy, attribution, and reasoned action. *Paper presented at the annual meeting of the American Educational Research Association*, Boston, MA.

Enochs, L. G., Posnanski, T., & Hagedorn, E. (1999). Science teaching self-efficacy beliefs: Measurement, recent research, and directions for future research. *Paper presented at the National Association of Research in Science Education*, Boston, MA.

Evans, E. D., & Tribble, M. (1986). Perceived teaching problems, self-efficacy and commitment to teaching among preservice teachers. *Journal of Educational Research, 80*(2), 81–85.

Forsyth, P. B., & Hoy, W. K. (1978). Isolation and alienation in educational organizations. *Educational Administration Quarterly, 14*, 80–96.

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

Glickman, C., & Tamashiro, R. (1982). A comparison of first-year, fifth-year, and former teachers on efficacy, ego development, and problem solving. *Psychology in Schools, 19*, 558–562.

Goddard, R. D., Hoy, W. K., & Woolfolk Hoy, A. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal, 37*, 479–508.

Greenwood, G. E., Olejnik, S. F., & Parkay, F. W. (1990). Relationships between four teacher efficacy belief patterns and selected teacher characteristics. *Journal of Research and Development in Education, 23*(2), 102–106.

Guskey, T. R. (1981). Measurement of responsibility teachers assume for academic successes and failures in the classroom. *Journal of Teacher Education, 32*, 44–51.

Guskey, T. R. (1982). Differences in teachers' perceptions of personal control of positive versus negative student learning outcomes. *Contemporary Educational Psychology, 7*, 70–80.

Guskey, T. R. (1984). The influence of change in instructional effectiveness upon the affective characteristics of teachers. *American Educational Research Journal, 21*, 245–259.

Guskey, T. R. (1987). Context variables that affect measures of teacher efficacy. *Journal of Educational Research, 81*(1), 41–47.

Guskey, T. R. (1988). Teacher efficacy, self-concept, and attitudes toward the implementation of instructional innovation. *Teaching and Teacher Education, 4*, 63–69.

Guskey, T., & Passaro, P. (1994). Teacher efficacy: A study of construct dimensions. *American Educational Research Journal, 31*, 627–643.

Hall, B., Burley, W., Villeme, M., & Brockmeier, L. (1992). An attempt to explicate teacher efficacy beliefs among first year teachers. *Paper presented at the annual meeting of the American Educational Research Association*, San Francisco.

Henson, R. K., Bennett, D. T., Sienty, S. F., & Chambers, S. M. (2000). The relationship between means-end task analysis and context specific and global efficacy in emergency certification teachers: Exploring a new model of efficacy. *Paper presented at the annual meeting of the American Educational Research Association*, New Orleans, LA.

Hoy, W. K., & Woolfolk, A. E. (1990). Socialization of student teachers. *American Educational Research Journal, 27*, 279–300.

Hoy, W. K., & Woolfolk, A. E. (1993). Teachers' sense of efficacy and the organizational health of schools. *The Elementary School Journal, 93*, 356–372.

Kerlinger, F. N. (1986). *Foundations of behavioral research.* New York: Holt, Rinehart & Winston.

Lee, V., Dedick, R., & Smith, J. (1991). The effect of the social organization of schools on teachers' efficacy and satisfaction. *Sociology of Education, 64*, 190–208.

Meijer, C., & Foster, S. (1988). The effect of teacher self-efficacy on referral chance. *Journal of Special Education, 22*, 378–385.

Midgley, C., Feldlaufer, H., & Eccles, J. (1989). Change in teacher efficacy and student self- and task-related beliefs in mathematics during the transition to junior high school. *Journal of Educational Psychology, 81*, 247–258.

Moore, W., & Esselman, M. (1992). Teacher efficacy, power, school climate and achievement: A desegregating district's experience. *Paper presented at the annual meeting of the American Educational Research Association*, San Francisco.

Newmann, F. M., Rutter, R. A., & Smith, M. S. (1989). Organizational factors that affect school sense of efficacy, community and expectations. *Sociology of Education, 62*, 221–238.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*, 533–578.

Parkay, F. W., Greenwood, G., Olejnik, S., & Proller, N. (1988). A study of the relationship among teacher efficacy, locus of control, and stress. *Journal of Research and Development in Education, 21*(4), 13–22.

Pintrich, P. R., & Schunk, D. H. (1996). *Motivation in education: Theory, research, and applications.* Englewood Cliffs, NJ: Merrill/Prentice-Hall.

Podell, D., & Soodak, L. (1993). Teacher efficacy and bias in special education referrals. *Journal of Educational Research, 86*, 247–253.

Raudenbush, S., Rowen, B., & Cheong, Y. (1992). Contextual effects on the self-perceived efficacy of high school teachers. *Sociology of Education, 65*, 150–167.

Riggs, I., & Enochs, L. (1990). Toward the development of an elementary teacher's science teaching efficacy belief instrument. *Science Education, 74*, 625–638.

Roberts, J. K., & Henson, R. K. (2001). A confirmatory factor analysis of a new measure of teacher efficacy: Ohio State Teacher Efficacy Scale. *Paper presented at the annual meeting of the American Educational Research Association*, Seattle.

Ross, J. A. (1992). Teacher efficacy and the effect of coaching on student achievement. *Canadian Journal of Education, 17*(1), 51–65.

Rose, J. S., & Medway, F. J. (1981). Measurement of teachers' beliefs in their control over student outcome. *Journal of Educational Research, 74*, 185–190.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs, 80*, 1–28.

Rubeck, M., & Enochs, L. (1991). A path analytic model of variables that influence science and chemistry teaching self-efficacy and outcome expectancy in middle school science teachers. *Paper presented at the annual meeting of the National Association of Research in Science Teaching*, Lake Geneva, WI.

Saklofske, D., Michaluk, B., & Randhawa, B. (1988). Teachers' efficacy and teaching behaviors. *Psychological Report, 63*, 407–414.

Skinner, E. A. (1996). A guide to constructs of control. *Journal of Personality and Social and Personality Psychology, 71*, 549–570.

Soodak, L., & Podell, D. (1993). Teacher efficacy and student problem as factors in special education referral. *Journal of Special Education, 27*, 66–81.

Soodak, L., & Podell, D. (1996). Teaching efficacy: Toward the understanding of a multi-faceted construct. *Teaching and Teacher Education, 12*, 401–412.

Stein, M. K., & Wang, M. C. (1988). Teacher development and school improvement: The process of teacher change. *Teaching and Teacher Education, 4*, 171–187.

Trentham, L., Silvern, S., & Brogdon, R. (1985). Teacher efficacy and teacher competency ratings. *Psychology in Schools, 22*, 343–352.

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

Veenman, S. (1984). Perceived problems of beginning teachers. *Review of Educational Research, 54*, 143–178.

Weiner, B. (1979). A theory of motivation for some classroom experiences. *Journal of Educational Psychology, 71*, 3–25.

Weiner, B. (1992). *Human motivation: Metaphors, theories, and research*. Newbury Park, CA: Sage.

Willower, D. J., Eidell, T. L., & Hoy, W. K. (1967). *The school and pupil control ideology. Penn State Studies Monograph No.* 24. University Park, PA: Pennsylvania State University.

Woolfolk, A. E., & Hoy, W. K. (1990). Prospective teachers' sense of efficacy and beliefs about control. *Journal of Educational Psychology, 82*, 81–91.

Woolfolk, A. E., Rosoff, B., & Hoy, W. K. (1990). Teachers' sense of efficacy and their beliefs about managing students. *Teaching and Teacher Education, 6*, 137–148.