Changes in teacher efficacy during the early years of teaching: 
A comparison of four measures

Anita Woolfolk Hoy\textsuperscript{a,*}, Rhonda Burke Spero\textsuperscript{b}

\textsuperscript{a}School of Education Policy and Leadership, Psychological Studies in Education, College of Education, The Ohio State University, 29 West Woodruff Ave. Columbus, OH 43210, USA
\textsuperscript{b}School of Education, University of Miami, 5202 University Drive, Coral Gables, FL 33146, USA

Abstract

Some of the most powerful influences on the development of teacher efficacy are mastery experiences during student teaching and the induction year. Bandura’s theory of self-efficacy suggests that efficacy may be most malleable early in learning, thus the first years of teaching could be critical to the long-term development of teacher efficacy. Yet few longitudinal studies exist that track efficacy across these early years. This study reports changes in teacher efficacy from entry into a teacher preparation program through the induction year. Multiple quantitative assessments of efficacy were used including Gibson and Dembo’s Teacher Efficacy Scale, Bandura’s assessment of Instructional Efficacy, and an instrument designed to reflect the specific context and goals of the preparation program studied. Results indicated significant increases in efficacy during student teaching, but significant declines during the first year of teaching. Changes in efficacy during the first year of teaching were related to the level of support received.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Teacher efficacy judgments; Preservice teachers; Teacher beliefs; Measurement of efficacy; Social context

1. Introduction

The role of self-efficacy in teaching and learning continues to interest researchers and practitioners alike. Teachers’ sense of efficacy—teachers’ judgments about their abilities to promote students’ learning—was identified almost 25 years ago as one of the few teacher characteristics related to student achievement in a study by the RAND corporation (Armor et al., 1976). Because teachers’ sense of efficacy appears to be a belief that affects teaching and learning, teacher educators, administrators, and policy makers are interested in the origins, supports, and enemies of efficacy.

Some of the most powerful influences on the development of teachers’ sense of efficacy are experiences during student teaching and the
induction year (Mulholland & Wallace, 2001). Bandura’s theory of self-efficacy suggests that efficacy may be most malleable early in learning, thus the first years of teaching could be critical to the long-term development of teacher efficacy. Yet few longitudinal studies exist that track efficacy across these early years. One goal of this study was to examine changes in teachers’ efficacy judgments from entry into a preparation program through the first year of teaching. In order to compare the results revealed by different measures, we used multiple quantitative assessments of self-efficacy for teaching, including Gibson and Dembo’s Teacher Efficacy Scale, Bandura’s Teacher Self-Efficacy Scale, and an instrument designed to reflect the specific context and goals of the preparation program studied.

Even though the early years of teaching likely are a critical time for developing efficacy beliefs, little is known about the kinds of context variables that affect efficacy. Social cognitive theory (Bandura, 1986, 1997) proposes that behavior, cognitive and other personal factors, and the environment interact to influence each other through the process of reciprocal determinism. The model of teacher efficacy presented by Tschannen-Moran, Woolfolk Hoy, and Hoy (1998) suggests that teachers make efficacy judgments, in part, by assessing the resources and constraints in specific teaching contexts. Resources in the form of support and feedback from colleagues, administrators, and parents could serve as social persuasion (Bandura, 1997). In addition the teaching resources available and the quality of the facilities could all impact teachers’ assessments about their ability to accomplish the tasks of teaching. However, research has not examined how such support is related to novice teachers’ developing efficacy judgments. Another goal of this study was to examine correlates of efficacy change during the first year of teaching.

2. Theoretical framework

Over a quarter century ago, Albert Bandura introduced the concept of self-efficacy or “beliefs in one’s capacity to organize and execute the courses of action required to produce given attainments” (Bandura, 1997, p. 3). Since that time, research in many arenas has demonstrated the power of efficacy perceptions in human learning, performance, and motivation. For example, efficacy beliefs are related to smoking cessation, adherence to exercise and diet programs, performance in sports, political participation, and academic achievement (Bandura, 1997).

Efficacy is a future-oriented judgment that has to do with perceptions of competence rather than actual level of competence. This is an important distinction because people regularly overestimate or underestimate their actual abilities, and these estimations may have consequences for the courses of action they choose to pursue and the effort they exert in those pursuits. “A capability is only as good as its execution. The self-assurance with which people approach and manage difficult tasks determines whether they make good or poor use of their capabilities. Insidious self-doubts can easily overrule the best of skills” (Bandura, 1997, p. 35). For example, Bouffard-Bouchard, Parent and Larivee (1991) found that children with the same level of skill development in mathematics differed significantly in their ability to solve math problems, depending on the strength of their efficacy beliefs. Children with higher efficacy more consistently and effectively applied what they knew; they were more persistent and less likely to reject correct solutions prematurely. In most cases, slightly overestimating one’s actual capabilities has the most positive effect on performance.

In the past two decades, researchers have found links between student achievement and three kinds of efficacy—the self-efficacy of students, the sense of efficacy of teachers, and the collective efficacy of schools (Goddard, Hoy, & Woolfolk Hoy, 2000; Pajares, 1996; Ross, 1992, 1994, 1998; Tschannen-Moran et al., 1998). Teachers’ sense of efficacy is the focus of this investigation.

2.1. Teachers’ sense of efficacy

In addition to being related to student achievement, teachers’ sense of efficacy has been associated with other student outcomes such as motivation (Midgley, Feldlaufer, & Eccles, 1989)
and students’ own sense of efficacy (Anderson, Greene, & Loewen, 1988). In addition, teachers’ efficacy beliefs appear to affect the effort teachers invest in teaching, their level of aspiration, and the goals they set. Teachers with a strong sense of efficacy tend to exhibit greater levels of planning, organization, and enthusiasm (Allinder, 1994) and spend more time teaching in subject areas where their sense of efficacy is higher (Riggs & Enochs, 1990), whereas teachers tend to avoid subjects when efficacy is lower (Riggs, 1995). Teachers with higher efficacy judgments tend to be more open to new ideas, more willing to experiment with new methods to better meet the needs of their students (Cousins & Walker, 2000; Guskey, 1988; Stein & Wang, 1988), and more committed to teaching (Coladarci, 1992). Efficacy beliefs influence teachers’ persistence when things do not go smoothly and their resilience in the face of setbacks (Gibson & Dembo, 1984). A greater sense of efficacy enables teachers to be less critical of students who make errors (Ashton & Webb, 1986), to work longer with a student who is struggling (Gibson & Dembo, 1984), and to be less inclined to refer a difficult student to special education (Meijer & Foster, 1988; Soodak & Podell, 1993).

How does teachers’ sense of efficacy affect student achievement? Ross (1994, 1998) reviewed 88 teacher efficacy studies and identified potential links between teachers’ sense of efficacy and their behaviors. Ross suggested that teachers with higher levels of efficacy are more likely to (1) learn and use new approaches and strategies for teaching, (2) use management techniques that enhance student autonomy and diminish student control, (3) provide special assistance to low achieving students, (4) build students’ self-perceptions of their academic skills, (5) set attainable goals, and (6) persist in the face of student failure.

Not all researchers agree that higher levels of efficacy are always positive influences, however. Wheatley (2002) identified a number of benefits for teacher learning and educational reform that might follow from having doubts about one’s efficacy. These include the possibility that doubts might foster reflection, motivation to learn, greater responsiveness to diversity, productive collaboration, and change provoking disequilibrium. We believe that a sense of efficacy for learning to teach would be necessary to respond to doubts in these positive ways, but the point is well taken that persistent high efficacy perceptions in the face of poor performance can produce avoidance rather than positive action.

### 2.2. The development of efficacy

Bandura (1977, 1997) postulated four sources of efficacy expectations: mastery experiences, physiological and emotional states, vicarious experiences, and social persuasion. Mastery experiences are the most powerful source of efficacy information. The perception that teaching has been successful raises efficacy expectations that teaching will be proficient in the future, unless the success required such massive work that the individual feels unable to sustain this level of effort. The perception that one’s teaching has been a failure lowers efficacy beliefs, contributing to the expectation that future performances will also be inept, unless the failure is viewed as providing clues about more potentially successful strategies. For novice teachers, mastery experience is an important source of efficacy beliefs (Mulholland & Wallace, 2001).

Attributions play a role as well. If the success is attributed to internal or controllable causes such as ability or effort, then self-efficacy is enhanced. But if success is attributed to luck or the intervention of others, then self-efficacy may not be strengthened (Pintrich & Schunk, 2001). The level of arousal, either of anxiety or excitement, adds to the feeling of mastery or incompetence, depending on how the arousal is interpreted. For example, feelings of tension can be interpreted as anxiety and fear that failure is imminent or as excitement—being “psyched” for a good class.

Vicarious experiences are those in which someone else models a skill. The degree to which the observer identifies with the model moderates the efficacy effect on the observer (Bandura, 1977). The more closely the observer identifies with the model, the stronger will be the impact on efficacy. When a credible model teaches well, the efficacy of the observer is enhanced. When the model performs poorly, the efficacy expectations of the observer decrease.
Social or verbal persuasion may entail a “pep talk” or specific performance feedback from a supervisor, colleague, or students. Student evaluations of instruction at the college level can be a form of verbal persuasion, for better or worse Heppner (1994). Social persuasion, though limited in its impact, may provide an “efficacy boost” to counter occasional setbacks that might have instilled enough self-doubt to interrupt persistence. The potency of persuasion depends on the credibility, trustworthiness, and expertise of the persuader (Bandura, 1986). For beginning teachers, a potent source of efficacy is feedback from students in the form of enthusiasm and engagement and verbal persuasion from experienced teachers in the form of encouragement and advice (Mulholland & Wallace, 2001).

Preservice and student teachers: The development of teacher efficacy beliefs among prospective teachers has generated a great deal of research interest because once efficacy beliefs are established they appear to be somewhat resistant to change. A fairly consistent finding is that sense of personal efficacy for teaching increases during college teacher preparation and student teaching (Hoy & Woolfolk, 1990; Spector, 1990). This trend has been identified in Canada (Housego, 1992), Korea (Gorrell & Hwang, 1995), and the United States (Wenner, 2001). But some researchers have found no change or a decline in level efficacy over the years of preparation (Israel: Romi & Daniel, 1999; Taiwan: Lin & Gorrell, 2001). It is possible that some of the discrepancies in findings can be traced to differences in the way efficacy is measured (Tschanne-Moran & Woolfolk Hoy, 2001).

Novice teachers: The first year of teaching can be a time of reality shock as the novice teacher is faced with all the role demands and expectations encountered by experienced teachers (Corcoran, 1981; Veenman, 1984). Weinstein (1988) described the “unrealistic optimism” of many beginning teachers in the United States and Friedman (2000) described the “shattered dreams of impeccable professional performance” (p. 595) experienced by first year teachers he interviewed in Israel. Friedman linked possible burnout in teaching to what he termed “professional efficacy discrepancy” (p. 597) or the distance between expected (if unrealistic) and actual levels of professional performance.

Although few studies have looked at the development of efficacy beliefs among novices, it seems that efficacy beliefs of first-year teachers are related to stress and commitment to teaching, as well as satisfaction with support and preparation. Novice teachers completing their first year of teaching who had a high sense of teacher efficacy found greater satisfaction in teaching, had a more positive reaction to teaching, and experienced less stress. Confident new teachers gave higher ratings to the adequacy of support they had received than those who ended their year with a shakier sense of their own competence and a less optimistic view of what teachers could accomplish. Efficacious beginning teachers rated the quality of their preparation higher and the difficulty of teaching lower than those who were less efficacious. And efficacious novices indicated greater optimism that they would remain in the field of teaching (Burley, Hall, Villeme, & Brockmeier, 1991; Hall, Burley, Villeme, & Brockmeier, 1992). Because efficacy beliefs are shaped early, it would be useful to better understand what supports and undermines efficacy in the early years.

Contextual influences on efficacy: Attention to the factors that support the development of a strong sense of efficacy among preservice and novice teachers is worthwhile because, as noted earlier, once established, efficacy beliefs of experienced teachers seem resistant to change. Evidence suggests that input during initial training has a different impact than input received after teachers are in the field (Tschanne-Moran et al., 1998). The new teachers interviewed in Friedman’s study (2000) who experienced “shattered dreams of idealist performance” (p. 595) pointed to criticisms from colleagues, isolation, work overload, lack of recognition or reward, and inappropriate initial teacher training as sources of stress and threats to efficacy. Longitudinal studies across teacher preparation programs and the first several years in the field could begin to map the development of efficacy beliefs. One purpose of this study was to initiate such a longitudinal study.
3. The measurement of teacher efficacy

The construct of teacher efficacy has been conceptualized in a number of ways, but the most pervasive is derived from two Rand Corporation evaluations of innovative educational programs funded by the Federal Elementary and Secondary Education Act (Armor et al., 1976; Berman, McLaughlin, Bass, Pauly, & Zellman, 1977). In these studies, teachers’ level of efficacy was determined by computing a total score for their responses to two 5-point Likert scale items: (a) "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," and (b) "If I try really hard, I can get through to even the most difficult or unmotivated students." The theoretical basis for these items was Rotter’s (1966) social learning theory.

3.1. Gibson and Dembo

In an attempt to improve on the validity and reliability of the Rand two-item scale, Gibson and Dembo (1984) developed a 30-item scale that yields two factors consistent with the Rand items. These researchers turned to Bandura’s cognitive social learning theory of self-efficacy to interpret the two factors. According to Bandura (1977), motivation is determined by people’s judgments of their capability to execute particular courses of action (called efficacy expectations) and their beliefs about the likely consequences of those actions (called outcome expectations). Gibson and Dembo labeled their first factor personal teaching efficacy (alpha = .75), and assumed this factor assessed self-efficacy. The second factor, teaching efficacy (alpha = .79), was assumed to capture outcome expectancy. Our research indicates, however, that this second dimension of efficacy does not represent an outcome expectation as defined by Bandura (1986). Instead it appears to reflect a general belief about the power of teaching to reach difficult children and may have more in common with teachers’ conservative/liberal attitudes towards education (Tschanne-Moran et al., 1998; Hoy & Woolfolk, 1990; Woolfolk & Hoy, 1990).

For this reason we have labeled the dimension, general teaching efficacy (GTE). The first factor appears to be the more accurate indicator of the teacher’s personal sense of efficacy and is labeled personal teaching efficacy (PTE) (Hoy & Woolfolk, 1993; Woolfolk & Hoy, 1990, Woolfolk, Rosoff, & Hoy, 1990).

Results of studies using either the Rand items or the Gibson and Dembo scale have consistently found that the two dimensions of efficacy, however conceptualized, are independent. (Anderson et al., 1988; Burley et al., 1991; Hoy & Woolfolk, 1993; Moore & Esselman, 1992; Saklofske, Michaluk & Randhawa, 1988; Soodak & Podell, 1993; 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 .15 to .25.

3.2. Bandura’s teacher self-efficacy scale

Bandura (1997) has offered his own Teacher Self-Efficacy Scale. He pointed out that teachers’ sense of efficacy is not necessarily uniform across the many different types of tasks teachers are asked to perform, or across different subject matter. In response, he constructed a 30-item instrument. This measure attempts to provide a multi-faceted picture of teachers’ efficacy beliefs without becoming too narrow or specific. Very little research is available, however, using Bandura’s scale.

3.3. Questions and hypotheses of the study

This study sought answers for the following questions:

(a) How does sense of efficacy change during student teaching and the first year of teaching for a group of students?
(b) Do different measures of teachers’ sense of efficacy reveal similar patterns of change?
(c) What are some possible factors in the first years of teaching that might be related to changes in efficacy?
We hypothesized that, consistent with much of the research, efficacy as measured by PTE would rise during student teaching whereas GTE would fall. We expected Bandura’s measure to perform similarly to PTE, though Bandura’s instrument has not been used in studies of student teaching. But because few studies have followed preservice teachers into their first year of teaching, we present no hypotheses about whether efficacy will increase, level, or fall during the first year of actual teaching for any of the measures. Based on the conception of efficacy as context specific and the first year of teaching is an important context for efficacy development, we hypothesized that, consistent with Bandura’s theory of self-efficacy, increases in efficacy would be related to perceptions of task mastery and support. We also hypothesized that efficacy as measured by the PTE, and Bandura’s measure would reveal similar patterns, whereas we expected GTE to be unrelated to the other two measures.

4. Method

The study was a longitudinal investigation that assessed the efficacy of prospective and novice teachers at the beginning of their preparation program, at the end of student teaching, and after their first year of employment as a teacher. All the members of the 1997–1998 elementary education Master’s of Education cohort at a major midwestern public university participated in the study.

4.1. Program and participants

Participants were 53 prospective teachers in the Master’s of Education initial teaching certification program. There were 38 females (71.7%) and 15 males (28.3%) in the group. Three of the females and two of the males were African American. The average age of the females was 24.7 (SD = 5.36) and of the males was 29.9 (SD = 8.85). Eight of the participants (15%) were married. Of the 40 graduates who were teaching 1 year after completing their degree, 29 returned questionnaires at the end of their first year of teaching. Thus, the final participant group included 53 students who provided responses at the beginning and end of their preparation program, 29 of whom went on to complete a year of teaching and return useable questionnaires.

The preparation program is based on a Holmes Group Professional Development School model. All students entered with an undergraduate degree and completed a Master’s of Education in five quarters. Students moved through their preparation as a cohort, taking classes and participating in all program activities together. The contents of the program included courses in PE/Health, Art Methods, Music Methods, Child Development/Learning, Technology, Educational Psychology/Pedagogy, Assessment, Mathematics Methods, Science Methods, Social Studies Methods, Children’s Literature, Literacy Methods, Equity and Diversity, and School & Society. Students were in school placements for most of the year, moving from visiting three days a week in the fall to 10 weeks of full time student teaching during the winter and spring. The program emphasized diversity and preparing teachers for urban settings.

4.2. Instrument and procedures

Students completed three instruments that yielded four measures of efficacy and also supplied background information about themselves and the schools in which they taught during their first year.

Gibson and Dembo short form: Factor analyses of all versions of the Gibson and Dembo Scale consistently have produced two independent dimensions of GTE and PTE (Hoy & Woolfolk, 1990; Woolfolk & Hoy, 1990). Students in the current study completed a 10-item version of the Teacher Efficacy Scale (Gibson & Dembo, 1984) adapted by Woolfolk and Hoy (1993). This scale consisted of four PTE, four GTE items, and the two original Rand items. These items were selected because they had the highest factor loadings in the earlier research. Example GTE items are: “If students aren’t disciplined at home, they aren’t likely to accept any discipline” and “If parents would do more for their children, I could do more.” Example PTE items are: “If a student did not remember information I gave in a previous
lesson, I would know how to increase his/her retention in the next lesson” and “When I really try, I can get through to most difficult students.” Response to each item is along a 6-point Likert scale from “strongly agree” to “strongly disagree.” For both the dimensions, the higher the score (the closer to 6), the more efficacious. In the current sample, alpha reliability coefficients were .73, .79, and .68 for PTE across the three administrations and .78, .78, and .84 for GTE. Thus the Gibson and Dembo instrument provide two measures of efficacy, GTE and PTE.

**Bandura Teacher Self-Efficacy Scale**: Bandura’s 30-item scale has 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.” Sample items include: “How much can you do to get through to the most difficult students?” “How much can you do to support learning when there is lack of support for the home?” “How much can you do to control disruptive behavior in the classroom?” “How much can you do to make parents comfortable coming to school?” In the current study, based on the average score for the entire 30-item scale, the alpha coefficients of reliability were .94, .95, and .92 across the three administrations (See Table 1). All items are scored such that a higher score indicates greater efficacy.

**Program-specific measure of efficacy**: In an attempt to identify an appropriate level of specificity for assessing efficacy in our preservice teacher preparation program, we surveyed all the instructors who worked with the prospective teacher cohorts, asking the instructors what students should be able to do after completing the coursework. After removing redundancies, the result was a list of 32 teaching skills such as manage classrooms, evaluate student work, use cooperative learning approaches, teach basic concepts of fractions, and build learning in science based on children’s intuitive understandings. We then designed a questionnaire, named the OSU Teaching Confidence Scale, that asked students to rate on a 6-point scale how confident they were in their ability to accomplish each skill, the higher the score, the more confident. In the current study, based on the average score for the entire 32-item scale, the alpha coefficients of reliability were .97, .95, and .70 across the three administrations (see Table 1).

**Data collection**: There were three phases of data collection: (a) Phase 1 during the first quarter of teacher preparation, before most course work had been completed; (b) Phase 2 at the end of the preparation program after student teaching was completed; (c) Phase 3 at the end of their first year of actual teaching. The first two data collections took place in classes during the teacher preparation program. Students completed all forms anonymously, but did create their own code (their mother’s birthday month and her maiden name). We asked participants to use this code on every administration so that we could match responses for participants across administrations. They were not required to participate and were assured that no part of their program would be affected by their answers to the questions. The final data collection was completed through the mail. Participants who did not return their first mailing were sent a second request.

### Table 1

**Correlations of efficacy measures and reliabilities for phase 1**

<table>
<thead>
<tr>
<th>Measures (n = 53)</th>
<th>BAN</th>
<th>GTE</th>
<th>PTE</th>
<th>OSU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandura (BAN)</td>
<td>.94</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GTE</td>
<td>.31*</td>
<td>.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTE</td>
<td>.52**</td>
<td>.26</td>
<td>.73</td>
<td></td>
</tr>
<tr>
<td>OSU</td>
<td>.52**</td>
<td>.21</td>
<td>.60**</td>
<td>.97</td>
</tr>
</tbody>
</table>

*p < .05; **p < .01.

Note: Reliabilities in italics on the diagonals.

### 5. Results

Factor analyses of the Gibson and Dembo 10-item scale yielded the two factors of GTE and PTE.
found in previous research. Thus two scales were created for these measures by calculating the means of the five items for each scale. Items were coded such that the higher the score (closer to 6), the higher the efficacy. Scores for Bandura’s instrument were based on the mean of all 30 items (the small sample size made factor analysis of this instrument unadvisable). Items were coded such that the higher the score (closer to 9), the higher the efficacy. This 30-item scale proved very reliable with alphas .92 or higher for each administration in this study. The OSU scale had high reliabilities for the total measure (alphas of .97, .95, and .70) across the three phases. Items were coded such that the higher the score (closer to 6), the greater the confidence.

In order to explore sources of efficacy during phase three, participants were asked to rate several aspects of the resources and support in their first year teaching assignment (quality of teaching resources provided, support from colleagues, support from administrators, support from parents, support from the community). Sources of support were rated on a scale from 1 (non-existent) to 9 (excellent). Factor analysis of these responses yielded one factor, so an index of support was created by calculating the mean of these five items. Items were coded such that the higher the score (closer to 9), the higher the degree of support. Reliability for this support scale was .76.

Finally, we also gathered the following data as part of the phase three surveys:

- **Perception of mastery 1**: participants’ ratings of their own success during the first year compared to other first year teachers in similar situations (1 = not successful at all, to 9 = highly successful);
- **Perception of mastery 2**: participants’ satisfaction with their professional performance in the first year (1 = not at all satisfied, to 9 = highly satisfied);
- **Assessment of the task difficulty**: their assessment of the difficulty of their teaching assignment (“Compared to other first year teachers’ classrooms, how challenging was your teaching assignment?” 1 = not at all, to 9 = much more challenging);
- **Assessment of the teaching context**: the SES level of their classroom (defined by the percent of students not on free and reduced lunch—the higher the score, the higher the SES of the class).

## 5.1. Changes in efficacy

The means and standard deviations for each of the four efficacy measures across the three phases of the study are given in Table 2; means in the same row that do not share the same subscript differ at \( p < .05 \). The means also are shown in Fig. 1. The figure is divided because Bandura’s measure has a different response range, thus means from the other 6-point scales cannot be compared directly to means based on Bandura’s 9-point scale. We did not change Bandura’s to a 6-point scale prior to administration because we wanted to maintain consistency with previous research and with Bandura’s recommendations for his scales. We decided not to mathematically transform the 9-point responses to 6-point

<table>
<thead>
<tr>
<th>Measures</th>
<th>Phase 1: beginning of preparation</th>
<th>Phase 2: after student teaching</th>
<th>Phase 3: after first year of teaching</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( M )</td>
<td>SD</td>
<td>( M )</td>
</tr>
<tr>
<td>Bandura</td>
<td>6.03(_a)</td>
<td>.82</td>
<td>6.60(_b)</td>
</tr>
<tr>
<td>GTE</td>
<td>3.85(_a)</td>
<td>1.03</td>
<td>4.50(_b)</td>
</tr>
<tr>
<td>PTE</td>
<td>4.12(_a)</td>
<td>.73</td>
<td>5.04(_b)</td>
</tr>
<tr>
<td>OSU</td>
<td>4.14(_a)</td>
<td>.96</td>
<td>5.17(_b)</td>
</tr>
</tbody>
</table>

*Note*: responses to Bandura were on a 9-point scale (1 = nothing, 9 = a great deal). Responses to GTE, PTE, and OSU were on a 6-point scale (1 = disagree, 6 = agree). Means in the same row that do not share the same subscript differ at \( p < .05 \).
responses, because this transformation would change the meaning of the scores; Bandura’s scale allows a midpoint response, whereas the 6-point scales do not have a midpoint response option.

We conducted a repeated measures analysis for each efficacy instrument with phase as the within-subject variable and efficacy score as the outcome variable. The results revealed statistically significant effects for Phase when efficacy was assessed using Bandura’s instrument \( F(2, 27) = 13.41, p < .0001, \eta^2 = .50 \), the GTE scale \( F(2, 27) = 7.06, p < .003, \eta^2 = .34 \), the PTE scale \( F(2, 27) = 20.69, p < .0001, \eta^2 = .61 \), and the OSU Teaching Confidence Scale \( F(2, 27) = 30.45, p < .0001, \eta^2 = .69 \), suggesting statistically significant changes in the participants’ efficacy scores over time.

Follow-up mean contrasts showed statistically significant differences for the Bandura instrument between Phase 1 and Phase 2 (the beginning and the end of student teaching) \( F(1, 28) = 25.32, p < .0001, \eta^2 = .48 \) and Phase 2 and Phase 3 (the end of student teaching to the end of the first teaching year) \( F(1, 28) = 13.59, p = .001, \eta^2 = .35 \). Significant differences also were found for the GTE scale between Phases 1 and 2 \( F(1, 28) = 9.62, p < .004, \eta^2 = .26 \) and Phases 2 and 3 \( F(1, 28) = 11.82, p = .002, \eta^2 = .30 \) and for the PTE scale between Phases 1 and 2 \( F(1, 28) = 40.24, p < .0001, \eta^2 = .59 \). But from Phases 2–3, the difference in PTE scores just exceeded the .05 level of significance \( F(1, 28) = 4.10, p = .052, \eta^2 = .13 \). For the OSU program specific measure of confidence, the difference from Phases 1–2 was significant \( F(1, 28) = 62.27, p < .0001, \eta^2 = .69 \), whereas the difference between Phases 2 and 3 was not significant \( F(1, 28) = 1.43, p = .242, \eta^2 = .05 \). The mean scores of the four instruments across the three time periods are shown in Fig. 1.

Thus for all four measures, the changes from the beginning of the program to the end of student teaching (phases 1–2) represented significant increases in efficacy. From the end of student teaching to the end of the first year of teaching (phases 2–3), the decreases in the Bandura and GTE scales were significant, the change in PTE approached significance, but the OSU scale level remained stable.

In order to compare the efficacy scores at the end of the first year of teaching to the entering scores, we conducted simple paired \( t \)-tests. Results indicated that from entry into the program to the end of the first year of teaching (phases 1–3), the increases in efficacy indicated by the PTE and OSU measures were significant. Thus, efficacy as assessed by the OSU measure rose during teacher preparation and held through the first year of teaching. Efficacy as assessed by the PTE scale rose and then fell (but still remained above the entering level), while efficacy as assessed by the Bandura and GTE scales rose significantly during preparation, but then after the first year of teaching returned to levels roughly equal to entering scores. A plot of individual cases for the Bandura measures showed that this pattern of increase then reversal held for 22 of the 29 cases. Of the seven not fitting this pattern, four decreased and three increased in efficacy, though not dramatically, over the three time periods.

**Measures of change during first year of teaching:**

We calculated change scores for the Bandura, OSU, GTE, and PTE scales by subtracting the individual’s phase 2 score from phase 3 score, so a positive change score indicated that efficacy increased during the first year of teaching, the
larger the change score, the greater the increase in efficacy. As shown in Table 3, change in the Bandura measure is correlated with change in all the other efficacy scales.

Correlates of change: Table 3 also shows correlations between changes in efficacy and the participants’ perceptions of mastery (ratings their success and satisfaction with their performance), perceived difficulty of their teaching assignment, the level of support they experienced (support index), and the SES of their students.

Surprisingly, one indicator of mastery, perception of success in relation to other teachers in similar first year settings, was not related to changes in efficacy, however assessed. Instead, perception of success was correlated with the SES of the class ($r = -.41, p < .05$; the lower the SES of the class, the higher the assessment of success). A second indicator of mastery, perceived satisfaction with performance in the first year, was correlated with changes in efficacy as assessed by the Bandura and GTE scales ($r = .43, p < .05$ and $r = .48, p < .01$, respectively—the greater the satisfaction, the larger the increases in efficacy).

Level of perceived support correlated positively with changes in efficacy as assessed by the Bandura, GTE, and PTE measures ($r$’s = .38, .37, and .37, respectively, $p$’s < .05). As perceptions of support also increased, efficacy increased as well. Finally, the greater the perceptions of support in the first year, the less difficult the teaching assignment was rated ($r = -.56, p < .01$).

Support was also correlated with the SES of the class, the higher the SES, the greater the perceptions of support ($r = .68, p < .01$).

6. Discussion

In general, three of the four measures of efficacy revealed similar patterns and changes over time. All scales were reliable, even with our small sample. The picture that emerged from these findings is that efficacy, however assessed, rose during teacher preparation and student teaching, but fell with actual experience as a teacher.

The only exception to this pattern came with the program specific OSU measure that asked teachers to assess their teaching activities and skills. Here confidence rose and held. It is possible that the teachers did feel capable of executing the strategies described in the OSU measure, but still were not sure that these actions or skills would lead to success as a teacher. Perhaps these first-year teachers were discovering that teaching is more than method and strategy, thus the drop in their other efficacy scores. Given these findings, we are reluctant to consider the confidence scale a true measure of perceived efficacy for influencing student outcomes. Rather this measure seems to assess confidence in the ability to enact different strategies, which may or may not lead to

Table 3
Means, standard deviations, and correlations matrix for efficacy change and context measures: phase 3 data ($n = 29$)

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
<th>Bch</th>
<th>GTE ch</th>
<th>PTE ch</th>
<th>OSU ch</th>
<th>MAS 1</th>
<th>MAS 2</th>
<th>DIFF</th>
<th>SUPT</th>
<th>SES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandura change (Bch)</td>
<td>-.68</td>
<td>.95</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>GTE change (GTEch)</td>
<td>-.66</td>
<td>1.03</td>
<td>.64**</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>PTE change (PTEch)</td>
<td>-.28</td>
<td>.77</td>
<td>.45*</td>
<td>.38</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>OSU change (OSUch)</td>
<td>-.13</td>
<td>.60</td>
<td>.45*</td>
<td>.41*</td>
<td>.36</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Mastery 1 (MAS 1)</td>
<td>7.38</td>
<td>1.35</td>
<td>.21</td>
<td>.22</td>
<td>.01</td>
<td>.34</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Mastery 2 (MAS 2)</td>
<td>6.69</td>
<td>1.75</td>
<td>.44*</td>
<td>.48**</td>
<td>.23</td>
<td>.32</td>
<td>.69**</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Difficulty (DIFF)</td>
<td>6.43</td>
<td>2.08</td>
<td>.03</td>
<td>.16</td>
<td>-.04</td>
<td>.02</td>
<td>-.12</td>
<td>-.15</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Support (SUPT)</td>
<td>5.28</td>
<td>1.52</td>
<td>.38*</td>
<td>-.02</td>
<td>.37*</td>
<td>.18</td>
<td>.07</td>
<td>.21</td>
<td>-.55*</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>SES</td>
<td>45.10</td>
<td>35.74</td>
<td>.28</td>
<td>-.11</td>
<td>.17</td>
<td>.05</td>
<td>-.41*</td>
<td>-.23</td>
<td>-.53*</td>
<td>.68**</td>
<td>—</td>
</tr>
</tbody>
</table>

*p < .05; **p < .01.

*RANGE = 1–9.

*Range = 1–100% of class not free and reduced lunch.
student learning. Change in confidence assessed by this program-specific measure did not correlate with self-perceived teaching success, satisfaction with professional performance, or perceptions of support.

As hypothesized, the GTE measure diverged from the other measures of efficacy, though scores followed a similar path of increase during student teaching and decrease in the first year. GTE was not correlated with the PTE or OSU measures and only moderately correlated with the Bandura measure.

One surprising finding was that, contrary to previous research (Hoy & Woolfolk, 1990), GTE rose during student teaching. It was not until the first year of actual teaching experience that GTE fell for the current group. For the prospective teachers in the Hoy/Woolfolk sample, student teaching was the first real experience in the classroom. Prospective teachers in the current study, however, were in a year-long internship accompanying their course work, so their immersion into teaching was gradual, until they assumed responsibility for their own class the following year as an employed teacher. Thus the current sample probably had more support and buffering during their student teaching experience than the Hoy/Woolfolk sample. When this support was withdrawn, efficacy fell. In fact, positive changes in efficacy as assessed by the Bandura and PTE measures correlated with level of support during the first year of teaching, so there is some evidence that support may be important in protecting efficacy during early teaching.

Why did efficacy go down with teaching experience? Student teaching and the teaching induction year provide opportunities to gather information about one’s personal capabilities for teaching. Prospective and novice teachers often underestimate the complexity of the teaching task and their ability to manage many agendas simultaneously. New teachers may either interact too much as peers with their students and find their classes out of control or they may grow overly harsh and end up not liking their “teacher self.” They become disappointed with the gap between the standards they have set for themselves and their own performance. Novice teachers sometimes engage in self-protective strategies, lowering their standards in order to reduce the gap between the requirements of excellent teaching and their self-perceptions of teaching competence, suggesting that the optimism of young teachers in some situations may be somewhat tarnished when confronted with the realities and complexities of the teaching task (Rushton, 2000).

The relationships between perceived support and efficacy increases are consistent with our hypothesis and with Burley et al. (1991) and Hall et al. (1992) who found that confident new teachers gave higher ratings to the adequacy of support they had received than those who ended their year with a shakier sense of their own competence. In the current study, the SES of the students in the class was related to teacher perceptions of support. Teachers in higher SES classrooms felt more supported and found their teaching assignment less difficult than teachers in lower SES classrooms. Clearly many teachers today work in lower SES classrooms. A challenge for teacher education is to prepare novice teachers to seek and create support for themselves in the early years of teaching. In addition, schools that serve low-income students need to make special efforts to support new teachers. One idea comes from Mulholland and Wallace (2001), who found that efficacy for using manipulatives in science lessons could be enhanced by letting new teachers work with half of the class at a time and by allowing the teachers to simply the managerial demands of activity-based lessons by including more teacher-led activities.

In terms of the assessment of efficacy, Bandura’s instrument correlated with the other two measures of efficacy and with the confidence measure at all three phases of the study, thus providing criterion-related concurrent validity evidence for the less frequently used Bandura measure. Because we lacked a sufficient number of participants to conduct a factor analysis on the Bandura measure, it is difficult to describe what exactly was being measured. We did try to create a subscale from Bandura’s nine instructional self-efficacy items (alpha = .84) and found that this subscale also correlated positively with GTE, PTE, and OSU. A larger sample will be needed to further examine
these instruments. But based on these preliminary findings, the PTE and Bandura's measure both provide useful measures of efficacy. The PTE is brief and has a long history of research. The Bandura measure is more specific to the many tasks of teaching. The positive correlation between perceptions of mastery (satisfaction with performance) and increases in efficacy as assessed by Bandura's measure provide some evidence that these novice teachers did find perceived mastery to be a source of efficacy. It is puzzling that PTE increases were not related to perceived mastery. Perhaps Bandura's measure better reflects self-efficacy for teaching, in that changes in this measure are related to mastery, as suggested by social cognitive theory.

Since collecting these data, the first author has worked with Tschannen–Moran to develop a measure of efficacy that is more specific to the tasks of teaching and that yields interpretable factors, called the 'Teachers' Sense of Efficacy Scale or TSES (Tschannen-Moran & Woolfolk Hoy, 2001). We believe that the TSES is superior to previous measures of teacher efficacy in that it has a unified and stable factor structure and assesses a broad range of 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. But even in its nascent form, the TSES opens new possibilities for research.

The results of this study are limited by the sample size and the fact that participants from only one preparation program were included. The results need to be replicated with larger samples and other programs. A challenge for future research is to identify characteristics of the schools that might affect the development of beginning teachers' beliefs and to map relationships between specific school characteristics and teachers' sense of efficacy. More work could be done with novice teachers to understand how their successes and disappointments in their classrooms interact with the supports available in their school to produce enduring efficacy beliefs (Hoy & Woolfolk, 1990; Pajares, 1992). And what kind of support is most helpful in the early years of teaching? Many schools have initiated mentoring programs for teachers in their induction year. Does mentoring provide the kind of support that protects and builds efficacy? What features of mentoring have the greatest impact on efficacy beliefs? 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? How can principals provide support? These and many other questions await investigation.

References


