Bridging the Gap: From Graduate Student Instructor Observation Protocol to Actionable Post-Observation Feedback

Sean Yee
University of South Carolina

Jessica Deshler
West Virginia University

Kimberly Rogers
Bowling Green State University

Robert Petrulis
EPRE Consulting

Christopher D. Potvin
Bowling Green State University

James Sweeney
University of South Carolina

In this study, two universities created and implemented a student-centered graduate student instructor observation protocol (GSIOP) and a post-observational Red-Yellow-Green feedback structure (RYG feedback). The GSIOP and RYG feedback was used with novice graduate student instructors (GSIs) by experienced GSIs through a peer-mentorship program. Ten trained mentor GSIs completed 50 sets of three observations of novice GSIs. Analyzing 151 GSIOPs and 151 RYG feedback meetings longitudinally provided insight to identify what types of feedback informed and influenced GSIOP scores. After qualitatively coding feedback along multiple dimensions, we found certain forms of feedback were more influential for GSI development than others with respect to change in GSIOP score. Our results indicate contextually-specific feedback leads to more observed changes and improvement across multiple observations than decontextualized feedback.

Keywords: Graduate Student Instructors, Feedback, Observation, Mentoring

Acknowledgments

This work was supported by Collaborative IUSE NSF grants (#1544342 & 1544346; #1725264, 1725295, & 1725230). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We also want to thank the University of South Carolina Reading and Research Group in Mathematics Education for providing insightful feedback.

Introduction

Mathematics graduate student instruction significantly impacts undergraduate courses and students (Belnap & Allred, 2009). Graduate student instructors (GSIs)[1] have been identified as a key component of success for collegiate mathematics departments for teaching undergraduate mathematics (Bressoud, Mesa, & Rassmussen, 2015, p. 117). As a result, mathematics departments and research in undergraduate mathematics education continue to focus on supporting and improving GSIs' student-centered instruction (Rogers & Yee, 2018; Speer & Murphy, 2009; Yee & Rogers, 2017). There are multiple methods of student-centered pedagogical support for GSIs (e.g. professional development, mentoring, pedagogically-focused courses; Speer, Gutmann, & Murphy, 2005; Yee & Rogers, 2017), but there is currently limited research on GSI teaching observation protocols and even less research on post-observation feedback (Reinholz, 2017). Multiple observation protocols exist to assess undergraduate mathematics instructors' classrooms (e.g. MCOP[2], RTOP, C-LASS, etc.), often with scalar

[1] GSI was used instead of TA (Teaching Assistant) because GSI references graduate students who are full instructors of record.

metrics such as 1-4, but many do not discuss how to make that assessment actionable so that it can be beneficial for the teacher.

To this end, we created a GSI observation protocol (GSIOP, Rogers, Petrulis, Yee & Deshler, under review) and a post-observation feedback structure at two universities to provide ongoing support for novice GSIs. Together, the GSIOP and feedback were implemented for two years as part of a peer-mentorship model[2] where novice GSIs were mentored by experienced (two or more years of experience) GSIs who had completed a mentor professional development (PD) seminar. This mentor PD included training with the GSIOP and post-observation feedback (See Rogers & Yee, 2018 and Yee & Rogers, 2017 for more information on peer-mentorship). The purpose of this paper is to help bridge the research gap between observations and post-observation feedback by identifying how feedback within this peer-mentoring model informed and influenced future observations. Our research questions for this study are:

- RQ1: In what ways (if any) did the feedback structure lead to changes in teaching observations throughout a semester?
- RQ2: How do those changes inform (if at all) methods for providing actionable feedback to influence observed teaching?

It is important to note that our study focused on GSIs but the observation protocol, feedback, and results are applicable to undergraduate mathematics instructors, not just GSIs.

Related Literature

Feedback

For over a century, psychology has long researched the importance of feedback as a means to change performance, cognition, and understanding in many professions (Kluger & DeNisi, 1996). Hattie and Timperley's (2007) meta-analysis looked at 500 articles of teachers providing feedback to students and found assessment-based feedback was one of the most dangerous forms because "rarely does such enhance the processes and metacognitive attributes of the task" (p. 101). White's (2007) research on 16 pre-service teachers showed that clear, concise, specific, and encouraging feedback were the most valuable forms of feedback. White's research also emphasized what Hattie and Timperley (2007) identified, that feedback (and thus observations) needs to be done regularly, not intermittently.

Although K-12 mathematics education research has extensively studied feedback within practicum courses (e.g. student teachers are observed regularly by their master teacher and university supervisor as a critical means of ongoing teacher development) our review of the literature has found few studies focusing on mathematics GSI peer feedback (Reinholz, 2017; Rogers & Steele, 2016; Yee & Rogers, 2017; Rogers & Yee, 2018). One exception is a recent study by Reinholz (2017) that explores peer feedback with mathematics graduate students as equal peers. Reinholz had six GSIs provide peer-feedback to one another and found that feedback not only helped the novice, but enhanced teacher noticing (Sherin, Jacobs & Philipp, 2011) and reflection in the observer, aligning with Reinholz's previous work (2016) where peer assessment led to improved self-assessment. Rogers and Steele (2016) concluded that novice instructors struggle to discuss teaching methods, which Reinholz (2017) argues could be aided by peer feedback. Thus, Reinholz's and Rogers and Steele's (2016) research supports post-observation feedback as a means of improving GSIs' teaching through discourse and reflection.

[2] Supported by a Collaborative National Grant

Complexities of Observations and Feedback

Reinholz (2017) reminds us that "how instructors engage with peer feedback is complicated" (p. 7) due to GSIs' beliefs about mathematics and its often-assumed relationship to innate intelligence. Kluger and DeNisi's (1998) meta-analysis of 607 studies on feedback interventions (i.e. providing people with some information regarding their task performance) showed that while overall feedback improves performance, it can also sometimes reduce performance, depending on the type of feedback and means by which it is delivered. Certain feedback was helpful for improving performance as long as attention was directed towards task-motivation and task-learning rather than praise, negative criticism, or focus on the person because deviating from the focus on the task requires effort that was found to decrease performance.

In light of the complexity that links observations and feedback, we questioned what type of feedback is most effective for GSIs. Cannon and Witherspoon (2005) provide a framework to navigate this complexity effectively using *actionable feedback* "that produces both learning and tangible, appropriate results" (p. 120). Actionable feedback provides a framework for examining undergraduate mathematics classrooms and providing feedback to help novices make changes to improve their teaching. We use this frame in our data analysis to determine how feedback affected the tangible result of novices' GSIOP scores over a semester.

Framework of Study

Our peer-mentorship research (Yee & Rogers, 2017; Rogers & Yee, 2018) and current literature (Reinholz, 2017) has found observational protocols need to have complementary feedback structure where novices are able to reflect more openly about how they can modify their teaching to achieve their goals. Hence, our design emphasizes post-observation feedback as reflective to complement the more evaluative observation protocol.

GSIOP

The initial goal of our peer-mentorship model was to provide feedback and facilitate discussions among novice GSIs around student-centered teaching strategies to improve undergraduate mathematics instruction (Rogers et. al., 2018, under review). The Mathematics Classroom Observation Protocol for Practices ($MCOP^2$, Gleason, Livers & Zelkowski, 2017) is an observation protocol designed for K-12 that originates from the STEM-based Reformed Teaching Observation Protocol (RTOP, Sawada et al., 2002), but unlike the RTOP, includes a means to observe student-centered investigations and collaborative learning environments focusing on mathematics. Thus, we modified the $MCOP^2$ to be applicable for use when observing GSIs and developed the GSIOP which focuses on both student and instructor actions. Similar to the $MCOP^2$, the GSIOP contains questions on an ordinal scale from 0 to 3 for four sections: classroom management, student engagement, teacher facilitation, and lesson design. A more thorough explanation of the GSIOP design can be found in Rogers et al's validation study (under review).

RYG Feedback

Mentors were educated through the mentor PD to use the GSIOP during their PD program (see Yee & Rogers, 2017) and to facilitate post-observation conversations using a Red-Yellow-Green feedback structure. Using this structure, mentors identify key points from the GSIOP that they could summarize for the novice in three categories: methods the novice is doing well (green), methods the novice could work on (yellow), and methods the novice needs to address (red). The mentor would summarize points of discussion from the GSIOP and keep the feedback

manageable by discussing at most two concerns within the yellow and red categories. Scenarios, role playing, and live observations helped prepare mentors to provide feedback in each category during post-observation meetings that occurred within a week of the observation. We refer to this post-observation feedback as the *RYG feedback*.

Methods

In this mixed-methods study, we quantitatively analyzed changes to GSIOP scores to answer our first research question. We then qualitatively coded the RYG feedback for types of actionable feedback and compared the types of feedback with the changes in GSIOP scores to answer our second research question.

Participants & Observations

This study included 10 mentor GSIs and 32 novice GSIs from two universities in the United States over two semesters. New novices were added between semesters while other novices completed their training after one semester. For this reason, we focused on sets of semester-long observations, which consisted of three observations with feedback for each novice on average (two novices were observed only twice while three novices were observed four times). This generated 50 sets of semester-long observations with feedback, totaling 151 observations with feedback. Mentors submitted novice teaching notes, videos of the novice's class, observation summaries, completed GSIOPs, and RYG feedback for analysis.

Data Analysis

As our research study emphasized student-centered instruction and RYG feedback, we focused only on the two sections of the GSIOP that emphasized student-centered instruction, the *student-focused* (student engagement) and *teacher-focused* (teacher facilitation) sections. One research assistant at each university longitudinally analyzed the GSIOP scores from both the student- and teacher-focused sections for each novice over an entire semester. Similarly, each research assistant analyzed the RYG feedback and observation summaries for student-focused feedback and teacher-focused feedback that aligned with the questions from appropriate sections of the GSIOP. This created 100 longitudinal data sets of semester-long observations and 100 data sets of semester-long feedback (50 student-focused and 50 teacher-focused).

To answer our first research question, we summed the questions on the GSIOP student-focused section (4 questions) and the GSIOP teacher-focused section (5 questions) separately. Thus, for each observation of each novice each semester, there was a teacher-focused GSIOP score and a student-focused GSIOP score. We looked at change in GSIOP scores over a single semester by looking for trends and subtracting novices' final GSIOP score from their initial GSIOP score for both the student- and teacher-focused sections.

To answer the second research question, we looked at the data collected by the mentor during each observation and the feedback each novice received from the mentor. We analyzed feedback through an *advice and improvement* framework. We looked at RYG feedback, GSIOP comments, and mentor observation summaries for suggestions that provided the novice with *advice* on teaching that focused on student learning or teacher facilitation. We then looked through the data sets at each novice to see if the mentor noted any observed *improvements* related to advice given previously in the semester.

Next, we coded each piece of advice and each noted improvement as *broad* or *specific*. To frame broad versus specific objectively, we used Nilsson and Ryve's (2010) definition of contextualization where the context of an event must be given to make a situation specific and

not referencing a context or event (often referred to as decontextualized) would be considered broad. Looking at feedback as advice or improvement concomitantly as broad or specific provides a categorization demonstrated on Table 1 with prototypical examples.

The last two categories, Advice Without Improvement (AWI) and No Advice Nor Improvement (NANI) took into account if advice and improvement were not given. AWI implied advice (broad or specific) was given, but improvement was not noted in subsequent observations. NANI lacked advice and therefore no improvement could be noted in subsequent observations.

To triangulate the qualitative coding of advice and improvement as broad or specific, after each research assistant qualitatively coded the results according to Table 1, two additional researchers went back and verified their work by comparing 75 of the 151 observations and post-observation feedback artifacts for both teacher-focused feedback and student-focused feedback. Interrater agreement was initially 94% and after discussion of the coding discrepancies, researchers agreed on the appropriate coding for the remaining 6%.

Table 1. Qualitative Coding Scheme for Feedback across an Entire Semester

Code	Description	Example
SASI	**Specific Advice Specific Improvement**: Feedback included at least one contextualized suggestion the novice could take to improve their teaching. In subsequent observations, the mentor noted that the novice had addressed the issues through particular contexts, actions, and/or strategies.	"Elaborate with the material and explain the importance of the concept. For example, one instance in which you could give a little more insight and explanation was when the student used P(A U B) = P(A)+P(B) - P(A cap B)"...(*later observation*) "You elaborated more than last time.. I felt that this was the perfect amount of elaboration. Also, you asked well thought out questions, and you rarely missed good opportunities to ask further questions."
BASI	**Broad Advice Specific Improvement**: Feedback included suggestions without context on when or how to improve the novice's teaching. In subsequent observations, the mentor noted that the novice had addressed the issues through particular contexts, actions, and/or strategies.	"Have tiny bits of student involvement through to keep students engaged" … *(later observation)* "Student questioning chosen was very effective in engaging students [with 2^x and log_2(x)]"
SABI	**Specific Advice Broad Improvement**: Feedback included at least one contextualized suggestion the novice could take to improve their teaching. In subsequent observations, the mentor noted that the novice had improved upon previous issues, but without referencing specific contexts.	"I encourage you to give more wait time before answering the questions yourself, this can have them participate more" … (*later observation*) "I saw great improvement since last time with student engagement….(*later observation*) "Great student interaction".
BABI	**Broad Advice Broad Improvement**: Feedback included suggestions without context on when or how to improve the novice's teaching. In subsequent observations, the mentor noted that the novice had improved upon previous issues, but without referencing specific contexts.	"Student engagement should be addressed" … (*later observation*) "Even though she ask[ed] many questions, students are not really active in this particular class"…(*later observation*). "She did not just answer but encourage[d] students to respond".
AWI	**Advice Without Improvement**: Feedback included suggestions, but the suggestions did not appear to be noted throughout the subsequent observations.	"For the next time, I hope that he can get more active participation during his lecture portions" *No follow up.*
NANI	**Neither Advice Nor Improvement**: Feedback was either statements extolling the novice's instruction or platitudes on teaching. Mentor did not provide advice nor improvements.	"He did a great job in his lesson of engaging the students, explaining material adequately and also giving his students problems to work on at the end of class". *No advice.*

Results

Longitudinally, each novice's three GSIOP scores from both the student-focused and teacher-focused sections determined how each set of three scores varied. We categorized the changes as *decrease* (each observation was at least two points less than the previous one), *steady* (each observation was within one point of the previous one), *moderate increase* (each observation was at least two points higher than the previous), *substantial increase* (each observation was at least three points higher than the previous), *hill* (middle score is at least two points higher than the other scores), and *valley* (middle score is at least two points lower than the other scores). Table 2 shows how many student-focused and teacher-focused sections (changes across a semester) fell into each category.

Table 2. Longitudinal Semester-Long Changes in GSIOP Scores by Student- and Teacher-Focused Sections

GSIOP Change Categories	Substantial Increase	Moderate Increase	Steady	Decrease	Hill	Valley	Grand Total
Number of Student-Focused Sections	9	12	15	10	2	2	50
Average GSIOP Change Per Student-Focused Section	5.00	2.50	0.20	-3.90	-1.00	-0.50	0.72
Number of Teacher-Focused Sections	9	14	18	5	3	1	50
Average GSIOP Change Per Teacher-Focused Section	5.11	2.21	0.28	-3.60	0.67	-1.00	1.30
Number of Student- and Teacher-Focused Sections	18	26	33	15	5	3	100
Average Change Per Student- and Teacher-Focused Sections	5.06	2.35	0.24	-3.80	0.00	-0.67	1.01

Results show that for both the student- and teacher-focused sections, on a 0-3 point scale, there was an average positive change of 1.01 points per section. We see that the number within each category had a fairly equal distribution between student- and teacher-focused sections, with the student-focused sections showing more decreases and the teacher-focused sections showing more steady or moderate increases. Although a majority of the GSIOP scores remained steady (33 out of 100), there were significantly more novices whose score increased moderately or substantially (44) than those that decreased (15) over a semester. Thus, our results indicated there was an observed change in teaching throughout a semester via the GSIOP score showing an overall increase in point value.

To answer our second research question, we wanted to understand the feedback at a more contextual (Nilsson & Ryve, 2010) level to determine how the feedback was actionable. We tallied the total change in score for all novices during a semester by taking the final GSIOP score for each section and subtracting it from the initial GSIOP score for that section. We then divided the total change by the number of novices to get the average change per novice.

Table 3. Inductive Analysis of Feedback Types Cross-Referenced with Change in GSIOP score

Feedback Types	SASI	BASI	SABI	BABI	NANI	AWI	Grand Total
Student-Focused Feedback	4	2	7	12	11	14	50
Average GSIOP Change Per Student-Focused Section	4.50	3.50	3.57	0.58	-0.73	-0.93	0.72
Teacher-Focused Feedback	10	4	4	8	5	19	50
Average GSIOP Change Per Teacher-Focused Section	3.40	3.00	-0.25	2.38	0.80	-0.16	1.3

Student and Teacher Feedback	14	6	11	20	16	33	100
Average GSIOP Change Per Student- and Teacher-Focused Feedback	3.71	3.17	2.18	1.30	-0.25	-0.48	1.01

Table 3 shows that of all 100 data sets of semester-long feedback, the one with the highest average change in GSIOP score was when mentors provided and noticed Specific Advice and Specific Improvement (SASI, *M=3.71)*. SASI feedback also resulted in the highest change in GSIOP scores for both student and teacher sections. BASI feedback provided high changes as well, but with fewer student-focused feedback (*N=2*) and teacher-focused feedback (*N=4*) sections. SABI feedback influenced the student-focused section more (*M=3.*57) than the teacher-focused section (*M=-0.25*) while BABI feedback influenced the teacher section (*M=2.38*) more than the student section (*M=0.58*). Both Advice Without Improvement (AWI, *M=-0.48*) feedback and No Advice and No Improvement feedback (NANI, *M=-0.25*) had the least change in GSIOP scores.

Discussion

In answering our first research question, we see from Table 2 that RYG feedback in our study led to both increases and decreases in GSIOP scores associated with student engagement and teacher facilitation, but that there were more increases than decreases in GSIOP scores over semester-long observation-feedback iterations. In answering our second research question, our coding of feedback (advice/improvement and broad/specific) illustrated how GSIOP scores on the teacher and student sections would change relative to the type of feedback. Moreover, feedback that included specific advice and specific improvements had the largest positive change in GSIOP observation score indicating that contextualizing feedback leads to more actionable feedback.

Limitations

The structure of the post-observation feedback and the overall design of the peer-mentorship model could have influenced the results of this study. Specifically, the training of mentors and the use of the peer-mentorship model may be critical factors in the results of this study. This in no way voids the results but is a limitation of implementing RYG feedback with another observation protocol or using the GSIOP with a non-RYG feedback structure.

Implications for Research and Practice

Tables 2 and 3 support Kluger and DeNisi's (1998) theory of feedback being "a double-edged sword" because Table 2 demonstrates overall growth to both the student and teacher sections, but it varies according to the type of feedback. Table 3 verifies Kluger and DeNisi's argument that change depends on the type of feedback. When mentors provided specific advice and noted specific improvement, or provided broad advice and noted specific improvement, novice GSIOP scores improved on observation questions focusing on student engagement and teacher facilitation of student-centered learning. However, if the mentor's feedback provided no advice nor improvements, or advice without improvements, there was a minor positive or negative change in GSIOP score for both student engagement and teacher facilitation of student-centered learning.

Our research provides undergraduate mathematics education with a framework for looking at post-observation feedback using a tested observation protocol (Rogers et al., under review) and a post-observation feedback structure. Our results (Table 3) indicate providing specific

improvements had the most actionable (Cannon & Witherspoon, 2005) results with respect to the observation protocol. Consider Roberto's yellow feedback and following green feedback which had a substantial increase in his novice's student- and teacher-focused GSIOP scores.

> *(Yellow Feedback)* Engage more with the students. Particularly, ask more questions. I see that you are using the PowerPoints…I will do a demonstration for you in the one-on-one for a slide that was in your lecture. The main thing is to actively think if this is a moment I can ask a constructive question to engage with the learning… *(Following Green Feedback)* You are asking more questions to your students and you are getting more participation! This is great. Keep it up but remember that you can also…

The specific advice to engage through questioning, followed by specific improvement that promoted continued development demonstrates actionable feedback that can positively frame post-observation feedback.

References

Belnap, J. K., & Allred, K. (2009). Mathematics teaching assistants: Their instructional involvement and preparation opportunities. In L. L. B. Border (Ed.), Studies in Graduate and Professional Student Development (pp. 11–38). Stillwater, OK: New Forums Press, Inc.

Boyle, P., & Boice, B. (1998). Systematic mentoring for new faculty teachers and graduate teaching assistants. *Innovative Higher Education*, *22*(3), 157–179. http://doi.org/10.1023/A:1025183225886

Bressoud, D., Mesa, V., & Rasmussen, C. (Eds.). (2015). *Insights and recommendations from the MAA national study of college calculus*. MAA Press.

Gleason, J., Livers, S., & Zelkowski, J. (2017). Mathematics Classroom Observation Protocol for Practices (MCOP2): A validation study. *Investigations in Mathematics Learning*, 9(3), 111-129.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81-112.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological bulletin*, *119*(2), 254.

Kluger, A. N., & DeNisi, A. (1998). Feedback interventions: Toward the understanding of a double-edged sword. *Current directions in psychological science*, *7*(3), 67-72.

Nilsson, P., & Ryve, A. (2010). Focal event, contextualization, and effective communication in the classroom. *Educational Studies in Mathematics, 74*(3), 241-258.

President's Council of Advisors on Science and Technology [PCAST]. (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics* (Report to the President) (p. 103). Washington, DC: Office of the President.

Reinholz, D. (2016). The assessment cycle: a model for learning through peer assessment. *Assessment & Evaluation in Higher Education*, *41*(2), 301-315.

Reinholz, D. L. (2017). Not-so-critical friends: Graduate student instructors and peer feedback. *International Journal for the Scholarship of Teaching and Learning*, *11*(2), n2.

Rogers, K. C., Petrulis, R., Yee, S., & Deshler, J. (under review). Mathematics Graduate Student Instructor Observation Protocol (GSIOP) Validation Study.

Rogers, K. C., & Steele, M. D. (2016). Graduate teaching assistants' enactment of reasoning-and-proving tasks in a content course for elementary teachers. *Journal for Research in Mathematics Education*, *47*, 372–419.

Rogers, K.C. & Yee, S.P. (2018, February). Peer mentoring mathematics graduate student instructors: discussion topics and concerns. Proceedings from *21st Conference of the Research in Undergraduate Mathematics Education* (RUME), San Diego, CA.

Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., & Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School science and mathematics*, *102*(6), 245-253.

Sherin, M., Jacobs, V., & Philipp, R. (2011). *Mathematics teacher noticing: Seeing through teachers' eyes*. New York: Routledge.

Speer, N. M., Gutmann, T., & Murphy, T. J. (2005). Mathematics teaching assistant preparation and development. *College Teaching*, *53*(2), 75–80.

Speer, N. M., & Murphy, T. J. (2009). Research on graduate students as teachers of undergraduate mathematics. In L. L. B. Border (Ed.), *Studies in Graduate and Professional Student Development* (pp. xiii–xvi). Stillwater, OK: New Forums Press, Inc.

White, S. (2007). Investigating effective feedback practices for pre-service teacher education students on practicum. *Teaching Education*, *18*(4), 299-311.

Yee, S.P. & Rogers, K. C. (2017, February). Mentor professional development for mathematics graduate student instructors. Proceedings from *20th Conference on Research in Undergraduate Mathematics Education* (RUME, pp. 1026-1034), San Diego, CA.