



Reliability of the Global Real-time Assessment Tool for Teaching Enhancement (G-RATE)

Ms. Nikitha Sambamurthy, Purdue University, West Lafayette

Mrs. Jeremi S London, Purdue University, West Lafayette

Jeremi S. London is a graduate student at Purdue University. She is pursuing a Ph.D. in Engineering Education. In 2008, she earned a Bachelor of Science in Industrial Engineering from Purdue, and a Master of Science in Industrial Engineering from Purdue in 2013. Her research interests include: the use of cyberlearning in science, technology, engineering, and mathematics (STEM) education; assessing the impact of cyberlearning; and exploring ways computer simulations can be used to understand complex education systems. In addition to conducting mixed methods research in formal academic settings, she has worked in various Industrial Engineering roles at Anheuser-Busch, Inc., and General Electric-Healthcare. She continues to conduct education policy research as part of her internship role in the Division of Undergraduate Education at the National Science Foundation.

Jeeyeon Hahn, Purdue University

Jiabin Zhu, Purdue University, West Lafayette

Dr. Monica Farmer Cox, Purdue University, West Lafayette

Monica F. Cox, Ph.D., is an Associate Professor in the School of Engineering Education at Purdue University and is the Inaugural Director of the Engineering Leadership Minor. She obtained a B.S. in mathematics from Spelman College, a M.S. in industrial engineering from the University of Alabama, and a Ph.D. in Leadership and Policy Studies from Peabody College of Vanderbilt University. Teaching interests relate to the professional development of graduate engineering students and to leadership, policy, and change in science, technology, engineering, and mathematics education. Primary research projects explore the preparation of engineering doctoral students for careers in academia and industry and the development of engineering education assessment tools. She is a National Science Foundation Faculty Early Career (CA-REER) award winner and is a recipient of a Presidential Early Career Award for Scientists and Engineers (PECASE).

Reliability of the Global Real-time Assessment Tool for Teaching Enhancement (G-RATE)

Abstract

Despite the integral role that instructors, particularly graduate teaching assistants, play in the success of higher education, they rarely receive multidimensional feedback on their pedagogical effectiveness. In response to the need for research-based assessment tools for effective teaching and to provide feedback to graduate teaching assistants about their instructional interactions in a classroom, the Global Real-time Assessment Tool for Teaching Enhancement (G-RATE) has been developed in the context of the “How People Learn” framework. Although the G-RATE is composed of many functions, this paper presents a revised version of one component of the G-RATE, the *Observer* function, which serves as the direct observation classroom portion in the G-RATE system. Therefore, it is important to closely examine the reliability of this function. Using both pen-and-paper and video vignettes of engineering classroom occurrences, researchers report the process for determining the inter-rater (inter-observer) reliability of this function and report possible changes to the G-RATE based upon reliability measures. Future work related to the *Observer* function is also presented.

Introduction

Classroom observation tools can be useful in providing real-time feedback to instructors’ teaching practice not only in a traditional lecture-based classroom, but also in more recent engineering courses, which often include team-based, laboratory activities and some problem-based and/or project-based learning activities. The latter often requires multiple pedagogical skills, such as the skills to motivate students and guide in their problem-solving processes, the ability to arrange team-based learning activities and other classroom management skills, which can be challenging for new instructors or graduate teaching assistants (GTAs) before they gain enough teaching experience and basic teaching skills¹. Many of these instructors can benefit from diverse forms of feedback about their teaching, particularly from real-time classroom observations given in a timely and effective manner.

In response to the need for effective assessment tools, the Global Real-time Assessment Tool for Teaching Enhancement (G-RATE) has been developed in the context of the How-People-Learn (HPL) framework². Revised from a tool used to collect pedagogical data among bioengineering undergraduate students³, the G-RATE incorporates HPL principles in ways that are easy for observers to capture in the classroom and for instructors to interpret once they receive HPL-based pedagogical profiles of their teaching. The HPL framework is useful in guiding the establishment of an effective learning environment that embraces the four important dimensions of the framework, that is, knowledge-centeredness, learner-centeredness, assessment-centeredness and community-centeredness². The HPL framework has been used to guide the design of learning modules or activities^{4,5}, and some assessment tools⁶. Using the HPL framework to guide the design and implementation of the *Observer* function can potentially improve the assessment process by framing the observation and feedback process in the said

framework.

The G-RATE has a flexible and easy-to-use interface that allows use in lecture and laboratory environments to collect data from a variety of stakeholders, including undergraduate students, instructors, researchers, and administrators. The G-RATE provides feedback from multiple sources, including survey responses from students, real-time classroom observations, and/or the instructors' self-reflections. Here, the authors focus on the Observer function, which provides real-time observations of pedagogical practices implemented by an instructor. For the purposes of this study, videos of classroom interactions were used to simulate the environment the improved tool will be used in for future studies. In the future, this tool can be used to conduct real-time observations or to conduct analysis using video recordings of previous classes.

Overview of the Observer Function

This paper presents a brief overview of the revised version of the *Observer* function of the G-RATE⁷. Details about the original version of the *Observer* function and the other G-RATE functions can be found in Authors (2011)⁷. Prior to developing the G-RATE, researchers used the VaNTH Observation System (VOS)³ to collect real-time, in-class data about GTAs' pedagogical practices in first-year engineering laboratories at a large Midwest university over the course of two semesters. However, some of the categories of the VOS were not intuitive to use and considerable training was required to achieve a desired 85% inter-rater reliability. Based upon VOS observations and upon findings that resulted from conducting pilot studies among GTAs in first-year engineering classrooms, researchers suggested that the following elements be incorporated into an improved direct observation system with the following characteristics⁸:

- The tool should be easy to use and should not require extensive observer training.
- A complete description of the codes and what they depict should be integrated into the tool so that users do not have to memorize the codes and their definitions.

After the G-RATE prototype was developed in 2011, three members of the research team tested and validated the *Observer* function of the G-RATE. The members were introduced to the HPL framework and its applicability within the G-RATE and received training on the *Observer* function prior to testing. Observations were performed in 15-second coding intervals using videotapes of engineering classrooms³. This resulted in the development of a revised version of the *Observer* function (Figure 1), which provided phrases representing possible actions that might occur within an observed classroom.

Figure 1. First revision of the *Observer* function of the G-RATE

Despite of the advances made in the *Observer* function of the G-RATE from the VOS, inter-rater reliability was hampered by two factors. First, the researchers determined that the 15-second time interval was still too long. While 15-second time intervals may be an enough time for the researchers to thoughtfully record responses, it is also likely that two or more events can happen during this time interval. For this reason, one observer can record the code string based on what was happening at the beginning of the time interval while others may code a response according to what is happening at the end of the time interval. This, of course, affects the consistency in the responses among researchers. Second, multiple selections of *how* category containing the four HPL dimensions (i.e., knowledge-centered, learner-centered, assessment-centered, and community-centered), along with the codes for classroom organization led to a lower inter-rater reliability. Due to the subjective nature of these HPL dimensions and to the integrative nature of the HPL framework, combinations of *how* items can vary between the observers. For example, distinguishing between observation codes for when an instructor “explained how to solve specific problems” and “provided guidance for to whom during problem-solving activities” was difficult for observers. As a result of these two factors that hampered inter-rater reliability, the *Observer* function of the G-RATE was revised again (Figure 2).

Observer

Start **End** **Analyze**

Organization

Who

Instructor 1 Instructor 2 Instructor 3

To Whom

Everyone Group Individual

Community-Centered

- engaged student(s) in team activities
- provided opportunities for students to learn from each other
- helped teams when they needed assistance

Assessment-Centered

- asked thought-provoking questions
- provided verbal or written feedback to students about their progress or performance in class
- confirmed that the class understood content before moving to a new topic

Learner-Centered

- acknowledged that learning engineering concepts can be hard at times
- acknowledged student's misunderstandings of concepts
- provided guidance for students during problem-solving activities

Knowledge-Centered

- shared his/her own practical experiences
- related course content to everyday situations
- shared skills students can apply later
- emphasized learning new skills
- helped students to understand key concepts

Figure 2. Updated Observer function of the G-RATE

Methodology

Figure 2. Updated Observer function of the G-RATE

Two researchers conducted two rounds of observations using the *Observer* function of the G-RATE to determine inter-rater reliability scores. Both rounds of observations were conducted on 3-minute videos of bioengineering classes, each with approximately 20-30 students. For the first round of observations, two videos of classroom instruction were randomly selected. Observers watched the videos once to familiarize themselves with the media, and then used the tool on the second viewing to record classroom interactions in 10-second intervals. Inter-rater reliability was calculated using Cohen's Kappa⁹. Cohen's Kappa is an index for measuring the extent of agreement between two coders. This index is often referred to as an inter-rater reliability score.

After this initial coding, the observers noted that instructors spent a significant portion of time explaining tasks in preparation for doing, and there was no way to record those interactions using the revised version of the G-RATE. As a result of this, a "transition" observation code was added to the tool, and another round of observations was conducted. For the second round, three 3-minute videos were randomly selected, and 10-second observations were recorded. Table 1 presents an overview of the context of the courses used to conduct reliability checks.

Table 1. Description of the video

	Instructor A	Instructor B	Instructor C
Clip 1	Small classroom 8 students	Midsized classroom About 20 students	Midsized classroom About 20 students

Results and Discussion

For the first round of observations, the inter-rater reliability was 86.8% for the first video, and 95.2 % for the second video. For the second round of observations, the inter-rater reliability was 96.8% for the first video, 100% for the second video, and 90.9% for the third video. The majority of coding discrepancies in the second round of ratings was due to timing issues, such as when an observer would carry over a code from the previous time period or code an observation in the time period prior to when it actually occurred. The perfect rating for the second video is attributed to the fact that the lecturer was exceptionally clear in providing feedback to students, and the majority of the video was spent asking students questions. Since raters received no training on the tool, and videos were selected randomly, the significance in the change of reliability scores can be attributed to the improvement of the *Observer* function.

Conclusion

This paper is meaningful in that it confirms the reliability of a direct observation tool that is based upon the HPL framework. It has the potential to frame classroom observations of teaching in a meaningful way for instructors. The G-RATE can be used to conduct real-time observations or by viewing video recordings of teaching experiences. The use of videos is practical, since future classroom observations will be videotaped and will be coded by trained observers. Future work will translate code strings collected within an entire class period into instructor profiles displaying the percentage of course time spent engaging in each HPL activity and will triangulate data collected in the *Observer* function with other components of the G-RATE.

The findings of this study are relevant to other researchers in a variety of ways. It presents an example of a classroom observation tool that can be used to assess and provide feedback on pedagogical practices. The elements of the observation tool and the feedback that results are grounded in the HPL framework, well-established theoretical framework that is useful for developing courses that smoothly integrate the major facets of course design. The results of this study also demonstrate a process for validating the contents of the instrument and a process for improving the rating consistency among coders.

References

- ¹Felder, R.M., Brent, R., & Prince, M (2011). Engineering Instructional Development: Programs, Best Practices, and Recommendations, *Journal of Engineering Education*, 100(1), 89 –122.
- ²Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds). *How People Learn: Brain, Mind, Experience, and School*.

Washington, DC: National Academy Press, 1999.

- ³Harris, A.H., and M.F. Cox. 2003. Developing an observation system to capture instructional differences in engineering classrooms. *Journal of Engineering Education* 92 (4): 329–36.
- ⁴Vanderbilt-Northwestern-Texas-Harvard/MIT Engineering Research Center for Bioengineering Educational Technologies (VaNTH), website. <https://repo.vanth.org/portal/public-content/how-people-learn/how-people-learn>, retrieved on Jan. 6, 2013.
- ⁵Cordray, D., Harris, T., & Klein, S. (2009). Research synthesis of the effectiveness, replicability, and generality of the VaNTH challenge-based instructional modules in bioengineering. *Journal of Engineering Education* 98(4), 335–348.
- ⁶Cox, M.F., Zhu, J., London, J., Hahn, J., & Ahn, B. (2012). Feedback about Graduate Teaching Assistants' Pedagogical Practices: Content Validation of a Survey Informed from Principles of the “How People Learn” Framework. In Greta Gorsuch (Ed.), *Working Theories for Teaching Assistant and International Teaching Assistant Development*. Stillwater, OK: New Forums Press.
- ⁷Cox, M.F., Hahn, J., McNeill, N., Cekic, O., Zhu, J., & London, J. (2011), Enhancing the Quality of Engineering Graduate Teaching Assistants through Multidimensional Feedback, *Advances in Engineering Education*. 2 (3)
- ⁸Cox and Cordray, 2008 Assessing pedagogy in bioengineering classrooms: Quantifying elements of the “How People Learn” model using the VaNTH Observation System (VOS) *Journal of Engineering Education*, 97(4), 413-431
- ⁹Cohen, J.A. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-45.