Measuring the Teaching Self-Efficacy of Science, Technology, Engineering, and Math Graduate Teaching Assistants

Sue Ellen DeChenne and Larry Enochs

Oregon State University

Presented at American Educational Research Conference

Denver, CO

April 2010

Abstract

An instrument to measure the teaching self-efficacy of science, technology, engineering, and mathematics (STEM) GTAs is adapted from a general college teaching instrument (Prieto Navarro, 2005) for the specific teaching environment of the STEM GTAs. The construct and content validity and reliability of the final instrument are indicated. The final instrument has two subscale factors, instructional strategies and positive learning environment but can also be used as an overall measure of STEM GTA teaching self-efficacy. Correlations to GTA training and teaching experience are demonstrated and further research and training implications are discussed.

Introduction

Graduate teaching assistants in science, technology, engineering, and mathematics (STEM) disciplines have a large influence on the teaching of undergraduates. These GTAs often have more contact hours with the students than the professors, especially in large introductory undergraduate courses where the GTAs are usually responsible for teaching laboratory or recitation sections (Fagen & Wells, 2004; Golde & Dore, 2001). Many of the first experiences STEM undergraduates have in college are closely associated with their GTAs. It is important for the university, department, and students that these GTAs are competent instructors (Svinicki, 1995-96). It is also important for the GTA that they are competent instructors. They are paying for their education with the GTA experience and many GTAs are interested in an academic faculty career. They will be the future professors teaching the next generation of undergraduate students. Even if the GTA is not interested in an academic career, the presentation and interpersonal skills learned in teaching can benefit other careers. Therefore it is important that GTAs become good instructors.

Although it is important for GTAs to become good instructors, many GTAs receive no training in teaching (DeChenne et al., 2009; Golde & Dore, 2001; Meyers, Lansu, Hundal, Lekkos, & Prieto, 2007; Piccinin & Fairweather, 1996-97; Rushin et al., 1997). Even GTAs that receive teaching training often indicate it is of little or no help in improving their teaching (Jones, 1993). They are told that teaching is not important, and are not supported in their efforts to improve their teaching (Luft, Kurdziel, Roehrig, & Turner, 2004). The GTA training provided does little to help the GTAs teach. Although there are individual programs that show promising results (e.g. Davis & Kring, 2001) much GTA training is ineffectual (Luft et al., 2004; Shannon, Twale, & Moore, 1998). A good deal of the literature about GTA training is descriptive with little measure of the effectiveness of the training (DeChenne, 2009). Reliable and valid measures of constructs related to improvement are needed. One possible construct to measure is GTA teaching self-efficacy. Teaching self-efficacy refers to a teachers' belief that they will be able to effectively teach a given population of students a specific subject. In the K-12 environment teaching self-efficacy has been shown to be valuable predictor for student achievement, teacher retention, and persistence in the face of teaching difficulties (for a review see Tschannen-Moran, Hoy, & Hoy, 1998). Using teaching self-efficacy could also provide a measure of STEM GTAs' teaching effectiveness. However, before this can be accomplished a valid and reliable measure of STEM GTA teaching self-efficacy is needed.

Literature Review

GTA Training

Studies of GTA training in the literature show mixed results in improving the teaching of GTAs. Many individual programs report success in improving the teaching of the GTAs or demonstrate that the GTAs have learned the material presented in the class (e.g. Carroll, 1977; Daniels, 1970; Davis & Kring, 2001; Hadre & Chen, 2005a, 2005b; Tubb, 1974). Bray and Howard (1980) demonstrate that differences in how GTAs are trained affect their self-reports of teaching, student ratings of teaching, and expert observations of teaching. However, studies that look at GTA training across multiple programs find that such training is usually not effective in changing the teaching of the GTAs (Jones, 1993; Luft et al., 2004). Shannon, Twale and Moore (1998) find that GTA training across a university has no significant effect on student evaluations of teaching. The only type of training that provides a significant effect is an undergraduate degree in education. This literature indicates that the quality of GTA training is highly variable among and across universities. While there are individual programs that report successful training of GTAs, many GTAs are not receiving any training (DeChenne et al., 2009; Golde & Dore, 2001; Rushin et al., 1997; Shannon et al., 1998) or if they are in GTA training, that training by implication from studies across universities, is poor (Jones, 1993; Luft et al., 2004; Shannon et al., 1998).

When GTA training is studied across multiple training programs, it is either measured as a dichotomous variable (GTA training or not) (e.g. Prieto & Altmaier, 1994) or as the amount of time in GTA training (e g. Shannon et al., 1998). This type of measure of GTA training yields minimal information regarding the quality of the GTA training program. The actual impact that GTA training has on a GTAs' teaching is important and having a narrow understanding limits

Table 1
Exploratory Factor Analysis of GTA-TSES subscales with all GTAs

		Factor Loadings[1]	
How confident am I in my ability too[2]…	Variable #	Factor 1: Instructional Strategies	Factor 2: Positive Learning Environment
Appropriately grade my students' exams/assignments?	1	.77	
Spend the time necessary to plan my classes?	2	.74	
Select appropriate materials for class activities?	3	.69	
Evaluate accurately my students" academic capabilities?	4	.66	
Prepare the teaching materials I will use?	5	.64	
Stay current in my knowledge of the subject I am teaching?	6	.61	
Provide my students with detailed feedback about their academic progress?	7	.61	
Specify the learning goals that I expect my students to attain?	8	.59	
Think of my students as active learners, which is to say knowledge builders rather than information receivers?	9		.79
Promote a positive attitude towards learning in my students?	10		.71
Ensure that my students consider themselves capable of learning the material in the course?	11		.67
Promote my students' confidence in themselves?	12		.66
Create a positive classroom climate for learning?	13		.66
Encourage the students to interact with each other?	14		.64
Let students take initiative for their own learning?	15		.64
Eigenvalue		4.13	3.95
Percent (%) of total variance explained		27.50	26.31
Cumulative percent (%) of variance		27.50	53.81
Factor Mean[2]		4.12	3.98
Cronbach α		.87	.86

[1] Principal components factor analysis with Varimax rotation.
[2]Items coded on a 5 point scale of 1=not at all confidence to 5=very confident.

the researchers' conclusions from the data. Rather than just measuring the presence/absence or time in GTA training, it is important to evaluate the learning the GTAs encountered. It should be possible to get a more complete understanding of how well the GTAs learned about teaching using survey items (Table 1) initially developed to determine what GTAs want to learn in their training (DeChenne et al., 2009). This should provide a better measure of the quality of the overall GTA teaching training than currently exists.

Teaching Self-Efficacy

The development of teaching skills is an important part of GTA training. Social cognitive theory (Bandura, 1986, 1997) offers a causal structure which provides for both the development of skills and the regulation of actions. Self-efficacy is a central component in this theoretical framework. Research has demonstrated that when training for a specific skill, high self-efficacy is positively correlated with performance (Bandura, 1997; Gist, Schwoerer, & Rosen, 1989; Pajares, 1996a). According to Bandura (1997)

> Perceived self-efficacy refers to beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments….[self-efficacy] beliefs influence the courses of action people choose to pursue, how much effort they put forth in given endeavors, how long they will persevere in the face of obstacles and failures, their resilience to adversity, whether their thought patterns are self-hindering or self-aiding, how much stress and depression they experience in coping with taxing environmental demands, and the level of accomplishments they realize (pg. 3).

Self-efficacy, as conceived by Bandura, is the theoretical base for teaching self-efficacy and the development of a STEM GTA Teaching Self-Efficacy Scale (GTA-TSES) used in this study.

Teaching self-efficacy research has followed two strands, the first resulting from the powerful predictive results of just two questions on an early study of teacher characteristics (RAND study) and student learning (Berman & McLaughlin, 1977) and the second strand based in Bandura's (1986) social cognitive theory. The two items from the RAND study proved so powerful in predicting student performance, teacher change, and continued use of methods and materials from federally funded projects that many different multiple item instruments were developed to capture teacher efficacy (see Tschannen-Moran et al., 1998, for a review). One of the most widely used is the Gibson and Dembo (1984) instrument which was based on the RAND study items but utilized the framework of self-efficacy from social cognitive theory (Bandura, 1986). Using instruments such as these, research with elementary, middle, and high school teachers demonstrated that teacher's self-efficacy beliefs impacted many student outcomes and teacher behaviors. In a review of the literature on teacher self-efficacy, Tschannen-Moran et al. (1998) indicated teachers' self-efficacy beliefs were related to student outcomes such as achievement, motivation, and the students' sense of efficacy. They also indicated teaching self-efficacy was related to teacher classroom behaviors, the goals set, persistence with students, enthusiasm for and commitment to teaching. Teachers with high teaching self-efficacy performed better and their students benefited.

At the post-secondary level, teaching self-efficacy has been studied in GTAs, mainly in psychology departments, although there are two studies with GTAs across the university. However, we find no published studies using teaching self-efficacy as a measure specifically for STEM GTA training. Teaching self-efficacy in GTAs has complex interactions with teaching assistant training, previous teaching experience, and supervision (Heppner, 1994; Meyers et al., 2007; Prieto & Altmaier, 1994; Prieto & Meyers, 1999; Prieto, Yamokoski, & Meyers, 2007; Tollerud, 1990).

In studies with psychology students, variable effects of GTA training, experience, and supervision are found. Tollerud (1990) first developed an instrument to measure the teaching self-efficacy of counseling psychology GTAs which was then later adapted for general use with GTAs (Prieto & Altmaier, 1994). Tollerud indicates that in a population of advanced counseling psychology GTAs and recent graduates, having no teaching experience decreases their teaching self-efficacy. Prior professional teaching experience does not significantly impact teaching self-efficacy but the amount of time spent teaching as a GTA does. GTA training has no significant effect on teaching self-efficacy. In contrast, Prieto and Meyers (1999) find a significant effect for GTA training but not supervision on teaching self-efficacy. In a further study of psychology GTA training, training and teaching experience are positively associated with teaching self-efficacy (Meyers et al., 2007).

Level of responsibility for teaching can have an effect on GTAs teaching self-efficacy. Prieto and Altmaier (1994) find there are no significant differences in teaching self-efficacy by amount of teaching responsibility (primary, assistant, or general instructor). However in two other studies, GTAs in non-instructive roles (e.g. graders) have significantly lower teaching self-efficacy scores than GTAs who have teaching roles (either assistant or primary instructors) (Prieto & Meyers, 1999); and there are significant differences between GTAs who are graders or assistants and those who have the primary responsibility for the classroom (Prieto et al., 2007).

In studies with GTAs from across institutions conflicting results are also evident. In one study, there are small significant correlations (Cohen, 1988) between teaching self-efficacy and prior training ($r=.22$) or previous experience ($r=.25$). However in a regression analysis, while teaching experience is significant, it only explains 5% of the variance. Prior training is not significant at the .05 level and when forced into the model explains 2% of the variance (Prieto & Altmaier, 1994). In a more detailed study including the effects of supervision, GTA training and supervision have variable effects depending on the teaching role of the GTA (Prieto et al., 2007). In the assistant/grader group there is no effect on teaching self-efficacy for training or supervision but prior teaching experience accounts for a significant amount of the variance. In the primary group prior teaching experience does not account for a significant amount of the variance in the teaching self-efficacy scores but there is a significant interaction effect for supervision and training. A further breakdown of the supervision and training interactions with the primary group shows supervision has no effect on teaching self-efficacy when there is training but has a positive effect on teaching self-efficacy when there is no training. Receiving both training and supervision produces lower teaching self-efficacy than just supervision alone. Training has no effect on the teaching self-efficacy of students who are not receiving supervision.

Situational specificity of self-efficacy

Bandura (1997) proposed that because self-efficacy beliefs were explicitly self-referent in nature and directed toward perceived abilities about given specific tasks, they were powerful predictors of behavior. Further he stated that self-efficacy refers to organizing and executing courses of action required to successfully accomplishing a specific teaching task in a particular context. Many measures of self-efficacy address specific tasks, yet fall short of providing the particular context (Enochs & Riggs, 1990; R. K. Henson, 2002; Pajares, 1996a).

In the development of teaching self-efficacy instruments the situational specificity of different teaching contexts and tasks is important (Bandura, 1997; R. K. Henson, 2002; Pajares, 1996a). There is a balance between the general and the specific. Overly specific and the instrument is not measuring teaching self-efficacy, just a specific task in a specific context. A measure of a genetics GTA’s belief in their ability to teach how to load DNA samples into an agarose gel for electophoresis to junior level biochemistry students at a research university is too specific. However, to broad and the instrument may simply be measuring general personality traits instead of the self-efficacy specific to the task (Pajares, 1996b). Since self-efficacy refers to a persons’ beliefs about their abilities to accomplish a behavior or task in a specific context it is important to design the instrument to the context in which the person in performing the task (Bandura, 1997).

College teaching self-efficacy has been measured mostly in psychology students, but teaching styles are different among the various disciplines. It has been recognized that teaching in STEM is fundamentally different from the other disciplines and that this difference should be recognized in the roles of GTAs (Golde & Dore, 2004; Lindblom-Ylanne, Trigwell, Nevgi, & Ashwin, 2006; Torvi, 1994; Verleger & Velasquez, 2007). STEM GTAs are rarely responsible for a course (DeChenne et al., 2009; Torvi, 1994) but instead teach laboratory and recitation sections, and so usually act as a conduit between the students and the course professor. GTAs need to understand complex grading rubrics and have skills that allow them to facilitate questions without giving the student the answer. STEM students often work independently or in small groups on complex projects that can span a term or more of coursework (Moore & Diefes-Dux, 2004; Pomalaza-Raez & Groff, 2003; Taylor, Heer, & Fiez, 2003). GTAs must understand these long-term projects, how to facilitate learning, help students at different points of scholarship, and with frustrating problems. All of these activities require the STEM GTA to have excellent interpersonal relationship skills.

Given that the STEM GTA teaching context is different from other university teaching contexts, measuring teaching self-efficacy in STEM GTAs requires a context specific instrument. Additionally, when comparing the GTA training across more than one program it is important to measure the GTAs' perception of the quality of that training rather than simply the number of hours of training or whether or not training occurred. According to social cognitive theory, the two measures should be correlated. The development and validation of these instruments is the focus of this paper. The research objectives are as follows:

1. Measure GTA perceptions of their learning in GTA training.
2. Adapt the College Teaching Self-Efficacy Scale (CTSES) for the STEM GTA context.
3. Determine the reliability of the GTA Teaching Self-Efficacy Survey (GTA-TSES).
4. Determine the content and construct validity of the GTA-TSES.

Methods

Instrument Modification

There were two college teaching self-efficacy instruments. The instrument used in most of the GTA teaching self-efficacy studies was the Self-Efficacy Toward Teaching Inventory – Adapted (SETI-A) (Prieto & Altmaier, 1994), which was adapted for general GTA use from a teaching self-efficacy instrument that was specific for counseling psychology educators (Tollerud, 1990). A second college teaching self-efficacy instrument, CTSES, was more recently developed (Prieto Navarro, 2005). After discussing what types of items should be included in a STEM GTA teaching self-efficacy scale; a team, including two science educators and two engineering education faculty, reviewed the items on the CTSES and the SETI-A. The CTSES was chosen since it required less extensive modification. The team collaborated to adapt the CTSES to the STEM GTA context.

As part of a larger study of STEM GTA teaching self-efficacy, the CTSES needed to be streamlined; items specific to STEM GTA teaching added or modified from the general college instructor context; and items not usually part of a STEM GTA duties removed. The CTSES was long (44 items) and contained two six point scales, one for self-efficacy and one measuring actual instructor action for each item. STEM GTAs were rarely involved in course design or were the primary instructors responsible for a course, therefore items related to overall course design and planning were removed (seven items). The CTSES contained items on reflective practice, many of which required teaching the same course repeatedly. Many GTAs, especially in engineering, did not teach the same course repeatedly so these items were removed (five items). Items that were unclear to the researchers or included technical pedagogical language were also removed (three items). There were two pairs of redundant items; one from each pair was removed. Four items were rewritten to be more specific to the STEM GTA context. Given the large amount of group work in STEM laboratory classes, one item relating to student interaction was added. Finally only the self-efficacy scale was retained but changed to a five point scale because of the limitations of data collection (the scantrons included A to E). The content validity of the items was reviewed by two additional social science faculty with knowledge of both social cognitive theory and instrument design. They were asked to evaluate whether each item represented an aspect of the GTA's teaching self-efficacy, to comment on clarity, and suggest revisions or additions. The instrument as administered contained 28 items, measured on a five point scale of A (no confidence) to E (complete confidence).

Participants

The GTA-TSES was administered to GTAs in various STEM departments at six universities; three in the Pacific Northwest, two in the Southwest and one in the Midwest. Five universities had a Carnegie basic classification of RU/VH (Research Universities with Very High research activity) and one was a DRU (Doctoral/Research University). Engineering and technology GTAs taught across the engineering disciplines including; aerospace, biological, biomedical, chemical, civil, computer, construction, electrical, environmental, industrial, manufacturing, mechanical, and petroleum. Science GTAs taught in biology, chemistry, geosciences, microbiology, molecular biology, and physics. Also included in the sample were GTAs who taught in mathematics.

Administration

The GTAs were administered the GTA-TSES once, near the end of the semester or quarter. Data was collected from Fall 2009 through Fall 2010. One of two administration techniques was used depending on location. The survey was distributed to the GTAs through the department mail system, collected in a sealed container in the departmental office, and returned to the researcher through the US mail (or collected directly by a researcher). Alternatively, the survey was administered during a GTA training class and collected by one of the researchers at that time. All protocols were approved by the institutional review board at the originating university.

Data Analysis

Study 1. The GTA-TSES was administered to STEM GTAs at five institutions outside the originating university as previously described. In addition to the items in the GTA-TSES, three additional questions were asked; two demographic questions (university and department affiliation) and a question indicating the GTA's primary role as; laboratory, recitation, lecturer, course instructor, or grader. There were 76 participants, including 48 engineering GTAs and 28 science GTAs. Twenty percent of the GTAs described themselves as graders, the rest indicated classroom instructional roles, with laboratory instructor the most common (56%). Using SPSS version 16, the 28 items on the self-efficacy instrument were analyzed using principle components factor analysis with varimax rotation.

Study 2. The survey instrument was administered to STEM GTAs at the originating university as previously described. The survey instrument contained the following measures: GTA-TSES, GTA training, GTA teaching experience, and additional demographic questions included gender, nationality, and career interest. The GTA-TSES items were analyzed using principle components factor analysis with varimax rotation, and a scree test.

There were 177 participants; 61% were engineering and 39% science or mathematics GTAs. Twenty six percent of the GTAs described their primary role as grading, the rest indicated classroom instructional roles, with laboratory instructor (36%) the most common. Twenty-seven percent of the GTAs were female and 64% were interested in college/university teaching as a career. The sample was split almost evenly between international teaching assistants (ITAs - 47%) and US teaching assistants (USTAs - 53%). Thirteen percent of the sample had no GTA training of any kind; university, department, or college coursework.

GTA training was measured two ways; (a) number of hours spent in teaching training measured in university-wide and department GTA training, and university coursework in teaching, and (b) by the GTA's overall perception of their GTA training. This was measured by rating the both overall GTA training and how well the GTAs learned different items directly related to teaching. These items were generated from a prior study of GTA training (DeChenne et al., 2009). Confirmatory factor analysis (CFA) of the items about GTA training was used to examine whether the variables measuring this latent factor provided a good fit and demonstrated construct validity. EQS 6.1 software and Satorra-Bentler robust estimation to correct for multivariate non-normality was used for the CFA analysis (Byrne, 1994). Robust corrected comparative fit index (CFI), non-normed fit index (NNFI), and root mean square error of approximation (RMSEA) were used to assess model fit. CFI and NNFI values ≥ 0.90 and RMSEA values ≤ 0.08 suggested acceptable fit (Browne & Cudeck, 1993).

Teaching experience was measured by the number of quarters taught and the amount of responsibility for the course the GTAs had in their teaching (from none to complete responsibility). Additionally, two items that asked the GTAs to rate their own experience were used. One item asked the GTAs to compare themselves to other GTAs in the department and the other asked them to rate their own experience (from beginner to expert).

Validity and Reliability. The GTAs from both studies were combined and the construct validity of the GTA-TSES was determined and a second-order factor analysis of the data done. Confirmatory factor analysis (CFA) was used to examine whether the variables measuring the latent factors provided a good fit and demonstrated construct validity. The possibility of a second order factor structure was examined using both second order CFA and principle components factor analysis with varimax rotation forcing one factor.

Internal consistency of multiple-item indices measuring these concepts was examined with Cronbach alpha reliability coefficients. An alpha coefficient ≥ 0.65 indicated that items were measuring the same concept and justified combining items into a single index (Cortina, 1993). Using the GTAs from Study 2, Pearson's Product Moment Correlations (r) between the GTA-TSES and teaching experience or GTA training measures were determined. The differences in teaching self-efficacy by nationality, teaching experience, gender, teaching responsibility, GTA training, and career plans were compared using t-tests and effect sizes were determined using point biserial correlations.

Results

Study 1

The 28 items were subjected to exploratory factor analysis as described above. Five factors that explained 69% of the variance in the GTAs scores were found. The rotation failed to converge in 25 iterations so the unrotated factors were examined. The first factor had an eigenvalue of 13.312 and an explained 48% of the variance. Costello and Osborne (2005) suggested, for small sample sizes in EFA, to select strong factor loadings when determining what items to retain in a factor; therefore items with factor loadings below .60 were removed (two items). Additionally, upon reading the items again one more item was removed due to poorly worded language. There were 25 items available for further analysis.

Study 2

The 25 items were subjected to exploratory factor analysis as described above. Five factors that explained 62% of the variance in the GTAs scores were found. A scree test (Cattell, 1966) suggested that two or three factors could be found in the data. The two factor analysis found items relating to a positive learning environment loading on one factor with items relating to instructional strategies loading on the other. The three factor solution split up the instructional strategies factor into two with some of the items from the positive learning environment also loading on the third factor. Given that the two factors were more defined and clearly represented the tasks of teaching for STEM GTAs, the two factor structure was used. Additionally, the two factors found were similar to two of the factors determined for pre-service and in-service teachers (Tschannen-Moran & Hoy, 2001). They found three factors, efficacy for instructional strategies, student engagement, and classroom management. We did not include any items relating to classroom management in this instrument since GTAs are teaching adults not children, but the other two factors were similar.

The initial two factor solution explained 48% of the variance in the GTAs scores. To strengthen the factors, all of the items that cross-loaded between the two factors and the items with the lowest factor loadings within each factor were removed (Costello & Osborne, 2005). This left 15 items. Another exploratory factor analysis with those 15 items gave two clean factors that explained 53% of the variance. The factors were labeled self-efficacy for instructional strategies (*is*=8 items), and positive learning environment (*ple*=7 items).

Validity and Reliability of the GTA-TSES

The data from the two studies of GTAs was combined to explore the validity, reliability, and possible 2nd order factor structure of the teaching self-efficacy instrument. An exploratory factor analysis of the combined data gave two clean factors that explained 54% of the variance in GTAs scores (Table 1). All of the variables loaded between .59 to .77 for instructional strategies and .64 to .79 for positive learning environment. Both factors were highly reliable (*is* α=.87, *ple* α=.86); all the variables met the criterion of item total item correlations being greater than .40 and deletion of any item did not improve the reliability. A CFA for each factor demonstrated good model fit, indicating good construct validity for each factor (*is*, NNFI=.916, CFI=.940, RMSEA=.064; *ple*, NNFI=.949, CFI=.966, RMSEA=.059). All variables loaded between .61 to .78 for instructional strategies and .62 to .74 for positive learning environment and were significant at the p<.05 level. Means for each factor were high (*is*=4.13, *ple*=3.95), indicating that in each factor the GTAs were confident in their ability to carry out these teaching duties and responsibilities.

Given the strong correlation (Cohen, 1988) between the instructional strategies and positive learning environment with all the GTA data (r=.69), a exploratory factor analysis forcing one component was conducted to determine if these factors measured a single concept of GTA teaching self-efficacy. All items loaded on one factor with factor loadings ranging from .61 to .76, explaining 45% of the variance in the GTAs responses. To confirm the measure of GTA teaching self-efficacy, a 2nd order CFA was performed on the items (Figure 1). There was a good fit of the second order structure (NNFI=.910, CFI=.924, RMSEA=.052). All the

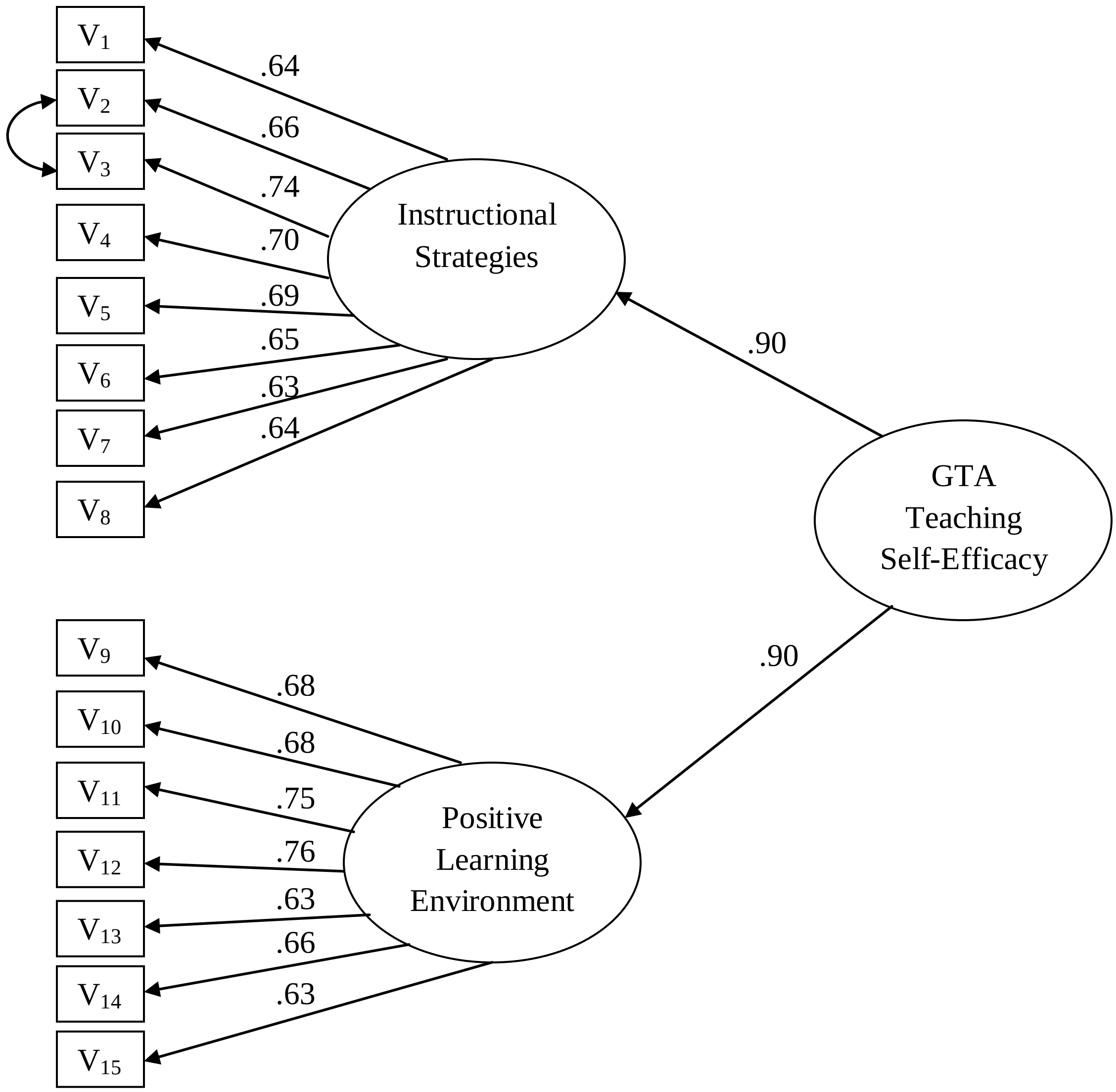

Figure 1: 2nd Order Confirmatory Factor Analysis of GTA Teaching Self-Efficacy

See Table 2 for variables corresponding to codes (e.g., V_1). All factor loadings indicated are significant at *p*<.05. Model fit indices are NNFI=.910, CFI=.924, & RMSEA=.052. To achieve these fit indices the errors between V_2 and V_3 were allowed to covary.

variables loaded between .63 to .74 for instructional strategies, .63 to .76 for positive learning environment, and each factor loaded on the second order GTA teaching self-efficacy at .90. All factor loadings were significant at the *p*<.05 level. The reliability of the single factor structure was .91 with a mean of 4.06. All the variables met the criterion of item total item correlations being greater than .40 and deletion of any item did not improve the reliability. These results suggested that this instrument could be used to measure the underlying concept of GTA teaching self-efficacy and that a total teaching self-efficacy as well as subscale self-efficacies could be measured.

Table 2

Confirmatory Factor Analysis of GTA Perception of Training with Study 2 GTAs

Of the following teaching topics and skills, please rate how well you have learned these in GTA training?[1]	Factor Loadings[2]
Facilitating group discussions	.86
Learning styles	.83
Motivating students	.82
Teaching students with different skill/knowledge	.82
Managing disruptive students	.81
Interacting professionally one-on-one with your students	.81
Teaching styles	.81
Teaching culturally diverse students	.80
Power/authority relationships in the classroom	.77
Assisting distressed students	.75
Communicating with course lead instructor	.71
Presenting material to large groups of students	.70
Harassment	.67
Grading	.60
Developing quizzes/exams	.60
Overall Questions on GTA training[3]	
Overall, how effective has the TA training you have received been in preparing you to work with students?[4]	.69
Overall, how effective has the TA training you have received been in preparing you to teach? [4]	.66
Mean	3.15
Cronbach α	.96

[1]Items coded on a 5 point scale of 1=never learned to 5=learned very well.

[2]All factor loadings are significant at $p<.05$. Model fit indices are NNFI=.925, CFI=.935, & RMSEA=.083.

[3]Items coded on a 5 point scale of 1=not effective to 5=very effective.

[4]Errors allowed to covary to achieve fit indices.

Validity and Reliability of GTA Perception of Training Instrument

Using the study 2 GTAs, a CFA demonstrated an acceptable model fit and supported the construct validity of the GTA perception of training factor (NNFI=.925, CFI=.935, RMSEA=.083) (Table 2). All of the variables loaded between .60 to .86 and were significant at $p<.05$. The GTA perception of training factor was highly reliable ($\alpha=.96$); all the variables met the criterion of item total item correlations being greater than .40 and deletion of any item did not improve the reliability. The GTAs indicated they learned various skills in their GTA training moderately well (M=3.15 on a 5 point scale of never learned to learned very well).

Correlational and Comparison Analysis

The overall GTA teaching self-efficacy scale (GTA-TSES) and the instructional strategies and positive learning subscales correlated significantly with several measures of teaching training and teaching experience (Table 3). These measures, with the study 2 GTA sample, indicated significant small to moderate correlations (Cohen, 1988) of the GTA-TSES and both subscales with the GTAs perception of GTA training. There was also a small correlation (Cohen, 1988) of the number of hours reported in university GTA training with the positive learning scale and amount of university coursework in teaching with the GTA-TSES. There were significant moderate correlations (Cohen, 1988) of the GTA-TSES and both subscales with measures of teaching experience that asked the GTAs to rate themselves compared to their peers and on a scale from beginner to expert. There was a small significant correlation (Cohen, 1988) with the instructional strategies subscale and the number of quarters the GTAs taught and the amount of course responsibility.

Table 3
Correlational Analysis with Study 2 GTAs

	Measures	Mean	GTA[2] TSES	A	B	C	D	E	F	G	H	I
Self-Efficacy												
A	Instructional Strategies[1]	4.16	.91**									
B	Positive Learning Environment[1]	4.00	.91 **	.65**								
Training												
C	Perception[1,3]	3.16	.33**	.32**	.25**							
D	University[4]	2.21	.13	.08	.15*	.23**						
E	Department[4]	7.67	.07	.06	-.01	.11	.23**					
F	Course Credit[4]	9.01	.15*	.14	.14	.15	.06	-.03				
Experience												
G	Compare[1,5]	3.34	.35**	.38**	.24**	.04	.10	.07	.10			
H	Rating[1,6]	2.93	.44**	.43**	.35**	.22**	.12	.12	.17*	.67**		
I	Quarters[7]	5.22	.12	.16*	.03	.00	-.05	.23*	.03	.46**	.46**	
	Responsible[1,8]	3.87	.12	.17*	.10	.06	.03	.03	-.03	.19*	.25**	.17*

**p*<.05 (2 tailed), ***p*<.01 (2-tailed)
[1]All scales were rated on a scale of 1 to 5, with 5 being the best in each scale.
[2]GTA-Teaching Self-Efficacy Scale
[3]GTA ratings of how well they learned teaching skills, see Table 3.3.1
[4]Hours of training
[5]Item asking; Compared to other GTAs how much teaching experience do you have?
[6]Item asking; Rate your own teaching experience?
[7]Numbers of quarters as a GTA
[8]Item-For the course in which you had the most responsibility as a GTA, how much responsibility did you have?

The data collected with Study 2 GTAs indicated no differences in mean scores on the GTA-TSES or the *is* and *ple* subscales. There was no significant difference by gender (GTA-TSES males=4.10, females=4.05; t=.51, p=.61, r_{pb}=.04), career goals (GTA-TSES academic=4.12, other=4.02; t=1.46, p=.25, r_{pb}=.09), GTA training (no training=4.08, training=4.09; t=.02, p=.99, r_{pb}=.001) or nationality (GTA-TSES ITA=4.16, USTA=4.02; t=1.63, p=.11, r_{pb}=.12). When the GTAs from both studies were combined there were no significant differences in GTA-TSES or the *is* and *ple* subscales by classroom role (GTA-TSES graders=4.13, classroom instruction=4.06; t=.86, p=.39, r_{pb}=.06) or college of instruction (GTA-TSES science/math=4.00, engineering=4.09; t=1.03, p=.28, r_{pb}=.07).

Discussion

The purpose of this study was to develop an instrument to measure the teaching self-efficacy of STEM GTAs and to explore some of the relationships between STEM GTA teaching self-efficacy and GTA training and teaching experience. Essential to this process was to work toward the establishment of the reliability and validity of the measure. Exploration of GTA training also necessitated developing an instrument that would give a better picture of the GTA training received by STEM GTAs than a simple record of training or time in training. Assertions related to instrument validity and correlations must be viewed as sample dependent.

Instrument Modification and Development

The teaching self-efficacy instrument developed during this process has two subscales; instructional strategies and positive learning environment. These subscales can be used individually or they can be combined into a measure of the teaching self-efficacy of the STEM GTAs. This structure is not unlike the Ohio State Teacher Efficacy Scale (Tschannen-Moran & Hoy, 2001). They found when they developed a new teaching self-efficacy instrument using their theory of teaching self-efficacy (Tschannen-Moran et al., 1998) that there were three factors; student engagement, instructional strategies, and classroom management, which could be used to measure an overall teaching self-efficacy. In the GTA-TSES the two factor structure provides more flexibility in the use of the instrument. It provides a global score of teaching self-efficacy. However, if the self-efficacy of the STEM GTAs relating to their classroom instruction or their ability to create a positive learning environment is needed then this instrument offers that option. When this instrument is used to evaluate GTA training, the subscales could be very useful in determining where changes have occurred in the GTAs' teaching self-efficacy.

Consistent with Prieto and Altmaier (1994) teaching self-efficacy in STEM GTAs also does not vary in this study by their gender, career plans, instructional role, or college of teaching appointment. In two other populations the self-efficacy of GTAs varied by instructional role (Prieto & Meyers, 1999; Prieto et al., 2007). Unlike these studies, STEM GTAs are rarely the course instructor (5% in this study), which is the group reported to have the different teaching self-efficacy in one of these studies (Prieto et al., 2007) and make up almost half of the combined assistant/full responsibility group in the other study (Prieto & Meyers, 1999). Like in Tollerud (1990), there was no difference in the teaching self-efficacy between GTAs who had attended GTA training and those who had not, although other studies have shown an effect on teaching self-efficacy for GTAs with GTA training (Meyers et al., 2007; Prieto & Altmaier, 1994; Prieto & Meyers, 1999; Prieto et al., 2007). Additionally, the means of the GTA-TSES with this sample of STEM-GTAs is higher (4.0 to 4.2) than those reported by Prieto and Altmaier (3.2 to 3.4) but neither varies much. These means indicate that the STEM GTAs have a relatively high teaching self-efficacy. ITAs have a significant, slightly higher instructional strategies self-efficacy than the USTAs. But the overall teaching self-efficacy of ITAs and USTAs was similar.

The GTA perception of training factor also had good reliability and validity in this study. It provides a new measure of GTA training beyond the simpler measures of presence/absence of GTA training or reports in amount of time of GTA training. This factor is an evaluation of the teaching training received by the GTAs around certain teaching skills. As such the items may reflect our biases towards important teaching skills and certainly cannot cover all the skills required for effective teaching. However, the items were generated by a broad group of individuals (DeChenne et al., 2009) and provide a way for the GTAs to assess their training. This measure gives researchers another way to estimate GTA training in quantitative studies. If the training received by the GTAs was high quality and the GTAs learned about teaching then this measure should be high. A large amount of time in teaching training is not a guarantee of learning. Not surprisingly then, the GTA perception of training does not correlate significantly with all the time measures of GTA, although it does show a small correlation with university-wide training. It also has a small correlation with the GTAs' rating of their own teaching experience (from beginner to expert).

One of the possible limitations of this study is sample size. Costello and Osborne (2005) demonstrate in exploratory factor analysis that a ratio of at least ten participants to each item in the instrument provides an average of less than one (0.70) item misclassified on the wrong factor. In Study 2 there is a ratio of seven participants to each item. However, by Costello and Osborne's own categorization, the individual factors (Table 2) in the GTA-TSES are solid; "a factor with…5 or more strongly loaded items (.50) are desirable and indicate a solid factor" (pg. 5). Additionally, the CFA of the combined GTAs indicates a solid first and second order factor structure. Although the strong first order factor structure could be expected since 70% of the sample is the same as used in the exploratory factor analysis, the second order factor analysis indicates that these two more specific self-efficacy scales collapse into broader teaching self-efficacy factor. This could be predicted from social cognitive theory which indicates that teacher efficacy should be task specific (Bandura, 1997; Pajares, 1996a). As indicated by Henson (2002), it is not surprising to find a higher order factor structure within this instrument. However, further research with this instrument should use CFA to further validate the two factor structure of the instrument.

Related Relationships

A correlational analysis suggest that the GTA-TSES and the *is* and *ple* factors are related to measures that theory and prior research indicate should correlate. There is a high correlation between the two subscale factors in this instrument. Since these two factors measure related activities in the classroom, instructional strategies and positive learning environment they should be correlated. Prior research (Meyers et al., 2007; Prieto & Altmaier, 1994; Prieto et al., 2007; Tollerud, 1990) shows a positive effect for GTA teaching experience except in one study where for GTAs with primary responsibility it did not (Prieto et al., 2007). The *is* subscale correlates with all measures of teaching experience. However the *ple* factor and the GTA-TSES only correlate with measures of teaching experience that are GTA self-reports and not with the number of quarters or amount of responsibility. Examining the SETI-A shows that most of the items are more similar to the *is* factor than the *ple* factor which may account for the different results of the *ple* and GTA-TSES with teaching experience. According to social cognitive theory (Bandura, 1997), training in a task should increase the self-efficacy of the person in performing that task. As a measure of the quality of their teaching training, GTA perception of teaching should correlate with the GTAs teaching self-efficacy. It has a moderate correlation with both subscales and the complete instrument. Prior research shows variable results for the effects of GTA training on teaching self-efficacy (Meyers et al., 2007; Prieto & Altmaier, 1994; Prieto & Meyers, 1999; Prieto et al., 2007; Tollerud, 1990). This may be because of the measure of training used. In this study only GTA perception of training correlated with both subscales and the GTA-TSES. Two of the rest of the training variables (measured in hours) had a small correlation with either the *is* subscale or the GTA-TSES.

This research includes the initial development of the GTA-TSES instrument. This instrument needs continued study using a variety of STEM GTAs to determine if the factor structure is stable and reliable with different populations of STEM GTAs. With this population of GTAs there appears to be little difference between groups in their teaching self-efficacy; the discriminatory power of this instrument needs to be further investigated. The initial correlational study shows promising results but, as Henson (2001) articulates, it is time for teaching self-efficacy studies to move beyond correlations. Research into the longitudinal effects of GTA training and teaching experience needs to be done. Using this instrument to track changes in the long-term self-efficacy of engineering GTAs is underway, as is a study investigating the relationships between teaching self-efficacy and faculty and peer interactions. Neither instrument developed to measure GTAs teaching self-efficacy has yet explored the GTAs' interpretation on the items included in the instrument. So further research on this instrument should include qualitative studies to determine the way STEM GTAs interpret the items in the instrument.

References

Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory (1st ed.). Englewood Cliffs, NJ: Prentice-Hall, Inc.

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

Berman, P., & McLaughlin, M. (1977). Federal programs supporting educational change: Vol. 7 Factors affecting implementation and continuation (R-1589/7-HEW). Santa Monica, CA: Rand Corporation.

Bray, J. H., & Howard, G. S. (1980). Methodological considerations in the evaluation of a teacher-training program. Journal of Educational Psychology, 72(1), 62-70.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In Bollen, K. A. & Long, J. S. (Eds.), Testing structural equation models (pp. 136-162). Newbury Park, CA: SAGE Publication, Inc.

Byrne, B. M. (1994). Stuctural equation modeling with EQS and EQS / Windows. Thousand Oaks, CA: Sage.

Carroll, J. G. (1977). Assessing the effectiveness of a training program for the university teaching assistants. Teaching of Psychology, 4, 135 - 138.

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

Chou, C. P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In Hoyle, R. H. (Ed.), Structural equation modeling (pp. 37-55). Thousand Oaks, CA: Sage.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (Second ed.): Lawerence Erlbaum Associates, Publishers.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. . Journal of Applied Psychology, 78, 98-104.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Practical Assesment, Research & Evaluation, 10(7), Available online: http://pareonline.net/pdf/v10n17.pdf.

Daniels, J. W. (1970). Effect of interaction analysis upon teaching assistants and student achievement in introductory college mathematics. Unpublished Dissertation, Indiana University.

Davis, S. F., & Kring, J. P. (2001). A model for training and evaluating graduate teaching assistants. College Student Journal, 35(1), 45 - 51.

DeChenne, S. E. (2009). Model of factors that contribute to the teaching effectiveness of graduate teaching assistants Paper presented at the Association for the Study of Higher Education, Vancouver, BC.

DeChenne, S. E., Anderson, S., Lesseig, K., Li, S., Staus, N., Barthel, C., et al. (2009). Graduate student teaching training: Differences in graduate students' training needs, American Educational Research Association. San Diego, CA.

Enochs, L. G., & Riggs, I. M. (1990). Further development of an elementary science teaching efficacy belief instrument: A preservice elementary scale. School Science and Mathematics, 90(8), 695 - 706.

Fagen, A. P., & Wells, K. M. S. (2004). The 2000 national doctoral program survey. In Wulff, D. H. & Austin, A. E. (Eds.), Paths to the professoriate: Strategies for enriching the preparation of future faculty. San Francisco, CA: Jossey-Bass.

Gibson, S., & Dembo, M. (1984). Teacher efficacy: A construct validation. Journal of Educational Psychology, 76(4), 569-582.

Gist, M., Schwoerer, C., & Rosen, B. (1989). Effects of alternative training methods on self-efficacy and performance in computer software training. Journal of Applied Psychology, 74(6), 884 - 891.

Golde, C. M., & Dore, T. M. (2001). At Cross Purposes: What the experiences of doctoral students reveal about doctoral education, 2007, from http://phd-survey.org/report%20final.pdf

Golde, C. M., & Dore, T. M. (Eds.). (2004). The survey of doctoral education and career preparation: The importance of disciplinary contexts. San Francisco: Jossey-Bass.

Hadre, P. L., & Chen, C.-H. (2005a). A Case Study Analysis of the Role of Instructional Design in the Development of Teaching Expertise. Performance Improvement Quarterly, 18(1), 34 - 58.

Hadre, P. L., & Chen, C.-H. (2005b). Teaching assistants learning, students responding: Process, principles, products and perspectives. Journal of Graduate Student Teaching Assistant Development, 10(1), 25 - 51.

Henson, R. K. (2001). Teacher Self-Efficacy: Substantive Implications and Measurement. Paper presented at the Educational Research Exchange, College Station, Texas.

Henson, R. K. (2002). From adolescent angst to adulthood: Substantive implications and measurement dilemmas in the development of teacher efficacy research. Educational Psychologist, 37(3), 137 - 150.

Heppner, M. J. (1994). An empirical investigation of the effects of a teaching practicum on prospective faculty. Journal of Counseling and Development, 72(May/June), 500-507.

Jones, J. L. (1993). TA training: from the TA's point of view. Innovative Higher Education, 18(2), 147 - 161.

Lindblom-Ylanne, S., Trigwell, K., Nevgi, A., & Ashwin, P. (2006). How approaches to teaching are affected by discipline and teaching context. Studies in Higher Education, 31(3), 285 - 298.

Luft, J. A., Kurdziel, J. P., Roehrig, G. H., & Turner, J. (2004). Growing a garden without water: Graduate teaching assistants in introductory science laboratories at a doctoral/research university. Journal of Research in Science Teaching, 41(3), 211 - 233.

Meyers, S. A., Lansu, M. L., Hundal, J. S., Lekkos, S. K., & Prieto, L. R. (2007). Preparing new psychology instructors to teach undergraduates: developing confidence and competence. Journal of Faculty Development, 21(1), 45 - 54.

Moore, T., & Diefes-Dux, H. (2004, October 20 - 23). Developing model-eliciting activities for undergraduate students based on advanced engineering content. Paper presented at the ASEE/IEEE Fontier in Education Conference, Savannah, GA.

Pajares, F. (1996a). Assessing self-efficacy beliefs and academic outcomes: The case for specificity and correspondence., American Educational Research Association. New York.

Pajares, F. (1996b). Self-efficacy beliefs in academic settings. Review of Educational Research, 66(4), 543-578.

Piccinin, S., & Fairweather, D. (1996-97). A local analysis of TA training needs: Implications for the planning and development of TA training programs. The Journal of Graduate Teaching Assistant Development, 4(1), 23 - 34.

Pomalaza-Raez, C. A., & Groff, B. H. (2003). Retention 101: Where robots go...students follow. Journal of Engineering Education, 92(1), 85 - 90.

Prieto, L. R., & Altmaier, E. M. (1994). The relationship of prior training and previous teaching experience to self-efficacy among graduate teaching assistants. Research in Higher Education, 35(4), 481 - 497.

Prieto, L. R., & Meyers, S. A. (1999). Effects of training and supervision on the self-efficacy of psychology graduate teaching assistants. Teaching of Psychology, 26(4), 264 - 266.

Prieto, L. R., Yamokoski, C. A., & Meyers, S. A. (2007). Teaching Assistant Training and Supervision: An examination of optimal delivery modes and skill emphasis. Journal of Faculty Development, 21(1), 33 - 43.

Prieto Navarro, L. (2005). Las creencias de autoeficacia docente del profesorado universitario. Madrid: Universidad Pontifica Comillas.

Rushin, J. W., De Saix, J., Lumsden, A., Streubel, D. P., Summers, G., & Berson, C. (1997). Graduate teaching assistant training: A basis for improvement of college biology teaching and faculty development? American Biology Teacher, 59(2), 86 - 90.

Shannon, D. M., Twale, D. J., & Moore, M. S. (1998). TA teaching effectiveness: The impact of training and teaching experience. The Journal of Higher Education, 69(4), 440-446.

Svinicki, M. (1995-96). A dozen reasons why we should prepare graduate students to teach. The Journal of Graduate Teaching Assistant Development, 3(1), 5 - 8.

Taylor, R. L., Heer, D., & Fiez, T. S. (2003). Using an Integrated Platform for Learning (TM) to reinvent engineering education. IEEE Transactions on Education, 46(4), 409 - 418.

Tollerud, T. (1990). The perceived self-efficacy of teaching skills of advanced doctoral students and graduates from counselor education programs. Unpublished Dissertation, The University of Iowa, Iowa City, Iowa.

Torvi, D. A. (1994). Engineering graduate teaching assistant instructional programs: Training tomorrow's faculty members. Journal of Engineering Education, 83(4), 375 - 381.

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

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

Tubb, G. W. (1974). Heuristic questioning and problem solving strategies in mathematics graduate teaching assistants and their students. Unpublished Dissertation, Texas A & M University.

Verleger, M., & Velasquez, J. (2007). An engineering teaching assistant orientation program: Guidelines, reactions, and lessons learned from a one-day intensive training program, ASEE/IEEE Frontiers in Education Conference. Milwaukee, WI.