7-2016

Pilot Testing the Debriefing for Meaningful Learning Evaluation Scale

Cynthia Sherraden Bradley
University of Minnesota

Kristina Dreifuerst
Marquette University, kristina.dreifuerst@marquette.edu

Clinical Simulation in Nursing, Vol. 12, No. 7 (July 2016): 277-280. DOI.

Pilot Testing the Debriefing for Meaningful Learning Evaluation Scale

Cynthia Sherraden Bradley
Doctoral Candidate, School of Nursing, Indiana University School of Nursing, Indianapolis, IN
Kristina Thomas Dreifuerst
Assistant Professor, School of Nursing, Indiana University School of Nursing, Indianapolis, IN

Abstract

Background

Debriefing for Meaningful Learning (DML), an evidence-based debriefing method, promotes thinking like a nurse through reflective learning. Despite widespread adoption of DML, little is known about how well it is implemented. To assess the effectiveness of DML implementation, an evaluative rubric was developed and tested.

Sample
Three debriefers who had been trained to use DML at least 1 year previously, submitted five recorded debriefings each for evaluation.

Methods
Three raters who were experts in DML scored each of the 15 recorded debriefing session using DML Evaluation Scale (DMLES). Observable behaviors were scored with binary options. These raters also assessed the items in the DMLES for content validity.

Results
Cronbach's alpha, intraclass correlation coefficients, and Content Validity Index scores were calculated to determine reliability and validity.

Conclusion
Use of DMLES could support quality improvement, teacher preparation, and faculty development. Future testing is warranted to investigate the relationship between DML implementation and clinical reasoning.

Keywords
DML, debriefing, effective briefing, debriefing evaluation, measurement

Key Points
- The DMLES was developed to measure the presence of DML debriefing behaviors.
- Evaluation of debriefing practice is critical to ensuring learning outcomes.
- Pilot testing of the DMLES demonstrated internal consistency and validity.

Debriefing has been widely recognized as the most significant component of a simulation learning experience (Shinnick, Woo, Horwich, & Steadman, 2011). Although there is an agreement that mastery of debriefing skills is crucial to facilitating student learning (Neill & Wotton, 2011), measurement of this mastery is lacking. Although aspects of debriefing have been marginally evaluated (Fey, 2014, Waznonis, 2015), there are no reports of a psychometrically tested instrument of a debriefer's ability to adhere to a specific structured debriefing method. Therefore, a formative, evaluative, behaviorally anchored scale, the Debriefing for Meaningful Learning Evaluation Scale (DMLES), was developed based on the framework of Debriefing for Meaningful Learning© (DML) to evaluate the ability of a debriefer to implement this method.

Background
A skilled debriefer plays an essential role in a learner's reflective thinking process during debriefing (Decker et al., 2013). A debriefer guides learners in discussing and deepening their learning while facilitating reflection on decisions, actions, and emotions occurring during the learning experience. Because debriefing enhances the transfer of knowledge, skills, and attitudes to nursing practice, it is crucial that the debriefer is equipped through formal training and competency assessment in a

structured, theory-based debriefing method (Decker et al., 2013, Hayden et al., 2014). Yet, training in debriefing practice is reportedly limited in both quantity and quality, with even less reported competency assessment of debriefing practice (Fey, 2014, Waznonis, 2015).

Debriefing has become an increasingly common pedagogical method in which teaching can be evaluated. Traditionally, evaluation of the debriefing process has focused on examining elements of debriefing such as duration, environment, faculty experience, and learning objectives (Waznonis, 2015). The following four tools have been reported for evaluation of the debriefing experience: a tool developed by Gururaja, Yang, Paige, & Chauvin (2008) measures broad aspects of debriefing effectiveness, the Debriefing Assessment for Simulation in Healthcare© measures debriefing quality across disciplines and simulation learning environments (Simon, Raemer, & Rudolph, 2009), the Objective Structured Assessment of Debriefing measures debriefing quality (Arora et al., 2012), and the Debriefing Experience Scale is a participant self-reporting of individual learning and perception in debriefing (Reed, 2012). Absent from these reports of debriefing evaluation, however, is a measure of the debriefer's ability to consistently implement a specific structured debriefing method.

As simulation pedagogy is integrated increasingly within nursing education, psychometric reports of reliable and valid instruments are needed to evaluate the effectiveness of each component of the teaching and learning experiences (Adamson, Kardong-Edgren, & Willhaus, 2013). Specifically, evaluation of the debriefer's ability to engage students in a structured, theory-based debriefing is critical, as the practice of debriefing methods broadens throughout nursing curriculum. A structured debriefing method widely adopted across nursing education is DML. However, little is known regarding how well DML is being implemented by debriefers. Because of DML's increased use and subsequent reports of improved clinical reasoning in nursing students with the use of this method (Dreifuerst, 2012, Mariani et al., 2013), DML was chosen as the framework for development of a behaviorally anchored rating scale to assess a debriefer's adherence to this structured debriefing method.

Through guided reflection and Socratic questioning, DML supports the development of clinical reasoning among learners and the translation of thinking into clinical decisions and actioned decisions in future clinical situations (Dreifuerst, 2012). Learners are engaged in purposeful reflection, while empowered to challenge assumptions and uncover relationships between thinking processes and nursing actions. DML is based on six elements that facilitate distinct, yet integrated, thinking processes. These elements are: engage, explore, explain, elaborate, evaluate, and extend.

The purpose of this pilot study was to test if the DMLES measures a debriefer's ability to implement this specific debriefing method. The research questions guiding testing were: (a) does the DMLES demonstrate internal consistency? (b) does the DMLES demonstrate interrater reliability? (c) does the DMLES demonstrate face validity? (d) does the DMLES demonstrate content validity? and (e) can a debriefing be scored with the DMLES without observation of the simulation experience?

Sample

Three debriefers teaching within a Midwestern prelicensure nursing program were purposively solicited to submit debriefings for review based on the following criteria: each received training in DML by the developer of DML in conjunction with the National Council of State Boards of Nursing National Simulation Study, and each facilitated debriefing with prelicensure nursing students.

Method

During regularly scheduled simulation learning experiences, students and debriefers participated in a 20-minute simulation, followed by a 30-minute debriefing per the customary process within the program. Each debriefing was recorded; each of the three faculty submitted five recorded debriefings to be scored, for a total of 15 recorded debriefing sessions. Three expert debriefers with extensive DML experience were asked to individually and privately score 15 recorded debriefings with the DMLES after receiving training in use of the scale. Training included review of all components of DML, and instruction regarding the observational approach used to collect direct information about a debriefer's teaching behaviors (Dowdy, Twyford, & Sharkey, 2013). The observable behaviors of each of the six elements of the DML method were structured as 33 distinct items of the DMLES, allowing for independent rating of each item without overlap, redundancy, or ambiguity. The behavioral anchors of the scale include the iterative steps of the DML method; each item was scored with two binary options: (a) *present*, or (b) *not present*.

Results

After scoring was completed, performance of each item was evaluated by entering the data from the 45 completed scales (3 raters × 15 recordings) into Statistical Package for Social Scientists (SPSS) version 22 (IBM Corp, 2013). To assess the internal consistency of the scale, Cronbach's alpha was calculated. Cronbach's alpha has been widely used throughout the social sciences as an estimate of the internal consistency of a psychometric test. Internal consistency is considered excellent for a Cronbach's alpha ≥0.90, between 0.80 and 0.90 is good consistency, and between 0.70 and 0.80 is an acceptable level of consistency (Tabachnick & Fidell, 2013). Cronbach's alpha for the total DMLES scale was 0.88. Cronbach's alpha for each element of the scale was: engage (0.39), explore (0.51), explain (0.73), elaborate (0.79), evaluate (0.78), and extend (0.70). Items of zero variance within the subscales *engage* and *explore* were removed from the computation, yielding the low values. Theoretically, these subscales demonstrate greater internal consistency than reflected in the numerical data.

One-way random effects intraclass correlation coefficients (ICCs) were calculated for the total scale and each of the 33 items to determine item reliability and interrater reliability (Shrout & Fleiss, 1979). The ICC is a descriptive statistic used when quantifying units organized into groups; the ratio of variance between groups to variance within those groups is used to determine interrater reliability. Values from 0.40 to 0.75 are considered fair to good, whereas values >0.75 are recommended in health research. The ICC for the total scale was 0.86 ($p < .01$). The ICC for each phase of the DMLES was as follows: engage (0.27), explore (0.37), explain, (0.69), elaborate (0.79), evaluate (0.73), and extend (0.70). The values <0.40 for the subscales *engage* and *explore* are attributed to the removal of zero variance items and can be theoretically interpreted as demonstrating a higher level of reliability.

Face validity of the DMLES was determined by the developer of DML and two additional DML experts through a questionnaire, with subsequent item and total scale revision. All agreed that the scale was representative of the DML method. Each expert was asked to establish content validity by ranking the relevance of each item on the DMLES as it pertained to DML.

Content validity is the degree to which a group of items represents an operational definition of a construct. Content validity is quantified through the content validity index (CVI), numerically representing the content relevance of instrument items using a four-point rating scale, as rated by construct experts. An acceptable level is determined as an index of 0.80 or greater. The item-level CVI (I-CVI) of the DMLES was scored by three experts and subsequently computed as 1.0 for 20 of the 33 items, indicating that these 20 items were rated as highly relevant by all experts. The I-CVI of 8 of the 33 items was computed as >0.80, whereas the I-CVI of 5 of the 33 items was computed as <0.80. One of these five items was eliminated, and four were revised to increase clarity. The scale-level CVI mean score was computed as 0.92, by dividing the sum of I-CVIs by the total items.

A questionnaire was completed by the raters after scoring of the recordings to assess if a debriefing can be scored with the DMLES without observation of the simulation portion of the experience. There was agreement among the raters that each item could be scored with observation of the debriefing but not the actual simulation, because the purpose of scoring with the DMLES is to assess the debriefer, irrespective of scenario or program level of participants. That is, rating each item as *present* or *not present* did not require knowledge of the objectives of the simulation.

The testing of the DMLES revealed several limitations. Testing of the scale across multiple sites would demonstrate if the scale's psychometric properties could be replicated, as common to instrument validation. In addition, the small sample size and small number of recorded debriefing sessions may not be representative of all debriefers using DML. Further validity testing is needed to ensure that each item of the scale demonstrates the constructs embedded in DML.

Conclusion

This purpose of this study was to test the DMLES, a newly developed scale behaviorally anchored in the iterative process of DML. Each of the six elements of DML includes debriefer behaviors that guide learners through reflective thinking, facilitating sensemaking of a learning experience. In this pilot test, the DMLES demonstrated internal consistency (Cronbach's alpha = 0.88), interrater reliability (0.86, total scale ICC [$p < .01$]), face validity, content validity (scale-level CVI 0.92), and the ability of a rater to behaviorally score a debriefing without observing the simulation.

Testing of this scale is an initial attempt to assess how consistently DML is implemented by a trained debriefer. Further research could advance debriefing practice by investigating the relationship between a debriefer's implementation of DML, and the development of learners' clinical reasoning, an established outcome of DML use. This scale could support quality improvement in nursing programs by evaluating the effectiveness of debriefers engaging learners in a specific, commonly used, structured debriefing method. In addition, the DMLES could enrich teacher preparation and faculty development in evidence-based debriefing methods by identifying specific debriefing behaviors necessary for facilitating reflective thinking among learners.

References

Adamson et al., 2013. K.A. Adamson, S. Kardong-Edgren, J. Willhaus. **An updated review of published simulation evaluation instruments.** Clinical Simulation in Nursing, 9 (9) (2013), pp. e393-e400

Arora et al., 2012. S. Arora, M. Ahmed, J. Paige, D. Nestel, J. Runnacles, L. Hull, N. Sevdalis. **Objective structured assessment of debriefing: Bringing science to the art of debriefing in surgery.** Annals of Surgery, 256 (6) (2012), pp. 982-988

Decker et al., 2013. S. Decker, M. Fey, S. Sideras, S. Caballero, T. Boese, A.E. Franklin, J.C. Borum. **Standards of best practice: Simulation standard VI: The debriefing process.** Clinical Simulation in Nursing, 9 (6) (2013), pp. S26-S29

Dowdy et al., 2013. E. Dowdy, J. Twyford, J.D. Sharkey. **Methods of assessing behavior: Observations and rating scales.** D.H. Saklofske, V.L. Schwean, C.R. Reynolds (Eds.), The Oxford Handbook of Child Psychological Assessment, Oxford University Press, Oxford (2013), 10.1093/oxfordhb/9780199796304.013.0026

Dreifuerst, 2012. K.T. Dreifuerst. **Using debriefing for meaningful learning to foster development of clinical reasoning in simulation.** The Journal of Nursing Education, 51 (6) (2012), p. 326

Fey, 2014. Fey, M. K. (2014). *Debriefing practices in nursing education programs in the United States* (Doctoral dissertation). Retrieved from https://archive.hshsl.umaryland.edu/handle/10713/4051

Gururaja et al., 2008. R.P. Gururaja, T. Yang, J.T. Paige, S.W. Chauvin. **Examining the effectiveness of debriefing at the point of care in simulation-based operating room team training.** K. Henriksen, J.B. Battles, M.A. Keyes, M.L. Grady (Eds.), Advances in patient safety: New directions and alternative approaches, Agency for Healthcare Research and Quality, Rockville, MD (2008)

Hayden et al., 2014. J.K. Hayden, R.A. Smiley, M. Alexander, S. Kardong-Edgren, P.R. Jeffries. **The NCSBN national simulation study: A longitudinal, randomized, controlled study replacing clinical hours with simulation in prelicensure nursing education.** Journal of Nursing Regulation, 5 (2) (2014), pp. S1-S64

IBM Corp, 2013. IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

Mariani et al., 2013. B. Mariani, M.A. Cantrell, C. Meakim, P. Prieto, K.T. Dreifuerst. **Structured debriefing and students' clinical judgment abilities in simulation.** Clinical Simulation in Nursing, 9 (5) (2013), pp. e147-e155

Neill and Wotton, 2011. M.A. Neill, K. Wotton. **High-fidelity simulation debriefing in nursing education: A literature review.** Clinical Simulation in Nursing, 7 (5) (2011), pp. e161-e168

Reed, 2012. S.J. Reed. **Debriefing experience scale: Development of a tool to evaluate the student learning experience in debriefing.** Clinical Simulation in Nursing, 8 (6) (2012), pp. e211-e217, 10.1016/j.ecns.2011.11.002

Shinnick et al., 2011. M.A. Shinnick, M. Woo, T.B. Horwich, R. Steadman. **Debriefing: The most important component in simulation?** Clinical Simulation in Nursing, 7 (3) (2011), pp. e105-e111

Shrout and Fleiss, 1979. P.E. Shrout, J.L. Fleiss. **Intraclass correlations: Uses in assessing rater reliability.** Psychological Bulletin, 2 (1979), pp. 420-428

Simon et al., 2009. R. Simon, D.B. Raemer, J.W. Rudolph. **Debriefing assessment for simulation in healthcare—Rater version.** Center for Medical Simulation, Boston, MA (2009)

Tabachnick and Fidell, 2013. B.G. Tabachnick, L.S. Fidell. **Using multivariate statistics.** (6th ed.), Pearson, Boston, MA (2013)

Waznonis, 2015. A.R. Waznonis. **Simulation debriefing practices in traditional baccalaureate nursing programs: National survey results.** Clinical Simulation in Nursing, 11 (2) (2015), pp. 110-119, 10.1016/j.ecns.2014.10.002