JET
The Journal of Effective Teaching
an online journal devoted to teaching excellence

Toward a Measure of Professional Development for Graduate Student Teaching Assistants

Sue Ellen DeChenne[1a], Kristin Lesseig[b], Shawn M. Anderson[c], Sissi L. Li[c], Nancy L. Staus[c], and Celeste Barthel[d]
[a]*University of Nebraska – Lincoln, Lincoln, NE 68583-0858*
[b]*Washington State University – Vancouver, Vancouver, WA 98686*
[c]*Oregon State University, Corvallis, OR 97331-6508*
[d]*Wilson College, Chambersburg, PA 17201*

Abstract

This study describes the development and validation of an instrument to measure graduate teaching assistants' (GTAs) learning about teaching during professional development. In the pilot study, exploratory factor analysis of data from 239 graduate students indicates a single factor structure. The second study, involving 177 science, technology, engineering, and mathematics (STEM) GTAs, confirms the single factor structure of the instrument. The instrument is highly reliable with both populations. The instrument is correlated to the hours STEM GTAs spend in professional development and their self-efficacy in teaching. It is sensitive to departmental differences between GTAs perceptions of their professional development. This instrument has multiple possible users including university faculty involved in GTA professional development as well as educational researchers. University faculty can use it for needs assessment during GTA program development, comparisons among departmental programs, and in improving current GTA programs. It also provides a sensitive measurement of the quality of GTA professional development in multiple program research studies.

Keywords: Graduate teaching assistants, professional development, scale development, scale validation; science, technology, engineering, and mathematics, STEM.

Professional development in teaching for graduate teaching assistants (GTAs) is vital to the instructional mission of universities because GTAs play a prominent role in current and future instruction of undergraduates. Universities are heavily dependent on GTAs (Johnson & McCarthy, 2000; Nyquist, Abbott, Wulff, & Sprague, 1991). For example, GTAs provide 91% of biology laboratory instruction at research universities (Sundberg, Armstrong, & Wischusen, 2005) and 84% of counseling psychology programs employ GTAs (Prieto & Scheel, 2008). Many of the first experiences that undergraduates have in their college classrooms are closely associated with their GTAs. In addition to their current teaching duties, GTAs will become the next generation of faculty members. Two-thirds of doctoral students are interested in a faculty career, many in smaller col-

[1] Corresponding author's email: sdechenne2@unlserve.unl.edu

leges where teaching is their primary responsibility (Golde & Dore, 2001). Given the economic state of higher education, this dependence on GTAs as university instructors is not likely to lessen. Costs of higher education, especially at public four-year institutions, are continuing to rise (College Board, 2007) and one way to cut university costs is to have more courses taught by part-time instructors and GTAs (Bettinger & Long, 2004).

Despite a concern for and research on GTA professional development that spans more than 40 years (eg. Costin, 1968; Sandi-Urena, Cooper, & Gatlin, 2011), there is evidence that most GTAs are still poorly or completely unprepared to teach. Although there are differences between academic fields, in a national study 51% of GTAs have access to a departmental workshop and 46% to university coursework in teaching (Golde & Dore, 2001). There appears to be little overall improvement in the amount of GTA preparation in the last 15 years. In national studies within specific disciplines, 37% of chemistry programs (Abraham et al., 1997) and 32% of psychology programs (Meyers & Prieto, 2000) had no GTA professional development available a little over a decade ago. More recently, that lowered to 18% of counseling psychology programs, however only 62% of the GTAs had received professional development in teaching and for an average of less than three days (Prieto & Scheel, 2008). Sundberg et al. (2005) found the number of biology programs which provided three or more days of GTA professional development had increased in the prior decade from 4% to 15% of research and 6% of comprehensive universities. However, the number of programs offering university coursework in teaching had dropped from 14% to 6% for research and 12% for comprehensive universities, indicating an overall status quo in GTA professional development offered.

Many of the GTAs who do receive professional development find it is not an adequate preparation for teaching (Commander, Hart, & Singer, 2000; Fagen & Wells, 2004; Jones, 1993; Luft, Kurdziel, Roehrig, & Turner, 2004; Prieto & Scheel, 2008; Shannon, Twale, & Moore, 1998). A common format for GTA professional development is a short, pre-term program that focuses on administrative details and university policies and procedures with little time set aside for instruction in teaching and learning (Kurdziel & Libarkin, 2003; Prieto & Scheel, 2008; Rushin et al., 1997). This information is important, but it does little to improve GTA instruction. Instead, best practices indicate that GTA professional development should include: active learning, peer interaction, practice and feedback, formative and summative assessment, reflection, generic and discipline specific teaching information, and should be of sufficient duration to ensure learning occurs (Park, 2004). There should also be an initial program needs assessment and continuing program evaluation (Park, 2004). The following sums up many GTAs' experience with GTA professional development:

> "I have always considered teaching my main reason for pursuing an academic degree. I am amazed at how little preparation I am receiving in how to teach. I am still planning on pursuing a teaching position but am filling in the gaps in my education and preparation on my own time with little encouragement from my academic program." (Molecular Biology Doctoral Student, Golde & Dore, 2001, pg. 22)

There are many individual professional development programs that demonstrate evidence for improved and/or quality teaching by the GTAs enrolled (e.g. Davis & Kring, 2001; Kurdziel & Libarkin, 2003; Young & Bippus, 2008). However, how are GTAs, faculty preparing and teaching professional development, university administrators, and educational researchers to recognize these programs? The purpose of this study is to present a valid and reliable instrument that can be used to measure GTA's perception of their learning about important topics in teaching during GTA professional development programs. Such an instrument can be used for GTA professional development needs assessment, evaluation and improvement of current programs as well as in educational research.

Literature Review

Improving the teaching preparation of GTAs has many benefits. It provides better instruction to students, demonstrates institutional commitment to the primary mission of excellence in education, and it is also beneficial for the graduate student; providing financial support and an apprenticeship in teaching for future faculty members (Park, 2004; Svinicki, 1995-96). Studies of individual GTA professional development programs indicate quality professional development can build graduate students' knowledge and skills in teaching (Belnap, 2005; Carroll, 1977; Davis & Kring, 2001; Hadre & Chen, 2005; Trouba, 2009), self-confidence in teaching (Salinas, Kozuh, & Seraphine, 1999), and improve their ability to obtain an academic position upon graduation (Svinicki, 1995-96). GTA professional development can also have a positive effect on student achievement (Childs, 2006; Ezrailson, 2004; Norris, 1991). Many graduate students are interested in academic faculty careers and professional development should help prepare them for future faculty positions (Austin, 2002; Golde & Dore, 2001; Park, 2004). Even if a GTA is not interested in an academic career, the presentation and interpersonal skills learned in teaching can benefit other career paths.

There is evidence that GTA professional development programs can also positively impact self-efficacy in teaching (Komarraju, 2008; Prieto & Altmaier, 1994; Prieto & Meyers, 1999; Young & Bippus, 2008), although not all studies show a significant effect (Liaw, 2004; Tollerud, 1990). Self-efficacy is a central component in social cognitive theory (Bandura, 1977, 1986, 1997) and research has demonstrated that when training for a specific skill, self-efficacy tends to be positively correlated with performance (Bandura, 1997; Gist, Schwoerer, & Rosen, 1989; Pajares, 1996). Self-efficacy in teaching is an instructors' belief that they will be able to effectively teach a given population of students a specific subject (Bandura, 1997) and has been shown to be a valuable predictor for student achievement, teacher retention, and persistence in the face of teaching difficulties (for a review see Tschannen-Moran, Hoy, & Hoy, 1998).

Despite the studies of individual programs that improve GTA knowledge and skills, studies that look at GTA professional development across multiple programs often find that professional development is not very effective and GTAs have naïve conceptions about teaching and learning. In multiple program studies, there are no significant effects for professional development on student evaluations of teaching (Shannon et al., 1998) or in some self-efficacy in teaching studies (Liaw, 2004; Tollerud, 1990). Additionally, GTAs indicate a need for more and better professional development (Commander et al., 2000; Jones, 1993; Prieto & Scheel, 2008). In a national survey on doctoral education, 45% of graduate students feel they are not prepared to teach (Fagen & Wells, 2004). Luft et al. (2004) find that GTAs primarily work autonomously, use direct instruction, and have intuitive views of student learning and motivation. While Saroyan, Dagenais, and Zhou (2009) find that advanced doctoral students hold views of teaching as the transmission of information and preparing and managing instruction.

In quantitative studies, when GTA professional development in teaching is examined across multiple programs, it is either measured as a dichotomous variable (i.e., presence or absence) (e.g. Liaw, 2004; Prieto & Altmaier, 1994) or as the amount of time in professional development (e.g. Shannon et al., 1998; Tollerud, 1990). This type of measure of GTA professional development yields minimal information regarding the quality of the programs and can result in no impact (Liaw, 2004; Shannon et al., 1998; Tollerud, 1990) or low impact for GTA professional development (Prieto & Altmaier, 1994). The actual impact of GTA professional development on teaching is important and having a narrow understanding limits researcher conclusions from the data. It is possible that high quality professional development programs are having an effect in research that includes multiple programs, but low quality programs are washing out or diluting that effect. Rather than measuring the presence/absence or time in GTA professional development, it is important to evaluate the learning of the individual in the program and determine whether that has an effect on the outcome variables of the research. Used this way, the GTAs perception of their learning in professional development becomes a proxy for the quality of professional development programs.

An instrument that measures GTA professional development quality should be based on concepts important in teaching. The instrument should be sensitive to differences in quality between professional development programs and correlate to self-efficacy in teaching. It should also correlate to time spent in GTA professional development, although given the variable quality of such programs the correlation may be low. This study aims to provide such a valid and reliable measure of the GTAs' perception of their professional development in teaching.

Methods

Instrument Development and Refinement

The instrument was developed by a team of science and mathematics education doctoral students during a course on quantitative methodology, who were experienced teachers from middle school through graduate school (*M* = 8.5 years teaching experience). There were two goals in developing the items: a short and easy to administer survey, and a broad coverage of important topics in teaching. Ideas were generated during class discussions producing a list of items and categories that were important in learning to teach. Additional item ideas were generated by searching the GTA literature and from literature on effective teaching.

To make sure that no items important to graduate students were missed, five focus groups of three to four graduate students were interviewed. These graduate students were self-selected; flyers advertising a discussion about teaching were posted around the university and sent to departments. Interested GTAs signed up to join the discussion and food was provided. The focus group conversations were semi-structured. The graduate students were asked about their experiences in learning to teach and what they wished they had experienced in GTA professional development. Results from these focus groups were incorporated into the emerging categories of important teaching topics.

Categories of topics important to teaching were narrowed down and refined during classroom discussions among the doctoral students. Finally, specific items for each category were written. To determine face validity, the items were sent to science and mathematics education faculty, additional science and mathematics doctoral students, and other faculty that had experience with either GTA professional development or survey research. After incorporating suggested changes, the clarity of the items was tested with biochemistry graduate students at a university in the Southwest (N = 34). Responses from those graduate students were incorporated into the items.

The items from the development phase were then tested and further refined through two studies. First, the items on the instrument were pilot tested in a university-wide survey of graduate students' professional development needs. Then one item was further refined, by splitting it into two items to remove multiple concepts in an item (Tables 1 and 2). Additionally, three items were added; one that dealt with harassment and two that asked about overall professional development experience. The revised items were then used in a study of science, technology, engineering, and mathematics (STEM) GTA teaching to determine perception of learning about teaching during professional development.

Participants and Administration

Pilot Study. From May to July of 2008, graduate students from 45 departments at a Pacific Northwest university with a Carnegie classification of Research Universities with Very High research activity were contacted by e-mail four times about an on-line survey of graduate student teaching experiences. They were asked to respond to the survey, which included the items about GTA professional development and demographics. Responses were then collated and downloaded using on-line software (Survey Monkey).

STEM GTA Study. GTAs from nine STEM departments at the same university as the pilot study were administered the professional development items as part of a larger study of STEM GTA teaching. Data was collected from Fall 2008 through Fall 2009. One of two administration techniques was used depending on the department. Questionnaires were distributed to the GTAs through the department mail system, collected in a sealed container in the departmental office, and then the container was picked up by a researcher. Alternatively, questionnaires were administered during a GTA professional development course and collected by one of the researchers at that time. Additional questions about hours of professional development, self-efficacy in teaching (DeChenne, 2010), and demographics were also included in this analysis.

Analysis

The professional development items were analyzed in the pilot study using principle axis exploratory factor analysis (EFA) with Varimax rotation, Kaiser criterion (Guttman, 1954), and Scree test (Cattell, 1966). Confirmatory factor analysis (CFA) was then used to determine whether the variables measured a single factor. CFA of the refined professional development items from the STEM GTA study was also used to examine whether the variables measuring this latent factor provided 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).

Internal consistency of the professional development items in both the pilot and STEM GTA study were examined with Cronbach alpha reliability coefficients. An alpha coefficient ≥ 0.65 indicated that items measured the same concept and justified combining items into a single index (Cortina, 1993).

Using the STEM GTA data, Pearson's product moment correlations (r) between the STEM GTAs perception of professional development, hours spent in professional development, presence or absence of professional development, and self-efficacy in teaching were determined. According to Cohen (1988) correlations less than .10 are considered small/weak, those around .30 are moderate/medium and those greater than .50 are large/strong. A second comparison of perception of professional development between STEM GTAs who had professional development and those who didn't was done using a Mann-Whitney U test. An ANOVA was used to compare the GTA perceptions of professional development between departments that had at least ten GTAs participating in the study.

Results

Pilot Study

239 graduate students responded to the survey (12% return rate). Departments reported by the graduate students were divided into three broad content areas – science, engineering, and liberal arts. 45% of the graduate students were from the sciences, 31% from engineering, and 24% from liberal arts. This distribution was not significantly different from the actual graduate student population distribution in each content area ($\chi^2 = 0.35$, $p = 0.838$).

EFA of the 12 items revealed one or two factors explaining 50% of the variance. Two of the factors had a Kaiser criterion (Guttman, 1954) greater than one. However, a Scree test (Cattell, 1966) and factor plot in rotated factor space both suggested that one factor could be found in the data. Additionally, approximately half of the items cross-loaded between the two factors and the two factors were highly correlated (r = .695) which also indicated a single factor.

A single factor was confirmed with CFA which demonstrated an acceptable model fit and supported construct validity (NNFI = .901, CFI = .924, RMSEA = .059, Table 1). All variables loaded between .53 and .74 and were significant at $p < .05$. The GTA perception of needs in professional development was highly reliable (α = .90), all the variables met the criterion of item total item correlations being greater than .40, and deletion of any item did not improve reliability. GTAs indicated that overall these items were helpful in learning to teach (M = 2.01 on a 5 point scale of very helpful to not at all).

Table 1. Factor Loadings for Confirmatory Factor Analysis of Graduate Student Perception of Need for Topics in Professional Development.

	Factor Loadings[1]
Of the following teaching topics which would be helpful to be included in TA training?[2]	
Assisting distressed students	.74
Power/authority relationships in the classroom	.74
Teaching students with different skills/knowledge[3]	.71
Motivating students[3]	.70
Managing disruptive students	.70
Interacting professionally one-on-one with your students[4]	.69
Teaching and learning styles[5]	.67
Facilitating group discussions[5]	.65
Presenting material to a large group of students	.62
Grading	.59
Communicating with course lead instructor[4]	.54
Developing assignments/laboratories/projects/exams	.53
Mean	2.01
Cronbach α	.90

[1]All factor loadings are significant at $p < .05$. Model fit indices are NNFI = .901, CFI = .924, RMSEA = .059.
[2]Items coded on a 5 point scale of 1 = very helpful to 5 = not at all.
[3-5]Errors allowed to co-vary to achieve fit indices.

STEM GTA Study

There were 177 returned surveys (54% response rate) from nine STEM departments. Engineering GTAs comprised 61% of the participants with 39% in science or mathematics. Twelve percent of the sample had no professional development of any kind. The STEM GTAs' perception of professional development was slightly above neutral, indicating that they learned some skills in their professional development (M = 3.18 on a 5 point scale of never learned to learned very well). CFA demonstrated an acceptable model fit and

Table 2. Factor Loadings for Confirmatory Factor Analysis of GTA Perception of Professional Development with STEM GTAs.

	Factor Loadings[1]
Of the following teaching topics and skills, please rate how well you have learned these in GTA training?[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	3.18
Cronbach α	.96

[1]All factor loadings are significant at $p < .05$. Model fit indices are NNFI = .925, CFI = .935, RMSEA = .083.
[2]Items coded on a 5 point scale of 1 = never learned to 5 = learned very well.
[3]Items coded on a 5 point scale of 1 = not effective to 5 = very effective.
[4]Errors allowed to co-vary to achieve fit indices.

supported construct validity of the STEM GTA perception of professional development factor (NNFI = .925, CFI = .935, RMSEA = .083; Table 2). All variables loaded between .60 and .86 and were significant at $p < .05$. The GTA perception of professional development 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 reliability.

There was a significant correlation between the hours the STEM GTA's spent in and their perception of professional development ($r = .21$, $p < .01$, Table 3). The more hours spent in professional development the higher their perception of their learning about teaching. There was a lower significant correlation to the presence or absence of professional development ($r = .17$, $p < .05$, Table 3). However, comparing STEM GTAs who had to

Table 3. Correlational Analysis of Teaching Professional Development Instrument.

Measures	Mean	Teaching PD Instrument[2]	Hours PD[3]	PD[4]
Teaching PD Instrument[1, 2]	3.18			
Hours PD[3]	21 hours	.21**		
PD[4]	88%[5]	.17*	.23**	
Self-efficacy in teaching[1]	4.12	.34**	.18*	.03

*Correlation is significant at the .05 level.
**Correlation is significant at the .01 level.
[1]All scales were rated on a scale of 1 to 5, with 5 being the best in each scale.
[2]Teaching Professional Development Instrument
[3]Hours spent in teaching professional development
[4]Presence/Absence of teaching professional development
[5]Percentage of GTAs with Teaching professional development

those who did not have any professional development was not significant, but showed a trend for higher scores in those STEM GTAs who had received professional development (No Professional development, M = 2.70, N = 14; Professional development, M = 3.24, N = 144; p = .095; Mann-Whitney U test). The GTA professional development instrument also showed higher correlations to GTA self-efficacy in teaching than hours of professional development or a dichotomous variable of professional development (Table 3).

There was a significant difference between the STEM GTAs perception of their learning about teaching in professional development by department (F = 3.025, df = 153, p = .008). A Levene's test indicated that the variances were equal and therefore a post-hoc Least Squares Difference test was used to indicate between group significant differences (Table 4). The Electrical Engineering department was sampled in both Fall '08 and '09 during their professional development course which was taught by different instructors each Fall. As indicated in Table 4 the Electrical Engineering department was significantly different than Geosciences, Mechanical Engineering, and Chemistry in '08 and Geosciences and Mechanical Engineering in '09.

Discussion

Development, Validity and Reliability

The purpose of this study was to develop an instrument to measure GTA professional development. Essential to this process was to work toward establishing reliability and validity of the GTA professional development measure. The GTA professional development instrument was developed as a result of an extensive dialogue among experienced educators, a review of the literature on GTA professional development, focus group interviews with graduate students about teaching, and reviewed by experienced educational researchers. Two studies were then used to determine the factorial structure, validity, and reliability of the items in the instrument. The first was a pilot study to determine reliability and the factorial structure of the items in the instrument. The second study was a

Table 4. Post-Hoc Comparisons of STEM GTA Perceptions of Teaching Professional Development[1]

Department	Comparison Department	N	Mean[2]	Significant *p* values
Electrical Engineering ('08) N = 38 *M* = 3.53	Geosciences	11	2.60	.005
	Mechanical Engineering	12	2.61	.004
	Physics	18	3.11	
	Chemistry	21	2.94	.024
	Math	14	3.10	
	Electrical Engineering ('09)	40	3.44	
Electrical Engineering ('09) N = 40 *M* = 3.44	Geosciences	11	2.60	.010
	Mechanical Engineering	12	2.61	.009
	Electrical Engineering ('08)	38	3.53	
	Physics	18	3.11	
	Chemistry	21	2.94	
	Mathematics	14	3.10	

[1]One-way ANOVA, $F = 3.025$, $df = 153$, $p = .008$
[2]Mean of items coded on a 5 point scales (1 = never learned to 5 = learned very well or 1 = not effective to 5 = very effective, see Table 2).

cross-validation study to confirm the factorial structure of the items, to provide reliability and further the validity of the measure. Assertions related to instrument and reliability must be viewed as sample dependent.

Although the response rate in the pilot study was low, the sample sizes in both studies were sufficient. Costello and Osborne (2005) argued that in exploratory factor analysis, a ratio of at least 10 participants to each item in the instrument provided an average of less than one (0.70) item misclassified on the wrong factor and in the pilot study the ratio was twice that (20 participants per item). It was also possible that some of the participants in the pilot study participated in the STEM GTA study, since they were drawn from the same university, but at different times. The number of GTAs who participated in both studies was probably low because of graduation of some of the graduate students between the studies, and the movement of graduate students out of GTA roles into graduate research assistantship roles which was prevalent in STEM fields. For example, in the largest department represented in both studies (and the university), GTAs were drawn almost exclusively from first year graduate students; therefore these GTAs could not have participated in the pilot study. Additionally, there was no reason to think that the responses of any possible redundant participants would be unique from their peers. Therefore, if they were not different from the rest of the participants, then including them in both studies should not have changed the results.

Factor analysis indicated that there was one dimension to this measure. EFA of the pilot study suggested the possibility of two factors, but given the Scree test, strong cross-loadings in the two factor structure, and the factor plot, a determination of one factor was made. This was confirmed by CFA with the pilot study data, which provided a good fit using a single factor (NNFI = .901, CFA = .924, and RMSEA = .059). Additionally all factor loadings were above 0.5. Costello and Osborne (2005) state “a factor with…5 or more strongly loaded items (.50) are desirable and indicate a solid factor (p. 5).” After further refinement of the items, the STEM GTA study also indicated a single strong factor using CFA (NNFI = .925, CFI = .935, RMSEA = .083) with high factor loadings. In both studies the factor was also internally highly reliable (r = .90 and .96 respectively).

The GTA professional development instrument significantly correlated with hours spent in professional development (r = .21) and a dichotomous variable of presence or absence of professional development (r = .17). GTAs should be learning about teaching during their professional development. Unfortunately, not all professional development was of the same quality; therefore it was not surprising that the correlations were small to moderate. This was further support by the results of the non-parametric mean comparison between those who had not received any professional development and those who had, which was not significantly different although there was a higher mean for those with some professional development. The sample sizes were vastly different (N = 14 vs. N=144) which may have contributed to the non-significance. It also could have been that the perceived quality of the professional development in some of the programs wasn’t very different from not receiving any professional development. This was supported by the departmental averages (Table 4). Two of the departments, Geoscience and Mechanical Engineering, had means (M = 2.60 and 2.61 respectively) lower than the mean for those GTAs reporting no professional development (M = 2.70).

Results from the ANOVA of perception of learning during professional development with the STEM GTAs had a significant effect (F = 3.025, df = 153, p = .008) indicating the instrument’s ability to distinguish differences among various professional development programs provided in the departments sampled. The post-hoc comparisons indicated significant effects between Electrical Engineering and Geosciences, Mechanical Engineering, and Chemistry (Table 4). Each of the STEM departments in the study had different types of professional development ranging from no required professional development to at least a one quarter course in teaching and learning. Those departments with the lowest scores had little or no required professional development, at most a couple of days before the quarter began. Those with the highest scores all required at least one quarter of a university course in teaching.

The sensitivity of the GTA professional development instrument was demonstrated by the differential correlations with self-efficacy in teaching. The GTA professional development instrument showed the highest correlations to self-efficacy in teaching, with the self-efficacy in teaching correlation to hours of professional development still significant but much lower (moderate versus low, Table 3). When using the dichotomous variable of presence or absence of professional development there was no correlation to GTA self-efficacy in teaching. This pattern is expected because of the variable types of professional development programs offered by the departments in the study. Good professional development programs should increase the GTAs knowledge and skills in teaching and therefore their self-efficacy in teaching. The quality of the professional development programs was variable (Table 4) resulting in a weaker self-efficacy in teaching correlation to hours and no correlation to a dichotomous measure of professional development. However, the GTA professional development scale distinguished between these differences and thus was moderately correlated.

Interestingly, the graduate students in the pilot study thought the items indicated in the measure would be helpful in learning to teach (M = 2.01 on scale of 1 very helpful to 5 not at all, Table 1) but the STEM GTA students were essentially neutral in how well they had learned the items (M = 3.18 on a scale of 1 never learned to 5 learned very well, Table 2). This result with the two populations of graduate student at the same university is similar to other research that indicate that GTAs want to learn to teach but don't feel that they have been taught about teaching and learning (Commander et al., 2000; Fagen & Wells, 2004; Jones, 1993; Piccinin & Fairweather, 1996-97; Williams & Schaller, 1994).

Implications

This instrument has various uses; at the university, program, and individual GTA levels as well as for educational research. At the university level this instrument could be used as part of a needs assessment when developing a teaching certification program for graduate students as it was at the university in this study (Table 1). The instrument could be used to determine what areas of teaching are most needed by the GTAs at the university. This information could then be used to help design a more comprehensive certification program. Determining needs is an important first step in developing a GTA program (Commander et al., 2000; Park, 2004). It could also be used to identify successful departmental GTA professional development programs within a university. Other departments could collaborate with these programs to detail what aspects of the programs contribute to their success which could then be adopted across departmental programs at the university. With a tightening of university budgets, it is fiscally sound to invest in GTA professional development that is successful. So having a more nuanced instrument to evaluate GTA professional development learning is particularly important. Additionally, this instrument is easy to administer and evaluate, making it especially useful across programs and with large numbers of GTAs.

The instrument is also useful for departmental programs and individual GTAs. The instrument can be used for a needs assessment within a department prior to developing a departmental GTA professional development program. It could be particularly valuable within a department as an evaluative tool for a current GTA professional development program. With the feedback from this tool, changes could be made to improve the program. Over time, the professional development program should then continue to better meet the needs of the GTAs enrolled. Faculty who supervise GTAs, especially in large numbers, could use this instrument on an individual level to determine where a GTA feels they may need more professional development activities. It could also be used to track an individual GTA's progress within a GTA professional development program.

This instrument is also highly valuable to educational researchers. Multiple program studies use a proxy for professional development, either presence or absence of professional development or time spent in professional development. Neither of these proxies gives any indication of the quality of the professional development in the various programs included in the study. Especially when conducting large cross-department or cross-university studies, there is no guarantee of similar quality in the professional development accessible to the GTAs. The measure developed in this study is more sensitive than presence/absence or time spent measures as indicated by the correlations with self-efficacy in teaching (Table 3) and the significant differences in department means (Table 4). Even for the studies that do have a significant effect for professional development, the increased sensitivity of this measure may show a greater effect for professional development, especially if there were many low quality programs in the study. This instrument gives the researcher a measure of the GTAs perception of their learning of important concepts in teaching which is a more sensitive measure of professional development and should provide better results in quantitative studies of GTAs.

Use of this instrument with graduate students at other universities and in other major areas of study will further validate the items in the instrument as will quantitative studies using this instrument. This instrument provided a valuable measure in a study of self-efficacy in teaching for STEM GTAs (manuscript under review) that was not apparent with the other measures of professional development. This measure will fulfill a need for universities, programs, and faculty providing GTA professional development as well as educational researchers studying GTA professional development.

Acknowledgements

We would like to thank the other members and professor of the quantitative class who helped generate items for the instrument: Greg Kise, Danielle Amatore, and Dr. Larry Flick. We would also like to thank the graduate students who took time from their busy schedules to attend a focus group conversation or respond to either survey. An earlier version of this article was presented at and greatly improved by comments from the 2009 annual conference of the American Educational Research Association in San Diego, CA.

References

Abraham, M. R., Cracolice, M. S., Graves, A. P., Aldhamash, A. H., Kihega, J. G., Gil, J. G. P., et al. (1997). The nature and state of general chemistry laboratory courses offered by colleges and universities in the United States. *Journal of Chemical Education, 74*(5), 591-594.

Austin, A. E. (2002). Preparing the next generation of faculty: Graduate school as socialization to the academic career. *The Journal of Higher Education, 73*(1), 94-122.

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

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

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

Belnap, J. (2005). *Putting TAs into context: Understanding the graduate mathematics teaching assistant.* Unpublished Dissertation, University of Arizona, Tucson.

Bettinger, E., & Long, B. T. (2004). Do college instructors matter? The effects of adjuncts and graduate assistants on students' interest and success. Working paper No. 10370. Cambridge, MA: National Bureau of Economic Research. Retrieved August 2010, from http://www.nber.org/papers/w10370.pdf.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (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.

Childs, K. M. (2006). *The relationship between graduate teaching assistant instructor expertise and algebra performance of college students.* Unpublished Dissertation, University of Missouri - Kansas City.

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

College Board. (2007). *Trends in College Pricing, 2007*. Retrieved August 2011, from http://www.careercornerstone.org/pdf/universities/tuition07.pdf.

Commander, N. E., Hart, L., & Singer, M. (2000). Preparing tomorrow's faculty: An assessment model to determine institutional needs. *The Journal of Graduate Teaching Assistant Development, 7*(2), 93 - 111.

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 Assessment, Research & Evaluation, 10*(7), Retrieved from: http://pareonline.net/pdf/v10n17.pdf.

Costin, F. (1968). A graduate course in the teaching of psychology: Description and evaluation. *Journal of Teacher Education, 10*, 425 - 432.

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. (2010). *Learning to teach effectively: Science, technology, engineering, and mathematics graduate teaching assistants' teaching self-efficacy.* Unpublished Dissertation, Oregon State University, Corvallis.

Ezrailson, C. M. (2004). *EMIT: Explicit modeling of interactive-engagement techniques for physics graduate teaching assistants and the impact on instruction and student performance in calculus-based physics.* Unpublished Dissertation, Texas A & M University, College Station.

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

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*. Retrieved September 2007, from http://phd-survey.org/report%20final.pdf.

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika, 19*, 149-161.

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

Johnson, B., & McCarthy, T. (2000). Casual labor and the future of the academy. *Thought and Action*, 107-120.

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

Komarraju, M. (2008). A social-cognitive approach to training teaching assistants. *Teaching of Psychology, 35*, 327-334.

Kurdziel, J. P., & Libarkin, J. C. (2003). Research methodologies in science education: Training graduate teaching assistants to teach. *Journal of Geoscience Education, 51*(3), 347 - 351.

Liaw, E. C. (2004). *"How are they different?" A comparative study of native and nonnative foreign language teaching assistants regarding selected characteristics: Teacher efficacy, approach to language teaching/learning, teaching strategies and perception of nativeship.* Unpublished Dissertation, The Ohio State University, Colombus.

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., & Prieto, L. R. (2000). Training in the teaching of psychology: What is done and examining the differences. *Teaching of Psychology, 27*(4), 258 - 261.

Norris, T. (1991). Nonnative English-speaking teaching assistants and student performance. *Research in Higher Education, 32*(4), 433 - 448.

Nyquist, J. D., Abbott, R. D., Wulff, D. H., & Sprague, J. (1991). *Preparing the professoriate of tomorrow to teach: Selected readings in TA training*. Dubuque, IA: Kendall/Hunt Publishing Co.

Pajares, F. (1996). *Assessing self-efficacy beliefs and academic outcomes: The case for specificity and correspondence.* Paper presented at the American Educational Research Association.

Park, C. (2004). The graduate teaching assistant (GTA): Lessons from North American experience. *Teaching in Higher Education, 9*(3), 349 - 361.

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.

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., & Scheel, K. (2008). Teaching assistant training in counseling psychology. *Counseling Psychology Quarterly, 21*(1), 49 - 59.

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.

Salinas, M. F., Kozuh, G., & Seraphine, A. E. (1999). I think I can: Improving teaching self-confidence of international teaching assistants. *The Journal of Graduate Teaching Assistant Development, 6*(3), 149 - 156.

Sandi-Urena, S., Cooper, M. M., & Gatlin, T. A. (2011). Graduate teaching assistants' epistemological and metacognitive development. *Chemistry Education Research and Practice, 12*, 92-100.

Saroyan, A., Dagenais, J., & Zhou, Y. (2009). Graduate students' conceptions of university teaching and learning: formation for change. *Instructional Science, 37*(6), 579-600.

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.

Sundberg, M. D., Armstrong, J. E., & Wischusen, E. W. (2005). A reappraisal of the status of introductory biology laboratory education in U.S. College and Universities. *The American Biology Teacher, 67*(9), 525-529.

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.

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, IA.

Trouba, J. C. (2009). *The design, implementation, and evaluation of a teacher training workshop for mathematics graduate teaching assistants.* Unpublished Dissertation, Montana State University, Bozeman.

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

Williams, D. E., & Schaller, K. A. (1994). A cross-disciplinary study of GTAs' views on topics used in training programs. *The Journal of Graduate Teaching Assistant Development, 2*(1), 13 - 20.

Young, S. L., & Bippus, A. M. (2008). Assessment of graduate teaching assistant (GTA) training: A case study of a training program and its impact on GTAs. *Communication Teacher, 22*(4), 116-129.