

Essay

A Conceptual Framework for Graduate Teaching Assistant Professional Development Evaluation and Research

Todd D. Reeves,^{†*} Gili Marbach-Ad,[‡] Kristen R. Miller,[§] Judith Ridgway,^{||}
Grant E. Gardner,[¶] Elisabeth E. Schussler,[#] and E. William Wischusen[@]

[†]Educational Research and Assessment, Northern Illinois University, DeKalb, IL 60115; [‡]College of Computer, Mathematical and Natural Sciences, University of Maryland, College Park, MD 20742; [§]Division of Biological Sciences, University of Georgia, Athens, GA 30602; ^{||}Center for Life Sciences Education, Ohio State University, Columbus, OH 43210; [¶]Department of Biology, Middle Tennessee State University, Murfreesboro, TN 37132; [#]Ecology & Evolutionary Biology, University of Tennessee, Knoxville, TN 37996; [@]Biological Sciences, Louisiana State University, Baton Rouge, LA 70803

Submitted October 28, 2015; Revised March 16, 2016; Accepted March 19, 2016
Monitoring Editor: Marilyne Stains

Biology graduate teaching assistants (GTAs) are significant contributors to the educational mission of universities, particularly in introductory courses, yet there is a lack of empirical data on how to best prepare them for their teaching roles. This essay proposes a conceptual framework for biology GTA teaching professional development (TPD) program evaluation and research with three overarching variable categories for consideration: outcome variables, contextual variables, and moderating variables. The framework's outcome variables go beyond GTA satisfaction and instead position GTA cognition, GTA teaching practice, and undergraduate learning outcomes as the foci of GTA TPD evaluation and research. For each GTA TPD outcome variable, key evaluation questions and example assessment instruments are introduced to demonstrate how the framework can be used to guide GTA TPD evaluation and research plans. A common conceptual framework is also essential to coordinating the collection and synthesis of empirical data on GTA TPD nationally. Thus, the proposed conceptual framework serves as both a guide for conducting GTA TPD evaluation at single institutions and as a means to coordinate research across institutions at a national level.

Biology graduate teaching assistants (GTAs) have important instructional roles in undergraduate education at colleges and universities. Rushin *et al.* (1997) reported that of 153 surveyed graduate schools, 97% used GTAs in some form of undergraduate instructional role. In another study, Sundberg *et al.* (2005) reported that biology GTAs teach 71% of laboratory courses at comprehensive institutions and 91% of laboratory courses at research institutions. More recently, a national

survey of 85 faculty and staff providing teaching professional development (TPD) to biology GTAs found that 88% of those surveyed were preparing GTAs to teach introductory-level biology courses (Schussler *et al.*, 2015). Thus, GTAs have a potentially powerful impact on undergraduate student learning at many colleges and universities, especially in introductory laboratories and introductory-level lecture courses.

Introductory science courses are often the “gateway” to the attainment of undergraduate science degrees, and progression through the degree and beyond often depends on undergraduate student performance in these early courses (Seymour and Hewitt, 1997). This makes these courses uniquely important for student retention as the nation attempts to increase the number of science, technology, engineering, and mathematics (STEM) graduates (President's Council of Advisors on Science and Technology, 2012). Biology education researchers have argued that because of the smaller and more intimate class size of introductory-course laboratory and discussion sections, GTAs contribute meaningfully to retention efforts, because they have

CBE Life Sci Educ June 1, 2016 15:es2

DOI:10.1187/cbe.15-10-0225

*Address correspondence to: Todd D. Reeves (treeves@niu.edu).

© 2016 T. D. Reeves *et al.* CBE—Life Sciences Education © 2016 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Non-commercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

more personal contact with first-year students than do most faculty members (Rushin *et al.*, 1997). Providing biology GTAs with opportunities to develop instructional expertise that maximizes student learning outcomes should be a priority for the universities that employ them, yet GTA teaching responsibilities are often relegated to secondary status or sometimes even actively discouraged (Nyquist *et al.*, 1999; Gardner and Jones, 2011).

Currently, there is wide variation among universities and departments vis-à-vis biology GTA TPD. A recent national survey found that 96% of responding TPD practitioners provided some formal TPD to their biology GTAs (e.g., TPD workshop) but that these programs varied extensively in terms of total contact hours (2–100 h per academic year). Because many of these contact hours are delivered as one-time presemester workshops between 2 and 5 h in length (Schussler *et al.*, 2015), GTA TPD does not generally meet research-based TPD standards (Garet *et al.*, 2001; Desimone *et al.*, 2002). Institutional differences in the levels of funding and support for TPD programs (Schussler *et al.*, 2015) suggest that university and/or department contextual variables may impact TPD design quality (as suggested by Park, 2004; Seymour *et al.*, 2005).

The current state of biology GTA TPD highlights the need for further research on biology GTA TPD that accounts for the diverse institutional contexts in which these TPD programs are implemented. This mirrors recent calls for “biology education research 2.0” to better consider contextual factors (Dolan, 2016). The current literature base for GTA TPD is primarily limited to small-scale evaluation studies concerning individual TPD programs (Abbott *et al.*, 1989; Marbach-Ad *et al.*, 2015a). Though these studies can be used to suggest practices that TPD leaders may adopt, there is no guarantee that what worked at one institution will effectively transfer to a different context. At the same time, existing studies often do not compare the efficacy of different TPD practices and frequently use different assessment tools, making cross-institutional and cross-study comparisons difficult. A systemic approach to evaluation and research is needed to identify evidence-based practices in biology GTA TPD.

This article proposes a conceptual framework for GTA TPD evaluation and research suggesting that the most important TPD program outcomes to measure (as determined by our BioTAP¹ working group and the current literature) are GTA cognition, GTA teaching practice, and undergraduate student outcomes. The framework also highlights key contextual variables that should be considered in broad-scale examinations of GTA TPD and potential moderators of TPD impact. It builds on the model put forth by DeChenne *et al.* (2015) but is more global in nature, positing the importance of multiple categories of relevant GTA TPD variables. The intent of this framework, then, is to support TPD practitioners in the evaluation of their programs (on their own or with assistance from an educational researcher/evaluator). At the

same time, the framework provides a structure for cross-institutional collaborations focused on the conduct, synthesis, and dissemination of research related to evidence-based biology GTA TPD practices. Crucially, this essay also offers categories of instrumentation and examples of specific instruments that GTA TPD practitioners might use in local and large-scale GTA TPD evaluation and research.

EVALUATION OF GTA TPD PROGRAMS

Given long-standing concerns that GTA TPD is inadequate (Boyer Commission on Undergraduates in the Research University, 1998; Gardner and Jones, 2011), evaluation of GTA TPD programs is critical. Such efforts can ensure TPD program effectiveness and/or the refinement of programs to support and enhance the quality of GTA teaching and, as a result, the learning outcomes of undergraduates. When discussing evaluation, the literature recognizes two overarching types of evaluations that are differentiated by their purpose: formative and summative (Patton, 2008; Yarbrough *et al.*, 2010).

In this context, at its core, *formative* evaluation endeavors to inform iteratively the quality of GTA TPD program design and implementation. As an example of a formative evaluative activity, a GTA TPD program staff member might collect data after the first of two TPD sessions to identify content GTAs would like to revisit during the third TPD session (Marbach-Ad *et al.*, 2015a). *Summative* evaluation, on the other hand, aims to summarize what happened as a result of GTA TPD program implementation. For example, researchers might seek to describe whether a GTA TPD program was associated with increased inquiry-based teaching in laboratories (e.g., Ryker and McConnell, 2014). It is also noteworthy that a particular GTA TPD program evaluation effort can serve *both* formative and summative purposes. For example, an end-of-TPD summative evaluation can inform the design of the next semester’s TPD program. Many of the key constructs in the conceptual framework proposed herein (e.g., GTA cognition, GTA teaching practice) can be examined for formative purposes, summative purposes, or both.

An expedient means of formatively evaluating a GTA TPD program is through the collection of data concerning GTA participants’ satisfaction. Measures of satisfaction capture how the respondent feels or thinks about the program. For example, an evaluation might ask GTAs who participated in a TPD program the degree to which they were satisfied with the program as a whole and/or with its particular components, activities, or processes (e.g., lectures, group activities, microteaching). Satisfaction is also commonly assessed at the end of GTA TPD programs, typically via post-TPD surveys, for summative evaluation purposes (e.g., Baumgartner, 2007; Vergara *et al.*, 2014). However, researchers have long criticized GTA satisfaction as an appropriate measure of outcomes in GTA TPD intervention research (Chism, 1998; Seymour, 2005), because the relationship between participants’ satisfaction and actual learning is equivocal at best (e.g., Gessler, 2009). Therefore, while we recognize the use of satisfaction in the GTA TPD literature, we do not include it in our evaluation and research framework, because we argue it is a fundamentally different variable than program outcomes such as GTA cognition, GTA teaching practice, and undergraduate student outcomes.

¹The Biology Teaching Assistant Project (BioTAP) and the Biology Teaching Assistant Project: Advancing Research, Synthesizing Evidence (BioTAP 2.0) are, respectively, a National Science Foundation-funded Research Coordination Network Incubator (DBI-1247938) and a National Science Foundation-funded Research Coordination Network (DBI-1539903).

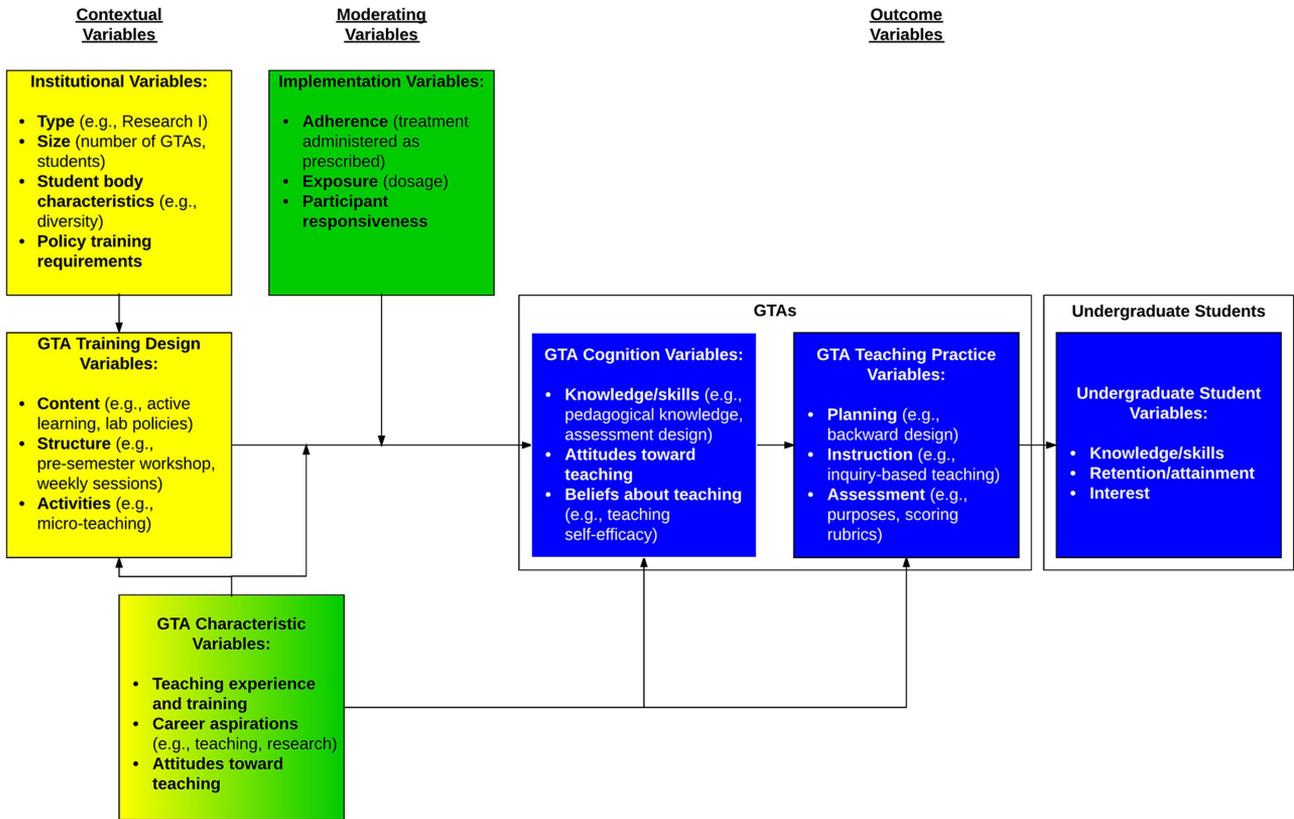


Figure 1. Framework for the relationships among GTA TPD outcome variables (blue), GTA TPD contextual variables (yellow), and GTA TPD moderating variables (green). The framework contains three main categories of outcomes at two levels, GTA and undergraduate student. These impacts (blue) are linearly (sequentially) related: GTA cognition, GTA teaching practices, and undergraduate student outcomes. GTA cognition pertains to GTAs’ knowledge, skills, attitudes, or beliefs about teaching. GTA teaching practices concerns the GTAs’ approaches to planning, instruction, and assessment. Undergraduate student outcomes centers on the knowledge and skills of GTAs’ students, as well as more distal student outcomes such as retention and graduation. The framework supposes that effective GTA TPD directly promotes changes in GTA cognition, which in turn impacts their instructional behavior (GTA teaching practices) and subsequent outcomes for undergraduates (undergraduate student outcomes). The framework contains three categories of contextual variables (yellow): GTA training design, institutional, and GTA characteristics. GTA training design variables pertain to the nature of the GTA training and are hypothesized to drive the most direct outcomes of GTA TPD: GTA cognition. Institutional and GTA characteristic variables are hypothesized to have effects on GTA training design. GTA characteristics are also hypothesized to directly impact GTA cognition (e.g., knowledge/skills, attitudes, and beliefs) and GTA teaching practices, independent of TPD. The final category of variables in the framework are moderating variables, that is, variables that impact or modify the relationship between two other variables (in this case, the relationship between GTA training design and GTA cognition). We first invoke Dane and Schneider’s (1998) implementation concepts of program adherence, exposure, and participant responsiveness as moderating variables. We also include GTA characteristics as moderators of the relationship between GTA training design and GTA cognition, given that some GTAs may change more than others during TPD. The reader will note that GTA characteristics serve as both contextual variables and moderating variables in the model. The framework is general in nature, in that it theorizes relationships between categories of variables (e.g., GTA training design and GTA cognition) rather than relationships between specific variables (e.g., GTA training length and GTA beliefs about teaching). Example specific variables in each category are not exhaustive and are provided for illustrative purposes. The framework does *not* posit that every specific variable represented within a particular variable category is associated with every specific variable represented within a related category. An arrow represents a direct impact of one category of variable on another category of variable (i.e., causal relationship).

CONCEPTUAL FRAMEWORK

Figure 1 presents our proposed conceptual framework for evaluation and research related to GTA TPD programs. The purpose of this framework is twofold: 1) to guide those who are planning to conduct empirical evaluation or research studies related to a particular GTA TPD program (at a particular department, college, or institution); and 2) to guide researchers interested in conducting, synthesizing, and disseminating large-scale and multisite research on GTA TPD.

The framework hypothesizes several categories of variables that are related to the operation of GTA TPD and is based on extant theory and research on GTA TPD (e.g., DeChenne

et al., 2015) and on broader conceptual frameworks for evaluation of professional development programs (e.g., Guskey, 2000; Wyse *et al.*, 2014). The framework contains three categories of variables: outcome variables, contextual variables, and moderating variables. In Figure 1, we provide nonexhaustive examples of key variables in each of these categories.

OUTCOME VARIABLES

An essential focus of GTA TPD program evaluation and research is on a program’s outcomes relative to its goals and objectives. The proposed framework contains three

main categories of outcomes (or impacts) that programs may measure (blue in Figure 1): GTA cognition, GTA teaching practice, and undergraduate student outcomes. Two of these outcomes pertain to GTAs and one outcome pertains to undergraduate students. Moreover, these outcome variable categories are linearly (sequentially) related, in that TPD directly impacts GTA cognition, which in turn impacts GTA teaching practice, which then impacts undergraduate student outcomes.

GTA Cognition

GTA cognition pertains to cognitive changes in GTAs' knowledge, skills, and attitudes toward or beliefs about teaching that directly result from the GTA TPD. For example, such outcomes might include GTA knowledge of active learning or inquiry-based teaching techniques or GTA teaching self-efficacy beliefs (e.g., Bowman, 2013; Connolly *et al.*, 2014). Hardré (2003) and DeChenne *et al.* (2015) reported evidence for a relationship between participation in TPD and GTA cognition (i.e., knowledge and self-efficacy).

GTA Teaching Practice

GTA cognition is linked to GTA teaching practice, which concerns GTAs' behavior related to planning, instruction, and assessment. Prior research, for example, documented improvements in GTA instructional planning and assessment practices as a result of TPD (Baumgartner, 2007; Marbach-Ad *et al.*, 2012), and Hardré (2003) linked GTA cognition (self-efficacy) and instructional practice in the context of GTA TPD. Generally, examination of GTA teaching practices will focus on teaching practices that were discussed in the GTA TPD. For example, if one of the TPD goals is to enhance inquiry-based teaching in the laboratory, part of the evaluation/research activities will focus on the level and adequacy of the implementation of inquiry-based instruction.

Undergraduate Student Outcomes

Finally, undergraduate student outcomes center on the gains in knowledge and skills made by GTAs' students, as well as more distal student outcomes such as retention and graduation. For example, one might expect that undergraduate students taught by GTAs who have received TPD would perform better on course exams. Indeed, research in K–12 settings has found that measures of teacher self-efficacy (a cognitive belief) are related to both teaching practices and student achievement (Tschannen-Moran *et al.*, 1998).

In sum, the framework uses existing literature to posit that GTA TPD directly promotes changes in participants (GTA cognition), which in turn affects their instructional behavior (GTA teaching practice), and, subsequently, outcomes for undergraduates (undergraduate student outcomes). Of these three GTA program outcomes, the first (GTA cognition) has been examined most often in GTA evaluation and research (unpublished data). Examination of the other two outcomes, GTA teaching practices and undergraduate student outcomes, is logistically more challenging and expensive, depending on the instrumentation used.

Multisite evaluation of these latter outcomes is furthermore challenging, owing to varying contextual factors (e.g., the roles of the GTAs, undergraduate course content). However, we contend that the most comprehensive and scientifically

rigorous GTA TPD evaluation should consider each of these three outcomes (and employ true experimental or quasi-experimental designs in order to confidently assess whether changes in these variables are due to GTA TPD rather than other variables). For those just starting evaluations of their programs, it would be reasonable to start with the most proximal GTA TPD outcome (i.e., GTA cognition), and once those effects are established, proceed to the evaluation of more distal outcomes (i.e., GTA teaching practice, then undergraduate student outcomes). In a later section, we offer practical guidance on how to elicit evidence of various GTA TPD outcomes.

CONTEXTUAL VARIABLES

As mentioned earlier, one limitation of the GTA TPD literature is that it largely comprises small-scale studies, each focused on a particular GTA TPD program at a particular institution. As such, the literature lacks large-scale, multi-institutional studies with the potential to compare the effectiveness of GTA TPD programs that systematically vary in their design, allowing for identification of evidence-based practices (Hardré and Chen, 2005; Hardré and Burris, 2012). The challenge of drawing comparisons among different TPD designs from the extant literature is furthermore compounded by considerable variation among institutional contextual factors (Schussler *et al.*, 2015). For example, findings from DeChenne *et al.*'s 2015 study underscored the importance of accounting for contextual variables, such as departmental teaching climate, when studying GTA TPD programs. Therefore, what might constitute an "effective" GTA TPD program for one institution/department might not be effective for another.

Given the generally fragmented nature of the body of GTA TPD literature, our framework considers three categories of contextual variables (in yellow in Figure 1): GTA training design variables, institutional variables, and GTA characteristic variables. These elements of the framework are intended for researchers interested in conducting research on GTA TPD program design and impact in diverse contexts. The categories also offer guidance for the types of information that individuals who publish outcomes of single GTA TPD programs should provide to situate the context of their program for their readers.

GTA Training Design Variables

The design of GTA TPD varies widely, in terms of training program content, structure, and activities (e.g., Hardré and Burris, 2012; DeChenne *et al.*, 2015). In the proposed conceptual framework, GTA TPD training design variables are hypothesized to drive the most direct outcome of GTA TPD—GTA cognition. As noted earlier, GTA cognition ultimately affects GTA teaching practices and, in turn, undergraduate student outcomes. Notably, K–12 professional development designs that translate to teacher and/or student outcomes are marked by a focus on subject matter content, coherence with teachers needs (content), an extended duration (structure), and opportunities for active learning (activities; Garet *et al.*, 2001; Desimone *et al.*, 2002).

There is also some published literature on the design of GTA TPD in terms of its content, structure, and activities. With respect to TPD content, TPD programs described in the literature have covered topics such as assessment, pedagogical

methods, policies and procedures, and multicultural issues (e.g., Luft *et al.*, 2004; Prieto *et al.*, 2007). In terms of TPD structure, GTA TPD programs discussed in the literature often take the form of a onetime workshop (Gardner and Jones, 2011; Schussler *et al.*, 2015); other designs or design elements such as GTA mentoring or receipt of teaching feedback are much more rare (Austin, 2002; DeChenne *et al.*, 2012). Relative to TPD activities, prior research has examined activities such as microteaching (Gilreath and Slater, 1994) and teaching skits (Marbach-Ad *et al.*, 2012). Published GTA TPD research even offers evidence for positive effects of some TPD design variables on GTA cognition, for example, the effect of training length on GTA self-efficacy related to teaching (e.g., Prieto and Meyers, 1999; Hardré, 2003; Young and Bippus, 2008).

Institutional Variables

The proposed conceptual framework incorporates institutional variables such as institutional type, size, student body characteristics, and policy training requirements. Institutional variables are hypothesized to have effects on the nature of the TPD provided to GTAs, although concrete empirical evidence for this is sparse and often indirect (Park, 2004; Lattuca *et al.*, 2014). As noted previously in the literature, TPD content and structure vary considerably from institution to institution and across different institutional contexts (Marbach-Ad *et al.*, 2015a; Schussler *et al.*, 2015), including institutional cultural differences with respect to how teaching is viewed (Serow *et al.*, 2002). Along these lines, Rushin *et al.* (1997) found differences between master's degree- and doctoral degree-granting institutions in terms of the GTA TPD models used. In their study, doctoral degree-granting institutions were more likely to employ a preacademic-year workshop, whereas master's degree-granting institutions were more likely to employ individualized GTA training led by the course professor. While the Rushin *et al.* (1997) findings are suggestive of a key role of institutional type (e.g., research-intensive university) in shaping GTA TPD design, (arguably) other variables are important as well. For example, the typical teaching role of the GTA at a particular institution (e.g., facilitating discussion sessions, coordinating laboratory sessions, or grading assignments) and the presence of a faculty development unit (e.g., Center for Teaching and Learning; Marbach-Ad *et al.*, 2015a) might also affect the design of a GTA TPD program, specifically its duration, structure, or content.

GTA Characteristic Variables

Finally, a third category of contextual variable in the proposed framework is GTA characteristics. The extant literature highlights considerable variation among GTAs both across and within institutions (Addy and Blanchard, 2010; DeChenne *et al.*, 2015). In particular, GTAs differ with respect to their prior teaching experiences and training (Prieto and Altmaier, 1994), relative prioritization of teaching versus research, aspirations for careers involving teaching (Nyquist *et al.*, 1999; Brownell and Tanner, 2012; Sauermann and Roach, 2012), and attitudes toward teaching (Tanner and Allen, 2006). In the framework, GTA characteristics are posited to impact the nature of the TPD provided to GTAs (i.e., TPD training design). A GTA population with varying levels of teaching experience, for example, might necessitate a differentiated TPD program (Austin, 2002; Schussler

et al., 2015). As another example, Marbach-Ad *et al.* (2015a,b) reported on three different TPD programs at their research-intensive university based on students' career aspirations. Thus, GTA characteristics can impact GTA training design variables such as duration (e.g., a longer course for those with teaching aspirations), structure (e.g., type and amount of homework assignments), and activities (e.g., developing a teaching philosophy and portfolio).

GTA characteristics are also hypothesized to directly impact GTA cognition (e.g., knowledge/skills, attitudes, and beliefs) and GTA teaching practice, independent of TPD. Prior research indicates large GTA-to-GTA variation even after participation in TPD (e.g., Bond-Robinson and Rodrigues, 2006; Addy and Blanchard, 2010), implying that other GTA-level variables besides training (i.e., GTA characteristics) impact GTA teaching cognition and practice. For example, research has shown a relationship between GTA level of teaching experience and teaching self-efficacy (Prieto and Altmaier, 1994) and that diverse GTA beliefs and prior experience impact their teaching practices (Addy and Blanchard, 2010). Moreover, these GTA characteristics should be considered in the interpretation of GTA evaluation findings. For instance, when comparing the effectiveness of two programs, one needs to consider the GTAs' input characteristics (e.g., prior TPD experience), because differential knowledge after training might be caused by those initial differences rather than differences in program effectiveness.

MODERATING VARIABLES

The proposed framework also includes two categories of moderator variables (in green in Figure 1): implementation variables and GTA characteristic variables. These variables are termed moderating variables, because they may impact or modify the relationship between two other variables (in this case the relationship between GTA training design and GTA cognition).

Implementation Variables

The success of any program in attaining its intended outcomes depends not only on the TPD program's intended design but also on how well it was implemented. Evaluation of program implementation involves examining the degree to which a GTA TPD program was enacted with fidelity, that is, as intended. We therefore also included implementation variables (i.e., Dane and Schneider's [1998] concepts of program adherence, exposure, and participant responsiveness) in the proposed framework as moderators of the relationship between TPD training design variables and GTA cognition outcomes. If null effects of GTA TPD are observed, implementation variable data (e.g., the number of times each GTA met with his or her mentor) can assist program staff in discerning whether effects were not observed because of a poorly designed program (i.e., theory failure) or poor program implementation (i.e., implementation failure). Examples of implementation variables that might be assessed include the GTAs' degree of participation/engagement in the TPD program, the degree to which all intended content was given sufficient attention during a TPD session, or whether protocols for collaborative learning activities for GTAs were followed appropriately. This information is

often collected through the use of external observers during the program, but it could also be collected from GTAs' self-reports during end-of-semester survey or interviews. For example, Marbach-Ad *et al.* (2015b) used an external evaluator to interview and survey GTAs who participated in a teaching certificate program. The design of the program included a component in which GTAs were observed and mentored by faculty members. GTAs reported that this component was not well implemented, mainly due to lack of faculty cooperation, suggesting that poor implementation might have moderated the relationship between GTA training design variables and TPD outcome variables.

GTA Characteristic Variables

The proposed framework also includes GTA characteristics as moderators of the relationship between GTA training design and GTA cognition. Simply put, this aspect of the framework pertains to possible *differential effects* of TPD on GTA cognition. Several studies have investigated the relationship between GTA prior teaching experience (e.g., number of semesters taught) and self-efficacy belief and attitudinal gains observed during TPD (e.g., Addy and Blanchard, 2010; DeChenne *et al.*, 2015). Other work has shown that GTAs' prior teaching experiences or knowledge is related to knowledge gains during TPD (Marbach-Ad *et al.*, 2012) and to the implementation of TPD content during GTAs' classroom practice (French and Russell, 2002; Hardré and Chen, 2005).

APPLYING THE CONCEPTUAL FRAMEWORK

Implicit in each of the proposed framework's directional paths are various evaluation and research questions/hypotheses about how GTA TPD programs operate to produce GTA and student outcomes and about the role of contextual variables in GTA TPD. These include 1) system-level questions, such as how institutional variables affect GTA TPD training design; 2) TPD program-level questions, such as how different TPD training designs translate to direct effects on GTAs' cognition and indirect effects on GTA teaching practices and undergraduate student outcomes; and 3) individual GTA-level questions, such as how GTAs with different characteristics respond differently to TPD. Through its inclusion of contextual variables, the framework also provides a structure for both small-scale, local (single program) evaluation and large-scale, cross-institutional GTA TPD research (looking across programs to identify evidence-based practices).

Even if a researcher is studying only a single, local GTA program and its outcomes, in reporting his or her findings, he or she should describe the program's design, implementation, and relevant contextual variables in terms of the institution and participating GTAs. This will afford the community more information to use in synthesizing findings across individual studies. At the same time, such information can help a reader weigh the applicability of a given study's findings to his or her local context. For example, findings derived from a TPD program for GTAs who want to enter industrial fields may not necessarily apply to a TPD program for GTAs who hope to attain positions at small, liberal arts colleges focused chiefly on teaching.

It bears noting that the framework is general in nature, in that it theorizes relationships between categories of variables

(e.g., GTA training design and GTA cognition) rather than relationships between specific variables (e.g., GTA training length and GTA beliefs about teaching). Specific variables are provided for illustrative purposes. The framework does *not* posit that every specific variable represented within a particular variable category (a box in Figure 1) is associated with every specific variable represented within a related category. Continued research is needed to empirically elicit the relationships between specific variables in each general category.

While the proposed framework is inclusive of several key categories of variables, it is not exhaustive in the sense that all determinants of GTA TPD design, implementation, and outcomes are included. For instance, in addition to institutional and GTA characteristic variables, TPD program staff variables (e.g., knowledge, beliefs) might also impact GTA TPD design. As additional evidence accumulates, other welcomed extensions to the general framework described here may include mediators or moderators of particular linkages (e.g., student population moderating the impact of certain GTA classroom practices on student achievement, or GTA curricular autonomy moderating the impact of GTA cognition on GTA practice). We hope that future research validates this framework and refines it as needed on the basis of evidence.

A PRACTICAL GUIDE FOR EVALUATING GTA TPD PROGRAMS

In Table 1, we offer practical guidance for those who wish to conduct evaluations of their own GTA TPD programs. In particular, we discuss how to elicit evidence of the three GTA TPD outcome variables implicit in the proposed conceptual framework (GTA cognition, GTA teaching practice, and undergraduate student outcomes). For each of these three GTA TPD outcomes, we enumerate some guiding evaluation questions, possible categories of instrumentation (e.g., surveys, tests), and examples of specific existing instruments (e.g., Smith *et al.*'s [2008] Genetics Concept Assessment)² that can be used in evaluation efforts. We caution that the specific instruments we reference are provided as examples but may not be the most appropriate for any given program.

In addition, we recommend that researchers interested in assessing GTA TPD outcomes across programs and institutions collect data concerning other variables in the framework besides outcomes (e.g., GTA characteristic variables, implementation variables), as they might be important covariates. To the best of the authors' knowledge, however, there are no known and broadly applicable instruments designed to elicit evidence of these other key categories of framework variables. The development of such instruments indeed constitutes a potential target of future scholarship. In particular, instruments could be designed to gather evidence concerning both GTA TPD contextual variables (i.e., institutional variables, GTA training design variables, and GTA characteristics) and implementation variables. These instruments could be administered to either TPD program staff or participating GTAs for data-collection purposes in the context of large-scale research.

²We refer the reader to Reeves and Marbach-Ad (2016) for information about how to select high-quality instruments.

Table 1. Possible instrumentation for collection of evidence concerning GTA TPD outcomes^a

GTA TPD outcome variable category	Specific GTA TPD program outcome variable	Example research/evaluation question	Possible categories of instrumentation	Example of existing instruments
GTA cognition	Knowledge/skills	Did participants acquire the intended knowledge and skills (in terms of pedagogy, assessment, and curriculum)?	Content tests; surveys	Pedagogy of Science Teaching Tests (Cobern <i>et al.</i> , 2014) ^b
	Attitudes toward teaching	Was the GTA TPD associated with changes in participants' valuing student-centered approaches?	Surveys; interviews	Survey of Teaching Beliefs and Practices (STEP; Marbach-Ad <i>et al.</i> , 2014)
	Beliefs about teaching	Was the GTAs' teaching self-efficacy increased following the TPD?	Surveys	Science Teaching Efficacy Belief Instrument (Smolleck <i>et al.</i> , 2006)
GTA teaching practices	Planning	Do GTAs who participated in the TPD use backward design to plan their classes?	Artifacts (e.g., lesson plans, assessments); surveys; interviews; focus groups	— ^c
	Instruction	Do GTAs who participated in the TPD spend more time interacting with students?	Surveys; student evaluations of instruction	End-of-semester student evaluations used in Marbach-Ad <i>et al.</i> (2012)
			Teaching observations	Reformed Teaching Observation Protocol (Piburn <i>et al.</i> , 2000); Classroom Observation Protocol for Undergraduate STEM (Smith <i>et al.</i> , 2013)
	Assessment	Following professional development, are GTA assessments more closely aligned to course learning outcomes?	Artifacts (e.g., assessments)	Rubric for examining objective-assessment alignment (Wyse <i>et al.</i> , 2014)
Undergraduate student outcomes	Knowledge/skills	Do students taught by TPD-trained GTAs demonstrate improved knowledge and skills?	Content tests/concept inventories; surveys; interviews; artifacts (e.g., student work)	Test of Scientific Literacy Skills (Gormally <i>et al.</i> , 2012); Genetics Concept Assessment (Smith <i>et al.</i> , 2008)
	Retention/attainment	Are students taught by GTAs who participated in TPD more likely to be retained in the biology major and graduate?	Official institutional and academic transcript data	Time to degree; first- to second-year retention; graduation
	Interest	Do biology students taught by TPD-trained GTAs demonstrate greater interest in learning biology?	Surveys; interviews; focus groups	Colorado Learning Attitudes about Science Survey (Semsar <i>et al.</i> , 2011)

^aFor each of three general GTA TPD outcomes (i.e., GTA cognition, GTA teaching practice, and undergraduate student outcomes) and nine corresponding specific outcomes (e.g., GTA knowledge/skills, GTA planning, and undergraduate student retention), the table outlines example research/evaluation questions that might be asked by TPD program staff or researchers and possible categories of instrumentation and example specific instruments that might be used.

^bThis instrument is intended for formative use in grades K–8 science teaching and is included for illustrative purposes only.

^cTo the best of the authors' knowledge, there are no currently published instruments to systematically elicit evidence of backward design planning, which is a fruitful area for future research.

CONCLUSION

The proposed conceptual framework explicated in this article was created with two purposes in mind: 1) to offer a

guide for the evaluation of GTA TPD programs at individual institutions and 2) to offer a framework for how institutions can begin to coordinate evaluation and research efforts in order to build evidence-based biology GTA TPD practices.

Although we make no claims that the framework is comprehensive and complete, we believe that it can serve as a starting point for dialogue among practitioners and researchers about how to conduct large-scale, systemic research. The results generated from these coordinated efforts will, in turn, provide biology GTA TPD practitioners with empirical data that can be used to improve GTA teaching practices and undergraduate outcomes at their institutions.

For those who lead GTA TPD programs, we hope the conceptual framework provides insights to improve local programmatic evaluation practices. Program practitioners may realize, for example, that they have only been evaluating GTA satisfaction with their programs. In this case, they may use the information in this framework to begin to assess bona fide outcomes such as GTA cognition (e.g., knowledge of inquiry-based teaching methods). The conceptual framework could potentially be used as justification to department chairs or other administrators to provide additional resources to conduct these types of studies, particularly if the connection to undergraduate student outcomes is made clear.

The framework also provides practitioners with flexibility, a key factor given the multiple contexts in which biology GTA TPD is enacted. Practitioners may realize that they are only interested in probing the impact of GTA TPD enactment on only one particular outcome variable. Identifying the questions practitioners may wish to pursue and the resources they have available to pursue those questions will help them to build an evaluation plan that fits their particular needs. The example evaluation/research questions in Table 1 should guide those practitioners to identify specific questions and begin to think about the methods (instrumentation) they could use to assess them.

Finally, the conceptual framework proposes contextual variables that should be documented during dissemination of evaluation/research results for the purposes of more systematically comparing programmatic results across institutions. Ideally, researchers and practitioners at different institutions would coordinate their programmatic efforts as part of a designed research study, but we recognize that this may not be possible in practice because of the contextual variability in which programs at different institutions are enacted. Instead, collecting similar contextual variables and using some of the same instruments to measure program outcomes will allow institutions to compare their results and begin to hypothesize practices that may be beneficial at either particular types of institutions or at institutions more broadly. Comparisons such as these will greatly improve the ability of the field to move forward with identifying practices that maximize the impacts of TPD on GTAs and undergraduates (Schussler *et al.*, 2015).

Given the profound impact that biology GTAs have on teaching at undergraduate institutions, enhancing GTA TPD as a means to improve GTA teaching practices and undergraduate learning outcomes should be a priority for institutions of higher education. Particularly for gateway science courses, improved GTA teaching practices may be a key lever to improve degree attainment in the sciences (e.g., O'Neal *et al.*, 2007). As these GTAs move through their graduate programs, many will go on to become members of the professoriate; thus, providing effective biology GTA TPD programs may be one critical link to fully envisioning the promise of evidence-based teaching practices in biology courses.

REFERENCES

- Abbott RD, Wulff DH, Szego CK (1989). Review of research on TA training. *New Dir Teach Learn* 39, 111–124.
- Addy TM, Blanchard MR (2010). The problem with reform from the bottom up: Instructional practises and teacher beliefs of graduate teaching assistants following a reform-minded university teacher certificate programme. *Int J Sci Educ* 32, 1045–1071.
- Austin A (2002). Preparing the next generation of faculty: graduate school as socialization to the academic career. *J High Educ* 73, 94–122.
- Baumgartner E (2007). A professional development teaching course for science graduate students. *J Coll Sci Teach* 36, 16–21.
- Bond-Robinson J, Rodrigues RAB (2006). Catalyzing graduate teaching assistants' laboratory teaching through design research. *J Chem Educ* 83, 313–323.
- Bowman JS (2013). Graduate student teaching development: evaluating the effectiveness of training in relation to graduate student characteristics. *Can J High Educ* 43, 100–114.
- Boyer Commission on Undergraduates in the Research University (1998). *Reinventing Undergraduate Education: A Blueprint for America's Research Universities*, Stony Brook: State University of New York.
- Brownell S, Tanner KD (2012). Barriers to faculty pedagogical change: lack of training, time, incentives, and ... tensions with professional identity? *CBE Life Sci Educ* 11, 339–346.
- Chism NVN (1998). Evaluating TA programs. In: *The Professional Development of Graduate Teaching Assistants*, ed. M Marincovich, J Prostok, and F Stout, Bolton, MA: Anker, 249–262.
- Cobern WW, Schuster D, Adams B, Skjold BA, Muğaloğlu EZ, Bentz A, Sparks K (2014). Pedagogy of science teaching tests: formative assessments of science teaching orientations. *Int J Sci Educ* 36, 2265–2288.
- Connolly MR, Lee Y-G, Savoy JN, Hill L, Grettie J, Vandenberg J, Austin AE (2014). *The Longitudinal Study of Future STEM Scholars: An Overview*, Madison: Wisconsin Center for Education Research.
- Dane AV, Schneider BH (1998). Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clin Psychol Rev* 18, 23–45.
- DeChenne SE, Enochs LG, Needham M (2012). Science, technology, engineering, and mathematics graduate teaching assistants teaching self-efficacy. *J Scholarship Teach Learn* 12, 102–123.
- DeChenne SE, Koziol N, Needham M, Enochs L (2015). Modeling sources of teaching self-efficacy for science, technology, engineering, and mathematics graduate teaching assistants. *CBE Life Sci Educ* 14, ar32.
- Desimone LM, Porter AC, Garet MS, Yoon KS, Birman BF (2002). Effects of professional development on teachers' instruction: results from a three-year longitudinal study. *Educ Eval Policy Anal* 24, 81–112.
- Dolan E (2016). Biology education research 2.0. *CBE Life Sci Educ* 14, ed1.
- French D, Russell C (2002). Do graduate teaching assistants benefit from teaching inquiry-based laboratories? *BioScience* 52, 1036–1041.
- Gardner GE, Jones MG (2011). Pedagogical preparation of science graduate teaching assistant: challenges and implications. *Sci Educ* 20, 31–41.
- Garet M, Porter A, Desimone L, Birman B, Yoon K (2001). What makes professional development effective? Results from a national sample of teachers. *Am Educ Res J* 38, 915–945.
- Gessler M (2009). The correlation of participant satisfaction, learning success and learning transfer: an empirical investigation of correlation assumptions in Kirkpatrick's four-level model. *Int J Management Educ* 3, 346–358.

- Gilreath JA, Slater TF (1994). Training graduate teaching assistants to be better undergraduate physics educators. *Phys Educ* 29, 200.
- Gormally C, Brickman P, Lutz M (2012). Developing a test of scientific literacy skills (TOSLS): measuring undergraduates' evaluation of scientific information and arguments. *CBE Life Sci Educ* 11, 364–377.
- Guskey TR (2000). *Evaluating Professional Development*, Thousand Oaks, CA: Corwin.
- Hardré PL (2003). The effects of instructional training on university teaching assistants. *Perform Improv Q* 16, 23–39.
- Hardré PL, Burris AO (2012). What contributes to teaching assistant development: differential responses to key design features. *Instr Sci* 40, 93–118.
- Hardré PL, Chen C (2005). A case study analysis of the role of instructional design in the development of teaching expertise. *Perform Improv Q* 18, 34–58.
- Lattuca LR, Bergom I, Knight DB (2014). Professional development, departmental contexts, and use of instructional strategies. *J Eng Educ* 103, 549–572.
- Luft JA, Kurdziel JP, Roehrig GH, Turner J (2004). Growing a garden without water: graduate teaching assistants in introductory science laboratories at a doctoral/research university. *J Res Sci Teach* 41, 211–233.
- Marbach-Ad G, Egan L, Thompson KV (2015a). *A Discipline-Based Teaching and Learning Center: A Model for Professional Development*, New York: Springer.
- Marbach-Ad G, Katz P, Thompson KV (2015b). A disciplinary teaching certificate program for science graduate students. *J Centers Teach Learn* 7, 24–52.
- Marbach-Ad G, Schaefer KL, Kumi BC, Friedman LA, Thompson KV, Doyle MP (2012). Development and evaluation of a prep course for chemistry graduate teaching assistants at a research university. *J Chem Educ* 89, 865–872.
- Marbach-Ad G, Schaefer KL, Orgler M, Thompson KV (2014). Science teaching beliefs and reported approaches within a research university: Perspectives from faculty, graduate students, and undergraduates. *Int J Teach Learn High Educ* 26(2).
- Nyquist JD, Manning L, Wulff DH, Austin AE, Sprague J, Fraser PK, Calcagno C, Woodford B (1999). On the road to becoming a professor: the graduate student experience. *Change* 31, 18–27.
- O'Neal C, Wright M, Cook C, Perorazio T, Purkiss J (2007). The impact of teaching assistants on student retention in the sciences: lessons for TA training. *J Coll Sci Teach* 36, 24–29.
- Park C (2004). The graduate teaching assistant (GTA): Lessons from the North American experience. *Teach High Educ* 9, 349–361.
- Patton MQ (2008). *Utilization-Focused Evaluation*, 4th ed., Thousand Oaks, CA: Sage.
- Piburn M, Sawada D, Turley J, Falconer K, Benford R, Bloom I, Judson E (2000). *Reformed teaching observation protocol (RTOP) reference manual*, Technical Report No. IN00–3, Tempe: Arizona Collaborative for Excellence in the Preparation of Teachers.
- President's Council of Advisors on Science and Technology (2012). *Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics*, Washington, DC: U.S. Government Office of Science and Technology.
- Prieto LR, Altmaier EM (1994). The relationship of prior training and previous teaching experience to self-efficacy among graduate teaching assistants. *Res High Educ* 35, 481–497.
- Prieto LR, Meyers SA (1999). The effects of training and supervision on the self-efficacy of psychology graduate teaching assistants. *Teach Psychol* 26, 264–266.
- Prieto LR, Yamokoski CA, Meyers SA (2007). Teaching assistant training and supervision: an examination of optimal delivery modes and skill emphases. *J Faculty Dev* 21, 33–43.
- Reeves TD, Marbach-Ad G (2016). Contemporary test validity in theory and practice: a primer for discipline-based education researchers. *CBE Life Sci Educ* 15, rm1.
- Rushin JW, DeSaix J, Lumsden A, Streubel DP, Summers G, Bernson C (1997). Graduate teaching assistant training—a basis for improvement of college biology teaching and faculty development? *Am Biol Teach* 59, 86–90.
- Ryker K, McConnell D (2014). Can graduate teaching assistants teach inquiry-based geology labs effectively? *J Coll Sci Teach* 44, 56–63.
- Sauermann H, Roach M (2012). Science PhD career preferences: levels, changes, and advisor encouragement. *PLoS One* 7, 777–780.
- Schussler EE, Read Q, Marbach-Ad G, Miller K, Ferzli M (2015). Preparing biology graduate teaching assistants for their roles as instructors: an assessment of institutional approaches. *CBE Life Sci Educ* 14, ar31.
- Semsar K, Knight JK, Birol G, Smith MK (2011). The Colorado Learning Attitudes about Science Survey (CLASS) for use in biology. *CBE Life Sci Educ* 10, 268–278.
- Serow RC, Van Dyk PB, McComb EM, Harrold AT (2002). Cultures of undergraduate teaching at research universities. *Innov High Educ* 27, 25–37.
- Seymour E (2005). *Partners in Innovation: Teaching Assistants in College Science Courses*, Lanham, MD: Rowman & Littlefield.
- Seymour E, Hewitt NM (1997). *Talking about Leaving: Why Undergraduates Leave the Sciences*, Boulder, CO: Westview.
- Seymour E, Melton G, Wiese DJ, Pedersen-Gallegos L (2005). *Partners in Innovation: Teaching Assistants in College Science Courses*, Boulder, CO: Rowman & Littlefield.
- Smith MK, Jones FH, Gilbert SL, Wieman CE (2013). The Classroom Observation Protocol for Undergraduate STEM (COPUS): a new instrument to characterize university STEM classroom practices. *CBE Life Sci Educ* 12, 618–627.
- Smith MK, Wood WB, Knight JK (2008). The Genetics Concept Assessment: a new concept inventory for gauging student understanding of genetics. *CBE Life Sci Educ* 7, 422–430.
- Smolleck LD, Zembal-Saul C, Yoder EP (2006). The development and validation of an instrument to measure preservice teachers' self-efficacy in regard to the teaching of science as inquiry. *J Sci Teach Educ* 17, 137–163.
- Sundberg MD, Armstrong JE, Wischusen EW (2005). A reappraisal of the status of introductory biology laboratory education in US colleges and universities. *Am Biol Teach* 67, 525–529.
- Tanner KD, Allen D (2006). Approaches to biology teaching and learning: on integrating pedagogical training into the graduate experiences of future science faculty. *Cell Biol Educ* 5, 1–6.
- Tschannen-Moran M, Hoy AW, Hoy WK (1998). Teacher efficacy: its meaning and measure. *Rev Educ Res* 68, 202–248.
- Vergara CE, Urban-Lurain M, Campa H, III, Cheruvelil KS, Ebert-May D, Fata-Hartley C, Johnston K (2014). FAST—future academic scholars in teaching: a high-engagement development program for future STEM faculty. *Innov High Educ* 39, 93–107.
- Wyse SA, Long TM, Ebert-May D (2014). Teaching assistant professional development in biology: designed for and driven by multidimensional data. *CBE Life Sci Educ* 13, 212–223.
- Yarbrough DB, Shulha LM, Hopson RK, Caruthers FA (2010). *The Program Evaluation standards: A Guide for Evaluators and Evaluation Users*, 3rd ed., Los Angeles, CA: Sage.
- Young SL, Bippus AM (2008). Assessment of graduate teaching assistant (GTA) training: a case study of a training program and its impact on GTAs. *Comm Teach* 22, 116–129.