Adapting Best Practices for Peer Assessment in Army Training

Elizabeth R. Uhl
U.S. Army Research Institute
Fort Benning, GA
elizabeth.r.uhl.civ@mail.mil

Ronelle L. Koschny
Consortium of Universities of Washington
Fort Benning, GA
ronelle.l.koschny.ctr@mail.mil

Tatiana H. Toumbeva, Kristy Reynolds, Joshua Shireman
Aptima, Inc.
Woburn, MA
ttoumbeva@aptima.com, kreynolds@aptima.com, jshireman@aptima.com

Ashley Wittig, Celeste Sanders
U.S. Army Research Institute
Fort Benning, GA
ashley.h.wittig.civ@mail.mil,
celeste.n.sanders.civ@mail.mil

Frederick J. Diedrich
Independent Consultant
Groton, MA
Frederick.diedrich@gmail.com

Scott M. Flanagan
Sophia Speira, LLC
Carthage, NC
scott@sophiaspeira.com

ABSTRACT

The current study leveraged best practices in peer assessment and provision of feedback, as well as end user input, to improve the effectiveness of peer assessments in a junior leader course. Relative to a legacy process that focused primarily on collecting rankings and general comments, revisions focused on providing clear guidance for evaluation criteria in the form of behaviorally anchored rating scales for critical leadership attributes. This promoted the provision of constructive peer feedback within the framework of the Army's Leader Requirements Model. To examine the effectiveness of the revisions, the administration of revised and legacy peer assessments were counterbalanced at Times 1 and 2, and all students completed the revised peer assessment at Time 3. Students provided feedback on the assessment they completed at Times 1 and 2. After completing the assessment at Time 3, students completed a survey comparing the utility and usability of the two versions of the assessment. Though the legacy assessment was generally rated as easier and quicker to complete, the revised process was rated as more useful for student development. Furthermore, cadre evaluated the revised process better than the legacy process on all characteristics. The implications for student development and areas for future research are discussed as well as recommendations for how to more effectively use peer assessments in support of Soldier development.

Controlled by: DAPE-ARI
CUI Category: Unclassified
Distribution Statement: A
POC: Jennifer S. Tucker, 706-366-7312

ABOUT THE AUTHORS

Elizabeth R. Uhl, PhD is a Research Psychologist at the U.S. Army Research Institute, Fort Benning Research Unit. Her research focuses on training development, including the use of peer assessments for formative and summative feedback.

Ronelle L. Koschny, MEd is a Research Assistant at the U.S. Army Research Institute, Fort Benning Research Unit, where she assists with data collection, research analysis, and program design.

Tatiana H. Toumbeva, PhD is a Senior Scientist in the Learning and Training Systems Division at Aptima with expertise in assessment tool development and validation, training evaluation, and occupational health psychology.

Kristy Reynolds, MS is a Scientist in the Learning and Training Systems Division at Aptima with expertise in training needs analysis, competency modeling, training design and evaluation, and performance assessment.

Joshua Shireman, MEd is a Software Engineer in the Performance Augmentation Systems Division at Aptima with expertise in distributed systems and software for observational measurement**.**

Ashley Wittig, PhD is a Research Psychologist for the U.S. Army Research Institute, Fort Benning Research Unit. Her work has focused on instructor development.

Frederick J. Diedrich, PhD is a consultant who focuses on methods of instruction and assessment designed to deliberately support competency and attribute development.

Scott M. Flanagan is a retired Army Master Sergeant having served his career in USASOC now providing operational consulting focused on purposeful development and assessment of leader competencies and attributes.

Celeste Sanders, **PhD** is a Research Psychologist in the Army Research Institute, Fort Benning Research Unit. Her work has focused on leadership development and assessment.

Adapting Best Practices for Peer Assessment in Army Training

Elizabeth R. Uhl
U.S. Army Research Institute
Fort Benning, GA
elizabeth.r.uhl.civ@mail.mil

Ronelle L. Koschny
Consortium of Universities of Washington
Fort Benning, GA
ronelle.l.koschny.ctr@mail.mil

Tatiana H. Toumbeva, Kristy Reynolds, Joshua Shireman
Aptima, Inc.
Woburn, MA
ttoumbeva@aptima.com, kreynolds@aptima.com, jshireman@aptima.com

Ashley Wittig, Celeste Sanders
U.S. Army Research Institute
Fort Benning, GA
ashley.h.wittig.civ@mail.mil, celeste.n.sanders.civ@mail.mil

Frederick J. Diedrich
Independent Consultant
Groton, MA
Frederick.diedrich@gmail.com

Scott M. Flanagan
Sophia Speira, LLC
Carthage, NC
scott@sophiaspeira.com

INTRODUCTION

Peer assessments are generally defined as the evaluation of a person of equal standing that is being made regarding their behavior, learning outcome, or level of task execution (Downey & Duffy, 1978; Zazanis et al., 2001). Peer assessments are used across a variety of settings (e.g., military, educational, industrial) for both summative purposes and formative feedback (O'Donnell & Topping, 1998; Reinholz, 2016; Zazanis et al., 2001). There are several challenges to effectively conducting peer assessments in military contexts. For example, though peer assessments are used widely throughout the military, the goal of the peer assessments may not always match the peer assessment method used. Further, students may not take peer assessments seriously, especially if they have incentives to try to manipulate the system or if they don't understand the purpose of completing the assessments. Peer assessments can also be time-consuming, particularly when it comes to compiling the results. The purpose of this research was to address these concerns and to develop a peer assessment for a junior leader course that is efficient, useful, and sustainable.

In order to identify methods for revising the peer assessment, focus groups were conducted with cadre and students to obtain insight into the existing peer assessment practices in the course (i.e., legacy peer assessments). Reoccurring themes were identified in the focus group feedback. This feedback was used to develop a peer assessment process that reinforces key leadership attributes and competencies in junior leaders, while also addressing larger programmatic needs. The effectiveness of the revised peer assessment was examined in Study 1 and a digital version of the revised peer assessment was examined in Study 2.

Identified Limitations and Research-Based Recommendations for Peer Assessment Process

Prior to this research, peer assessments were conducted in the junior leadership course on three occasions spaced throughout the course. The first two peer assessments were conducted at the squad level (e.g., groups of 8-13) and the final peer assessment was conducted at the platoon level (e.g., groups of approximately 30-50). Each individual was asked to rank the other members of their squad or platoon from best performer to worst performer. Individuals were also asked to provide comments and feedback for each person they ranked, and to answer a trust question, such as, "Would you go to war with this person?" Typically, peer assessments were completed on the students' own time.

Cadre and students participated in focus groups and provided feedback on the strengths and limitations of the existing peer assessment process (henceforth referred to as the legacy peer assessment process). There were several limitations that were identified by both cadre and students.

One concern was that peer assessments were too subjective or could reflect an individual's popularity rather than performance. Students were not provided with clear criteria on which to evaluate each other. Therefore, ratings were perceived as being influenced by how likeable or popular an individual was. Further, since the peer assessments counted towards course grades, which could impact an individual's career beyond the course, there were concerns that the system was sometimes manipulated by groups of individuals to improve their scores (gaming). A related concern was about whether the peer assessments should count towards the course grade and standing, though there was not complete agreement on this issue among the cadre or the students – some were concerned that counting the peer assessment towards the course grade encouraged gaming while others were concerned students wouldn't take the assessment seriously if it wasn't counted towards their grade. In addition, cadre were concerned that the comments written by students often lacked constructive and/or actionable feedback, that the platoon-level assessment had limited utility, and that data management was time consuming. Other concerns noted by students included that it was difficult to meaningfully rate their entire platoon, that there was insufficient guidance on what criteria to use, and that students don't always take the peer assessments seriously.

These findings led to several research recommendations. In response to concerns that peer assessments were too subjective and that students weren't given sufficient guidance on what evaluation criteria to use, the first recommendation was to include a rubric based on the Leadership Requirements Model (LRM; U.S. Department of the Army, 2019) as guidance for evaluation. Peer assessment researchers have identified setting clear and understandable criteria as an important component of successful peer assessments (e.g., Dochy et al., 1999). The LRM outlines attributes (Character, Presence, and Intellect) and competencies (Leads, Develops, and Achieves) which provide a common framework that describes the fundamental traits desired for all leadership levels and cohorts. Each attribute and competency includes sub-attributes or sub-competencies (see Figure 1). The LRM rubric was created for the junior leader course and includes Behaviorally Anchored Rating Scales (BARS) for each sub-attribute and sub-competency in the LRM written at a level appropriate for this course (Toumbeva et al., 2018). This recommendation was implemented.
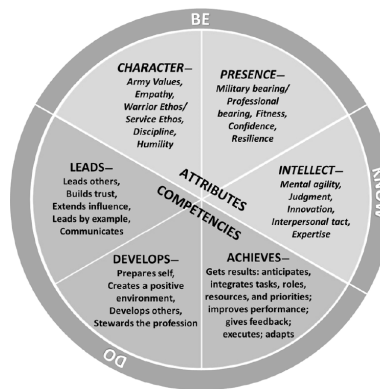

Figure 1. The Army Leadership Requirements Model (U.S. Department of the Army, 2019).

In response to concerns about students' ability to meaningfully rate the members of their platoon, the second recommendation was to drop the final platoon level peer assessment in order to reduce instances of rating individuals the rater did not know well. A variation of this recommendation was implemented, and the final peer assessment was changed from a platoon-level assessment to a squad-level assessment.

The third recommendation was to drop the peer assessment from the course grade in order to reduce gaming. This recommendation was not implemented for this study, though it was later implemented by the course. The inclusion of the peer assessment in the course grade could incentivize individuals who were particularly concerned about their grades to work with others to manipulate their ratings in order to improve their standing in the course. Farh et al.

(1991) found that when students were told that peer assessments were for developmental purposes, they were less lenient, and more differentiating, reliable, and valid than when students were told the peer assessments were for evaluative purposes. Though this study did not examine gaming specifically, it does suggest that peer assessments may be more accurate under developmental conditions than evaluative conditions.

The use of peer rankings along with comments was considered appropriate for the course goals for the peer assessment. There are three main methods of conducting quantitative peer assessments: rankings, nominations, and ratings (Kane & Lawler, 1978). In peer rankings, an individual ranks everyone in their peer group from best to worse on a specified characteristic. For nominations, an individual identifies the top (and/or bottom) *n* number of performers on the specified characteristic. Finally, for peer ratings, an individual rates everyone in their peer group on a Likert-type scale on several identified characteristics. Peer rankings and peer nominations have demonstrated better reliability and validity than peer ratings (Kane & Lawler, 1978; Love, 1981). However, peer nominations are limited because individuals who are not top or bottom performers, may not be assessed. Therefore, peer rankings were considered the quantitative method best suited for this course.

Liu and Carless (2006) made a distinction between peer assessment, a quantitative approach like those described above, and peer feedback, a qualitative approach that includes providing rich comments about performance. Quality peer feedback can help improve student performance (e.g., Gielen et al., 2010). Research has demonstrated that peer feedback can improve performance (e.g., Xiao & Lucking, 2008) through several possible mechanisms, including increased time on task, task engagement, and practice (Topping et al., 2000). There is also evidence that providing peer feedback can also lead to learning by improving knowledge building and better understanding of the criteria (Cho & Cho, 2011).

STUDY 1

The revised peer assessment process was developed based on the findings from the literature review, focus groups, and guidance from course leadership. Study 1 examined the efficacy of the revised peer assessment in comparison to the legacy peer assessment for the junior leader course. Qualitative and quantitative data were gathered on the utility and usability of each peer assessment for development.

Methods

Participants
Participants included 126 students from a junior leader course. Peer assessments were a required component of the course, but use of student course data and completion of the other components of the study were voluntary. The average reported age of the students was 27.09 years ($SD = 3.54$) and average time in service was 29.67 months ($SD = 39.37$ months; minimum = 3 months, maximum = 153 months). Of the sample, 65.7% reported their gender as male and 23.4% reported their gender as female. Some participants did not report their gender (10.9%). The highest level of education for the participants was Bachelor's degree (82.5%) and Master's degree or higher (16.7%). A few participants did not report their level of education (0.8%).

Materials
Legacy Peer Assessment. The legacy peer assessment process included two squad-level peer assessments (Times 1 and 2) and a platoon-level peer assessment (Time 3). However, for the purposes of this project, only the squad-level peer assessment was used. The legacy peer assessment required students to rank the members of their squad or platoon in order of performance and to write a comment about each peer without specific evaluative guidance. The legacy peer assessment also included a question about the degree to which each peer was trusted ("Would you go to war with this person?").

Revised Peer Assessment. The revised peer assessment process included three squad-level peer assessments. In addition to rating each member of their squad, students were asked to identify at least 1 and up to 3 strengths and weaknesses based on the LRM rubric and to write a comment about each strength and weakness they identified. Then they were asked to answer an accountability question (i.e., what have you done/could you do to help your peer improve?). The revised peer assessment included a revised trust question, "Would you trust this Soldier as your leader in combat?" The revised peer assessment included clear instructions on how to evaluate their peers: "For each other person in your squad, please identify <u>at least 1 and up to 3</u> attributes/competencies from the provided Leadership

Rubric as *Sustains* and *Improves*. For each *Sustain*, describe why you view the attribute/competency as a strength for the individual. For each *Improve*, explain why you identified the attribute/competency as needing improvement, and how the individual might be able to improve. Please write comments that are specific, concrete, constructive, and actionable, especially for the *Improves*."

LRM Rubric. Paper copies of the LRM rubric were provided with the revised peer assessment forms at the time of the peer assessment. The LRM rubric includes Behavioral Anchored Ratings Scales (BARS) for each of the leader attributes and competencies on a scale from needs improvement to outstanding for each of the attributes and competencies included in the Army's LRM (see Table 1 for an excerpt).

Table 1. Excerpt from LRM Rubric, BARS for Military and Professional Bearing

Needs Improvement	Satisfactory	Excellent	Outstanding
• Fails to have uniform squared away; has poor hygiene • Consistently fails to follow appropriate customs and courtesies; does not adhere to Army standard • Fails to convey information concisely, clearly, and logically; hesitates, pauses, and self-corrects to the point of being distracting	• "Looks the part" of a Soldier (e.g., cleanly shaven; clean haircut, appropriate uniform) • Follows basic customs and courtesies • Communicates clearly but has shaky voice, stumbles over words, or looks at the ground/notes when speaking	• Consistently follows customs and courtesies; adheres to Army standards • Communicates calmly and effectively	• Models appropriate customs and courtesies, even when not in the spotlight • Communicates calmly and effectively while motivating/energizing others, even when under stress • Explains to peers the implications for looking the part (e.g., that personal appearances reflect on the Army)

End-of-Cycle Survey. At Time 3, students were asked to complete a questionnaire that compared the legacy and revised peer assessments in terms of usefulness, effectiveness, and other characteristics. Students identified which peer assessment version they preferred based on different utility and usability criteria. The 7 point scale offered response options that ranged from (1) [legacy peer assessment] much better, (2) better, (3) slightly better, (4) same, (5) [revised peer assessment] slightly better, (6) better, (7) much better. The criteria students rated can be found in Table 2. Students were also asked to rate their agreement with statements about the different components of the peers assessment, such as the rubric, the self-assessment, and the accountability question (e.g., "The leadership rubric helped me to more meaningfully evaluate my peers") on a scale from 1 (strongly disagree) to 5 (strongly agree). A similar survey was completed by cadre after the course conclusion.

Procedure

Peer assessments were conducted three times during a 12 week course: week 4 (Time 1), week 9 (Time 2), and week 11 (Time 3). In order to ensure all students experienced both versions of the peer assessments, the delivery of the legacy and revised peer assessment versions were counterbalanced so that two of the four platoons completed the legacy peer assessment, while the other two completed the new peer assessment at Time 1. At Time 2 platoons completed the alternate version of the peer assessment. All four platoons received the revised peer assessment at Time 3, followed by a request to complete a survey comparing the revised and legacy peer assessments. After the course concluded, cadre provided their feedback regarding comparison and utility of the revised and legacy peer assessment processes, including considerations for further improvements. Though historically the legacy peer assessment was completed on the students' own time, for the purposes of this study, both peer assessments were completed during dedicated class time.

Results

The results of the comparison portion of the end-of-cycle survey completed at Time 3 are shown in Table 2 (select results are displayed due to space constraints). Students responded on a 7-point Likert scale: for ease of interpretation and presentation, results were condensed into three categories: legacy was better, revised was better, or both versions were the same. As can be seen in Table 2, students tended to rate the legacy peer assessment as easier to complete and more efficient. The only item without a significant difference between groups was "taken seriously," suggesting that the version of the peer assessment didn't appear to impact how seriously students took the assessment. The revised

peer assessment was rated as better on the remaining characteristics, including prompting students to think deeply about leadership attributes and competencies, being useful for students, providing clearer evaluation criteria, overall effectiveness, and providing actionable feedback.

Table 2. Comparison Survey Student Results

Characteristic	Legacy PA		Same		Revised PA		χ^2 (2)
	n	%	*n*	%	*n*	%	
Easy to complete.	79	71.20	7	6.30	25	22.50	75.89**
Efficient.	70	63.10	7	6.30	34	30.60	54.00**
Helped uncover my strengths and weaknesses.	14	12.50	35	31.30	63	56.20	32.38**
Helped me understand the purpose of PAs at [junior leader course].	11	9.80	43	38.40	58	51.80	30.88**
Prompted me to think deeply about leadership attributes/competencies.	12	10.60	23	20.40	78	69.00	66.39**
Useful for cadre.	11	9.90	38	34.20	62	55.90	35.19**
Useful for students.	15	13.30	24	21.20	74	65.50	53.65**
Overall effectiveness.	17	15.00	25	22.10	71	62.90	45.10**
Provided clearer evaluation criteria.	19	16.80	20	34.50	74	65.50	52.58**
Provided me with actionable feedback.	10	8.80	36	31.90	67	59.30	43.24**

Note. N ranges from 111-113 due to missing data. Expected *N* per response option = 37. PA = peer assessment. ** $p < .001$. * $p < .01$.

Students also provided feedback on the specific components of the peer assessments (e.g., rubric, self-assessment, and ranking) in the end-of-cycle survey. On average, students tended to agree that "the leadership rubric helped me to more meaningfully evaluate my peers" ($M = 4.10$, $SD = 0.88$) and that "the peer assessments made me think more deeply about critical leadership attributes and competencies" ($M = 4.06$, $SD = 0.83$).

Cadre also completed a survey comparing the legacy and revised peer assessment after the course was complete. The cadre did not rate the legacy peer assessment as better on any category, though there were some cadre that rated the two assessment types as the same, the revised peer assessment was rated as better on most characteristics (e.g., overall effectiveness, provided clearer evaluation criteria, useful for cadre, useful for students, helped students improve as leaders, and enabled students to provide actionable feedback).

Discussion

Overall, the legacy process was rated less time consuming, but was rated poorly for helping to provide actionable feedback. Meanwhile, the revised peer assessment process was rated highly on most other characteristics of interest, including usefulness and overall effectiveness. Students rated the leadership rubric that accompanied the revised peer assessment highly, but qualitative feedback from the end-of-cycle survey suggested it was time consuming to digest during the peer assessment session. Students and cadre alike recommended digitizing the new peer assessment to exploit its benefits, while addressing the issues of time and fatigue. The findings from Study 1 informed the development of an Excel-based peer assessment. This tool was iteratively tested and revised in Study 2.

STUDY 2

Feedback provided by students and cadre on the paper-based format of the revised peer assessment used in Study 1 suggested it was time-consuming and tedious to use. The goal of Study 2 was to digitize the revised peer assessment, making data management less tedious for cadre and students, while standardizing information gathered in the peer assessments for more actionable feedback.

Methods

Participants
The digital peer tool was tested by a class of approximately 120 junior leaders. Demographic data were not collected.

Materials
Digital Peer Assessment Tool. The digital peer assessment tool was built in Excel. Excel was chosen because it is readily available and approved for use on the Army network. See Figure 2 for a screenshot of the peer assessment tab in the digital tool. This tool included the elements of the revised peer assessment developed and evaluated in Study 1. Students were asked to identify at least one and no more than three sustains and improves for each member of their squad based on the LRM.

Figure 2. Peer Assessment Tab of Digital Peer Tool

The LRM rubric was included as a reference in the tool. However, instead of accessing the entire rubric at once, students could click on an attribute or competency of interest, and the rubric for that attribute or competency would be displayed (see Figure 3 for an example). Students were also asked to rank the members of their squad in order of overall performance and to answer the trust and accountability questions.

Figure 3. Example of Integrated LRM rubric

Utility and Usability Survey. Students were asked to rate their level of agreement with several statements regarding the utility and usability of the digital peer tool (e.g., "The peer assessment tool was easy to use") on a five point scale from 1 (strongly disagree) to 5 (strongly agree).

Procedure

The digital tool was used during a course to collect peer assessment data on three occasions. The Time 1 peer assessment was conducted one platoon at a time so that more researchers would be able to troubleshoot with individual students as issues arose. Time 2 and 3 peer assessments were conducted with the full company (e.g. all four platoons). As in Study 1, students were asked to assess their peers on their squad at all three time points. Feedback was gathered throughout the process and updates were made as issues were identified. Peer assessments were conducted using the digital tool at approximately week 4 (Time 1), week 9 (Time 2), and week 11 (Time 3) of the 12-week junior leader course. After Time 3, students were invited to complete the utility and usability survey. After the completion of the course, company leadership and cadre were asked how the digital peer assessment tool or process could be improved for long-term sustainability within the junior leader course and extend to larger programmatic needs.

Results

Fifty-three students completed the utility and usability survey at the end of Time 3 (see Table 3; results with means greater than or equal to 4.00 were included in the table due to space constraints). The tool was generally rated favorably with ratings on all items above three (the neutral point on the scale). In particular, students tended to agree that they could quickly learn how to use the tool ($M = 4.43$, $SD = 0.64$); that they could easily remember how to use the tool ($M = 4.40$; $SD = 0.74$); and that the content of the tool was easy to understand ($M = 4.30$, $SD = 0.77$). Students and cadre also provided verbal feedback that further supports the findings of an overall positive experience with the digital peer assessment. Additional positive feedback included usability of the digital tool, primarily the ability to click on the attributes and competencies within the digital tool to expand the definitions within the embedded rubric.

Table 3. Digital Tool Utility and Usability Survey Results

Survey Items	Mean	*SD*
1. I quickly learned how to use the peer assessment tool.	4.43	0.64
2. It was easy to remember how to use the peer assessment tool.	4.40	0.74
3. The tool content was easy to understand.	4.30	0.77
4. The integrated instructions were helpful.	4.25	0.84
5. The integrated leadership rubric was helpful.	4.23	0.85
6. It was easy to find what I was looking for in the tool.	4.23	0.85
7. I felt confident using the peer assessment tool.	4.21	0.95
8. The tool content was easy to read.	4.17	0.92
9. The steps I had to follow to complete the assessment were simple.	4.17	0.94
10. I knew exactly where to go within the tool to complete an assessment.	4.17	0.99
11. It was easy to navigate between the different sections of the tool.	4.13	0.86
12. The peer assessment tool was easy to use.	4.40	0.74
13. The tool interface was organized well.	4.08	0.98
14. The tool interface was user friendly.	4.04	1.07
15. It was easy to tell if I missed a step in the assessment process.	4.00	0.86

Note. Item responses ranged from 1 (strongly disagree) to 5 (strongly agree); SD = standard deviation.

Most of the students (n = 47) who completed the survey also responded to the open-ended question at the end of the survey. Students noted that the tool was easy to use and the process easy to follow (n = 23) and that the integrated LRM rubric was helpful for evaluating peers (n = 15).

GENERAL DISCUSSION

The research presented here describes the successful development of a useful and sustainable peer assessment process, with an efficient collection and delivery system for providing feedback. Additionally, the peer assessment process supports student growth while meeting larger programmatic needs, such as the need for a graded component. The survey data obtained through various phases of the research process suggests that the revised peer assessment process is an improvement over the legacy process in terms of both utility and usability. Though this study focused on a junior leader course, the improvements and recommendations modeled in this study can be used to develop tools and processes to support peer assessment processes in other courses or training contexts.

In Study 1, this project leveraged multiple research findings to revise the existing peer assessment process. Peer assessments that are standardized can provide clear guidance on the purpose of the peer assessments and the relevant evaluation criteria. Using a rubric (e.g., BARS) with the peer assessments can enable students to produce constructive, actionable, and useful feedback that is consistent with course success criteria. Using the LRM rubric in this study, students were able to select and rate their peers on attributes and competencies the Army has identified as important for leaders.

The revisions to the legacy peer assessment in Study 1 used a paper-based format to deliver the revised peer assessment and accompanying LRM rubric to students. Survey data suggested that the changes to the legacy peer assessment process were well received by cadre and students, with some identified areas of improvement. A digital version was recommended by cadre and students alike to address concerns of time, fatigue, and rushed peer feedback comments. An Excel-based version of the peer assessment tool was developed and evaluated in Study 2.

Study 2 focused on the development and testing of an Excel-based peer assessment tool for a junior leader course that would be compatible with the Army network, and would meet larger instructional and programmatic needs. Study 2 focused on digitizing the revised paper-based peer assessment forms and embedding the LRM rubric within the assessment criteria. The digital version of the revised peer assessment expedited the collection and synthesis of peer data. The digital peer assessment tools and process were iteratively tested and refined at all stages of the research to maximize their utility and sustainability. Feedback on the utility and usability of the digital tool was generally positive, though with further testing, refinements can be made to the tool.

There are several areas for future research and development. One area is to refine the LRM rubric to be more user friendly for students. Students found the rubric to be quite long, which made it difficult to review during the time allotted for the peer assessment. Further, the rubric included BARS at the excellent and outstanding levels, which were difficult for students to discriminate between. Revisions to the rubric could include shortening it and reducing to three levels of performance. Another potential area for future research is to examine methods for providing useful formative feedback to students in the course.

This project addressed several challenges to effectively conducting peer assessments in the military context. First, the goal of the peer assessment is clearly explained to students and students were provided with the LRM rubric for guidance on evaluation criteria. The use of the LRM rubric to provide clear evaluation guidance was well-received by both students and cadre. The impact of reducing incentives to game the system by using peer assessment only for developmental purposes has not been examined yet as that recommendation was implemented after the period covered by this study. Digitizing the peer assessment has reduced the amount of time required to complete the peer assessment as well as the amount of time required to compile the results. Together, the findings of this research suggest that best practices for peer assessment can be incorporated into military peer assessments in classroom settings.

REFERENCES

Cho, Y. H., & Cho, K. (2011). Peer reviewers learn from giving comments. *Instructional Science, 39*, 629-643.

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education (24)*3, 331- 350.

Downey, R. G., & Duffy, P. J. (1978). *Review of Peer Evaluation Research.* (ARI Technical Paper 342). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA 061780).

Farh, J. L., Cannella, A. A., & Bedeian, A. G. (1991). The impact of purpose on rating quality and user acceptance. *Group & Organization Studies, 16*(4), 367-386.

Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction, 20*, 304-315.

Kane, J. S., & Lawler, E. E. (1978). Methods of peer assessment. *Psychological Bulletin, 85*(3)*, 555-586.

Liu, N., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education, 11*(3), 279-290.

Love, K. G. (1981). Comparison of peer assessment methods: Reliability, validity, friendship bias, and user reaction. *Journal of Applied Psychology, 66*(4), 451-457.

O'Donnell, A. M., & Topping, K. (1998). Peers assessing peers: Possibilities and problems. In K. Topping, & S. Ehly (Eds.), *Peer-Assisted Learning (*pp.255-278). Mahwah, NJ: Lawrence Erlbaum Associates.

Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment & Evaluation in Higher Education, 25*(2), 149 – 169.

Toumbeva, T. H., Ratwani, K. L., Diedrich, F. J., Flanagan, S. M., & Uhl, E. R. (2018). *Development of a behaviorally anchored rating scale for leadership* (ARI Research Product 2018-06). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. AD 1048729)

U.S. Department of the Army. (2019). Army Leadership, (ADP 6-22).

Xiao, Y., & Lucking, R. (2008). The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment. *Internet and Higher Education*, *11*, 186-193.