Testing the Stability of Items in a Survey to Measure Relative Instructional Priorities Among Graduate Teaching Assistants

Eliza Gallagher
Clemson University
egallag@clemson.edu

Tony Nguyen
Clemson University
ttn@g.clemson.edu

Aubrie L. Pfirman
Clemson University
aubriep@g.clemson.edu

Khushikumari Patel
Clemson University
khuship@g.clemson.edu

The results presented in this paper are part of a larger mixed-methods study examining relative instructional priorities among mathematics graduate teaching assistants (MGTAs). In this paper we share some early results and observations from a limited test-retest analysis of a pilot survey administered to MGTAs in two large public institutions in the Southeast United States. This is not intended to be an exhaustive statistical analysis of the pilot survey results or test-retest analysis. Instead, we focus on specific items to serve as a lens for better understanding the complexity of the choices MGTAs make in instructional settings.

Keywords: Graduate student teacher identity, Instructional choices, Survey validation

Introduction and Background

The notion of *teacher identity*, the extent to which one identifies as a teacher, informs a large body of literature related to professional preparation and development of secondary mathematics teachers (Ward, Nolen, & Horn, 2011; Beauchamp & Thomas, 2011; Ball & Bass, 2004; Beauchamp & Thomas, 2009; Flores & Day, 2006; Hamman, Gosselin, Romano, & Bunuan, 2010; Heyd-Metzuyanim & Sfard, 2012; Horn, Nolen, Ward, & Campbell, 2008; Lasky, 2005; Sexton, 2008; VanZoest & Bohl, 2005; Hodges & Cady, 2012). Although novice secondary mathematics teachers and new mathematics graduate teaching assistants (MGTAs) share many characteristics such as undergraduate mathematics coursework, stage of life, and assumption of new professional teaching duties, the frameworks developed for understanding secondary teacher identity do not translate easily to work with MGTAs (Gallagher, 2016).

In this paper, we build on previous work that explored the ways in which experienced secondary teachers saw themselves as subject matter, didactical, and/or pedagogical experts, and then assigned those teachers locations within a "personal knowledge triangle" to represent the relative weight given to each of those types of expertise (Beijaard, Verloop, & Vermunt, 2000). A multi-year multiple case study following the development of teacher identity and instructional practice among four MGTAs led the lead author to develop a modified framework to represent the types of expertise and actions valued by MGTAs. For that population, instructional decision-making and value judgments were stratified into three zones: subject-matter concerns, class-level structures, and individual-level needs. For a more complete description of the boundaries of each of those zones, we refer the reader to (Gallagher, 2016).

We refer to the collective framework of these choices as Relative Instructional Priorities (RIP) and visualize each individual MGTA's balance as being situated within a triangle, where the vertices represent the components of the RIP (individual needs, subject-matter knowledge, and class-level considerations). *Individual needs* items are focused on the students as individuals

and prompt for instructor prioritization of individualized instruction, supporting diversity, and awareness of campus resources for students. Items in the *subject-matter knowledge* category focus on content and curriculum, including content mastery and preparation for subsequent courses. *Class-level consideration* items include the use of technology in the classroom, physical classroom arrangement, choice of instructional activity, and pacing of a class session.

The relative position of a point along an edge of the triangle represents the weight given to that RIP component. For instance, in Figure 1, the point along the I-S edge is closer to vertex I meaning that this respondent prioritizes the individual needs of students over subject-matter concerns while the point along the C-S edge reveals that the respondent is relatively balanced when choosing between class-level versus subject-matter. Triangulating the three edge points gives the respondent's overall RIP placement, the "star" inside the triangle.

Survey Development

Our goal with the RIP survey was to have students make choices between pairs of statements that were anchored strongly at one of the three vertices. For example, a descriptor such as "knowing how the course content is used in subsequent courses" would be anchored at S, while "having a range of strategies to encourage group discussion" would be anchored at C and "knowing what resources are available for a student who is upset about a non-academic issue" would be anchored at I. On the other hand, a descriptor such as *'adapting class activities based on students' prerequisite knowledge'* contains aspects of subject-matter knowledge (S), individual needs (I), and class-level decision-making (C).

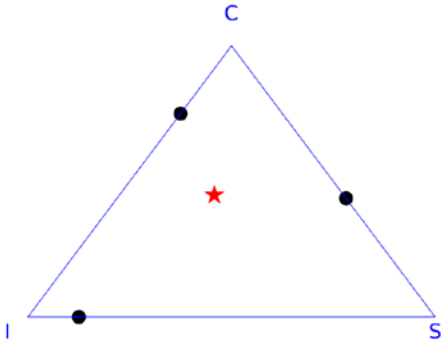

Figure 1. Visualization of an Example RIP

For our first implementation, we drew on our prior multiple case study interview and survey data to draft items capturing different aspects of each category, as shown in Table 1. Each item in a category was then paired with every item in the other two categories, generating a total of 48 pairings; all possible cross-category pairings were present and each pairing was prefaced with the prompt, *'For each item, circle the statement that is more important to you in your teaching.'*

Table 1. Early Version of RIP Items

Individual (I)
I1 Knowing how to help underrepresented students feel welcome in your class
I2 Knowing how to adapt instruction for individual students
I3 Having a range of strategies for encouraging a struggling student
I4 Being able to support the emotional needs of your students

Class-level (C)

C1 Having a range of strategies for evaluating class outcomes
C2 Knowing how to manage a classroom setting
C3 Having a range of instructional options for structuring a class session
C4 Knowing how to plan a class

Subject-matter (S)

S1 Knowing how to apply the mathematical content in other contexts
S2 Being able to answer questions about the content
S3 Being able to identify mathematics mistakes in student work
S4 Knowing the content of the course

This early paper-based version of the survey was administered to a group of 21 first-year MGTAs in a mathematics professional development course near the beginning of the semester at one large, public, southeastern university. In this version, students circled the item from each pair that was more important to them. Following administration of the paper version, we engaged in a focus group discussion to understand the students' perceptions of the individual prompts. In this discussion, it became apparent that some of these prompts were too broadly phrased, as some students saw an overlap of items. For example, one participant interpreted the item, *'Knowing how to help a struggling student,'* as an S item, as evidenced by her statement, "Knowing your content *is* helping your struggling student, they're the same thing!" Similar statements from other participants allowed us to refine the phrasing of prompts within the C, S, and I categories.

For the second version of the survey, which is the focus of this paper, we used results from the paper-based pilot administration and focus group discussion to generate a bank of 120 items, with each of the four authors generating ten potential items within each category. Each of the potential items was then classified by the other three members of the team as either C, S, or I. Items that did not reach unanimous agreement on classification were discarded or modified until they reached consensus of classification. That winnowing process resulted in 21 C items, 28 I items, and 20 S items. From those, each member of the team selected the eight that he or she felt best captured the range of archetypal prompts for the category. Those votes resulted in the selection of four prompts for each category.

We then created 27 pairings distributed evenly among C-S, C-I, and I-S pairings. Each of the four prompts within a category occurred 2-3 times and was never paired twice against the same prompt from another category. The nine prompts within each pairing were evenly distributed among three groupings to explore whether participants differentiated between preferences, importance, and appeal in choosing between prompts. Each group of nine pairings was prefaced by one of: *'Which is more important to you?' 'Which scenario appeals to you more?'* and *'Select the one you prefer.'* To avoid biasing generation of new archetypal prompts in the discussion portion of the session, we purposefully do not include here a list of all twelve prompts we selected. Some specific prompts are discussed below, and the full set is available from the corresponding author.

Survey Administration and Retest

We used Qualtrics® to develop, edit, and distribute the survey. At the beginning of the survey, participants were asked to enter a participant ID or code that was generated to protect their anonymity. Additional prompts after the item-selection blocked pairs asked demographic questions about their undergraduate and graduate majors, languages spoken, year of study, teaching background, etc. Participants were given the option of entering into a random drawing for a Visa® gift card incentive, and a separate option to enter their email address to indicate willingness to participate in a follow-up interview.

Thirty-six MGTAs from two large southeastern universities responded to the survey, out of 88 who were invited to participate (40.1% response rate); there was an incentive of two $25 gift card in a random drawing from among respondents. From the responses received, we eliminated those who did not complete the survey and those who completed it in under four minutes, leaving us with 28 responses (31.8% valid response rate). The participants at one university (n=25) could voluntarily enroll in one of several course-specific professional development seminar courses, while participants at the other university (n=3) were required to enroll in a general professional development course. We conducted single-session observations of each semester-long seminar course; each seminar leader indicated that the session we observed was typical of the course. From those observations, the seminars ranged from a focus on instructional planning, to a focus on content mastery, to a focus on planning instruction around common student misconceptions. Participants were a mix of domestic and international, and traditional and non-traditional. All were in their first, second, or third year of graduate school. They held a range of teaching duties including teaching assistant, classroom assistant, and instructor of record.

We sent follow-up emails to all 28 participants six weeks following their original survey completion, inviting each to retake the survey for test-retest comparison. Fifteen participants (53.57%) completed the re-test. It is worth noting that we had 12 retest responses from one institution (48%) and 3 retest responses from the other (100%). Agreement statistics were calculated using Cohen's κ in JMP® Pro 12 (DeVellis, 2016; McHugh, 2012) and we use these results to direct our attention to avenues for improvement of the survey.

Results and Discussion

We use McHugh's recommendations for Cohen's κ estimates of reliability (McHugh, 2012), and recognize that with only 15 respondents our values of κ are likely underestimates of stability. Under those guidelines, 4 of our 27 items were classified as 'moderate agreement ($0.60 \leq \kappa \leq 0.79$)' and another 9 as 'weak agreement ($0.40 \leq \kappa \leq 0.59$).' Six met the criteria for 'minimal agreement ($0.21 \leq \kappa \leq 0.39$)' and the remaining eight were classified as 'no agreement ($\kappa \leq 0.20$).' We discuss here two specific item pairings: one with weak (nearly moderate) agreement, and one with no agreement. We have selected these items not because they represent the extremes, but because they provide insight into issues with specific phrasings. It is worth noting that although we chose one C-I and one C-S item for this discussion, all three category pairings had pairs that scored well and pairs that scored poorly.

No Agreement, κ = -0.05. PROMPT: *Which is more important to you? Knowing how to evaluate class activities, OR writing a thorough test that assesses content mastery.*

Here, we have a C item paired against an S item. We suspect that *'evaluate class activities'* may be too broadly worded and could have also elicited elements of subject-matter although it was intended as a class-level item, which includes planning, delivering, and assessing

instructional activities. This is because one may perceive the evaluation of class activities as dependent on subject-matter knowledge.

Weak Agreement, κ = 0.59. PROMPT: ***Which scenario appeals to you more? You know how and when to adjust class pace and focus, OR you know what resources are available on campus for a student who is upset in class about an issue not related to content.***

This question is a pairing of a C item to an I item. This pairing shows weak agreement likely because each individual phrasing is a strong item in its own category and is not likely to be confused with the other categories. We anticipate that with a larger sample, this item might reach moderate or strong stability.

In general, the items with minimal or no stability are those that have one or both prompts conflated with a second category, rather than being strong anchors to just one category. For example, the prompt, *'Having multiple styles of teaching the same concept,'* was interpreted by some respondents as reflecting full mastery of the content. Although we built this item to exemplify C, it can also be seen as a strong S item. Confounding the stability test, several of the respondents were engaged in teaching and professional development during the six-week period between test and retest. Thus, some lack of stability may reflect an actual shift in instructional priorities. Since we do not have specific professional development enrollment linked to individual survey responses, we cannot clearly account for this effect.

Questions for the Audience

While most items in this version of the RIP survey failed to meet the criterion for moderate or strong stability, analysis of items that had minimal to no agreement as compared to items that had weak or moderate agreement has provided additional insight into the complexities of the choices MGTAs make as they develop their instructional practice. We continue to struggle with developing prompts that are anchored at the poles of our model in such a way that we can quantitatively capture the nuances that emerged in previous qualitative analysis. We welcome audience input to help guide our next steps. In particular:

- Our attempts to prime for different aspect of decision-making (importance, appeal, preference) do not appear to have produced meaningful differences in response category or test-retest stability. In our next iteration, we plan to prompt for what the MGTA's response has been, or would be, in specific scenarios. Do the additional insights we could potentially gain from pairing an action-based prompt with either an importance- or preference-based prompt outweigh the much higher participation rate we would need in order to analyze the data for differences in responses and stability between those two prompts?
- We have struggled particularly with crafting prompts that isolate aspects associated with the I category: meetings the needs of individual learners. Suggestions from the audience for specific prompts in this category would be most welcome.
- One eventual goal for the RIP survey is to use it to capture change in MGTA's instructional decision-making over time. Do you see an 'ideal' location in the triangle as an outcome for professional development for MGTAs?

References

Ball, D.L., & Bass, H. (2000). Interweaving content and pedagogy in teaching and learning to teach: Knowing and using mathematics. In J. Boaler, *Multiple Perspectives on Mathematics Teaching and Learning* (pp. 83-104). Westport: Greenwood Publishing Group.

Beauchamp, C., & Thomas, L. (2009). Understanding teacher identity: An overview of issues in the literature and implications for teacher education. *Cambridge Journal of Education*, 175-189.

Beauchamp, C., & Thomas, L. (2011). New teachers' identity shifts at the boundary of teacher education and initial practice. *International Journal of Educational Research*, 6-13.

Beijaard, D., Verloop, N., & Vermunt, J. D. (2000). Teachers' perceptions of professional identity: An exploratory study from a personal knowledge perspective. *Teaching and teacher education*, 749-764.

DeVellis, R. F. (2016). *Scale development: Theory and applications.* Thousand Oaks: Sage Publications.

Flores, M., & Day, C. (2006). Contexts which shape and reshape new teachers' identities: A multi-perspective study. *Teaching and Teacher Education*, 219-232.

Gallagher, E. D. (2016). On Becoming a Graduate Student Teacher of Mathematics. ProQuest Dissertations Publishing. (10151980).

Hamman, D., Gosselin, K., Romano, J., & Bunuan, R. (2010). Using possible-selves theory to understand the identity development of new teachers. *Teaching and Teacher Education*, 1349-1361.

Heyd-Metzuyanim, E., & Sfard, A. (2012). Identity struggles in the mathematics classroom: On learning mathematics as an interplay of mathematizing and identifying. *International Journal of Educational Research*, 128-145.

Hodges, T. E., & Cady, A. J. (2012). Negotiating contexts to construct an identity as a mathematics teacher. *The Journal of Educational Research*, 112-122.

Horn, I. S., Nolen, S. B., Ward, C., & Campbell, S. S. (2008). Developing practices in multiple worlds: The role of identity in learning to teach. *Teacher Education Quarterly*, 61-72.

Lasky, S. (2005). A sociocultural approach to understanding teacher identity, agency and professional vulnerability in a context of secondary school reform. *Teaching and Teacher Education*, 899-916.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia medica*, 276-282.

Sexton, D. M. (2008). Student Teachers Negotiating Identity, Role, and Agency. *Teacher Education Quarterly*, 73-88.

VanZoest, L., & Bohl, J. V. (2005). Mathematics Teacher Identity: A framework for understanding secondary school mathematics teachers' learning through practice. *Teacher Development*, 315-345.

Ward, C. J., Nolen, S. B., & Horn, I. S. (2011). Productive friction: How conflict in student teaching creates opportunities for learning at the boundary. *International Journal of Educational Research*, 14-20.