CONNECTING OBSERVATION PROTOCOLS AND POST-OBSERVATION FEEDBACK

SEAN YEE
UNIVERSITY OF SOUTH CAROLINA
YEE@MATH.SC.EDU

JESSICA DESHLER
WEST VIRGINIA UNIVERSITY
JMDESHLER@MAIL.WVU.EDU

KIMBERLY ROGERS
BOWLING GREEN STATE UNIVERSITY
KCROGER@BGSU.EDU

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 mathematics graduate student instructors (GSIs) by experienced GSIs through a peer-mentorship program. Ten trained mentor GSIs observed novice GSIs, completed a GSIOP, and provided RYG feedback as part of an observation-feedback cycle. This generated 50 semester-long data sets of three observation-feedback cycles of novice GSIs. Analyzing these 50 semester-long data sets helped identify how certain feedback influenced GSIOP scores.

Introduction

Mathematics graduate student instruction significantly impacts undergraduate courses and students (Belnap & Allred, 2009). Graduate student instructors (GSIs)[i] have been identified as a key component of success for collegiate mathematics departments (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 metrics such as point values 1-4, but few discuss how to connect that assessment with observer feedback.

To this end, we created a GSI observation protocol (GSIOP) 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[ii] 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 question for this study is in what ways (if any) did the feedback structure lead to changes in teaching observations throughout a semester?

Related Literature

Feedback

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 & Yee 2016, 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 and reflection in the observer, aligning with Reinholz's previous work (2016) where peer assessment led to improved self-assessment. Rogers and Yee (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 Yee's (2016) research supports post-observation feedback as a means of improving GSIs' teaching through discourse and reflection.

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. In light of the complexity that links observations and feedback, we questioned what type of feedback is most effective for GSIs.

Framework of Study

Our peer-mentorship research (Rogers and Yee 2016, 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 (Yee & Rogers, 2017). We modified the MCOP^2 (Gleason, Livers & Zelkowski, 2017) to observe GSIs to develop the GSIOP which focuses on both student and instructor actions. 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 and Yee's validation study (forthcoming).

RYG Feedback

Mentors were educated through the mentor PD to use the GSIOP and 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.

Methods

In this mixed-methods study, we quantitatively analyzed changes to GSIOP scores. 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 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. This generated 50 data sets of semester-long observations with feedback (totaling 151 individual 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 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. Additionally, 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

Due to limited space, we will briefly summarize the longitudinal trends. Each novice's three GSIOP scores from both the student-focused and teacher-focused sections determined how each set of three scores varied. 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. Although a majority of the GSIOP scores had less than a one point change from previous GISOPs (33 out of 100), there were significantly more novices whose score increased by more than one point (44) than those that decreased by more than one point (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.

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 2. 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 2 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. 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.

We provide a small excerpt demonstrating SASI semester-long feedback that generated a substantial increase in his novice's student- and teacher-focused GSIOP scores. Consider Roberto's yellow feedback and following green feedback which had a

> *(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… (*Coded SASI*)

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

Discussion

In answering our research question, we found that the RYG feedback in our study there were more increases than decreases in GSIOP scores over semester-long observation-feedback iterations, illustrating novices were attending to mentor feedback. Additionally, 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. 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

Table 2 verifies Kluger and DeNisi's (1998) 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 & Yee, forthcoming) and a post-observation feedback structure. Our results (Table 2) indicate providing specific improvements had the most actionable (Cannon & Witherspoon, 2005) results with respect to the observation protocol.

---

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

[ii] This work was supported by grants from the National Science Foundation (NSF DUE 1544342, 1544346, 1725295, 1725230 and 1725264). 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.

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.

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.

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.

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., & 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.

Rogers, K.C., Petrulis, B., Yee, S.Y., Deshler, J., Potvin, C., Sweeney, J. (forthcoming), Mathematics graduate student instructor observation protocol (GSIOP) validation study.

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.

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.

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.