

Development and Validation of a Theory-Informed Group Learning Environment Assessment Tool for Graduate Medical Education Programs

Cristina E. Welch, MD

Melissa M. Carbajal, MD

Shelley Kumar, MS, MSc

Satid Thammasitboon, MD, MHPE

ABSTRACT

Background Recent studies showed that psychological safety is important to resident perception of the work environment, and improved psychological safety improves resident satisfaction survey scores. However, there is no evidence in medical education literature specifically addressing relationships between psychological safety and learning behaviors or its impact on learning outcomes.

Objective We developed and gathered validity evidence for a group learning environment assessment tool using Edmondson's Teaming Theory and Webb's Depth of Knowledge model as a theoretical framework.

Methods In 2018, investigators developed the preliminary tool. The authors administered the resulting survey to neonatology faculty and trainees at Baylor College of Medicine morning report sessions and collected validity evidence (content, response process, and internal structure) to describe the instrument's psychometric properties.

Results Between December 2018 and July 2019, 450 surveys were administered, and 393 completed surveys were collected (87% response rate). Exploratory factor analysis and confirmatory factor analysis testing the 3-factor measurement model of the 15-item tool showed acceptable fit of the hypothesized model with standardized root mean square residual = 0.034, root mean square error approximation = 0.088, and comparative fit index = 0.987. Standardized path coefficients ranged from 0.66 to 0.97. Almost all absolute standardized residual correlations were less than 0.10. Cronbach's alpha scores showed internal consistency of the constructs. There was a high correlation among the constructs.

Conclusions Validity evidence suggests the developed group learning assessment tool is a reliable instrument to assess psychological safety, learning behaviors, and learning outcomes during group learning sessions such as morning report.

Introduction

Psychological safety has recently garnered more attention within medical education, as it has been shown to improve residents' perception of their work environment.¹ It refers to the perception that a learner is free to take interpersonal risks such as reporting mistakes or problems, or sharing new ideas without feeling they will be penalized for highlighting their vulnerability.^{1,2} Early work by Edmondson showed that observed medical teams displaying higher levels of teamwork disclosed more medical errors in order to encourage learning and improvement.³ These observations led to the development of "Teaming Theory" in which psychological safety and team learning behaviors are core constructs.³ Although learners in any environment naturally seek to minimize interpersonal risks, environments like morning report (typically a case-based teaching session for faculty and trainees at academic institutions⁴), which are marked by constant evaluation and

hierarchy, can make disclosing medical errors and knowledge deficits challenging.^{5,6}

Although there is extensive evidence that shows psychological safety improves team performance in the business literature, there is limited evidence in medical education. Studies suggested psychological safety may be relevant to resident perception of clinical experiences,¹ and it correlates with resident satisfaction survey scores.⁷ Additionally, one study reported "toxic" work environments—" (negative) interactions with faculty/attending" and "attending who berate resident physicians"—as the most critical factor influencing medical trainee burnout.⁸ Others have highlighted ensuring the psychological safety of learning environments as a key factor in facilitating learning.⁹

We posit that psychological safety experienced by medical trainees and faculty is an essential, but currently overlooked, component for assessing quality of group learning environments such as morning report. Using Teaming Theory as part of our theoretical framework,^{3,10} we aimed to develop a tool to assess the group learning environment

DOI: <http://dx.doi.org/10.4300/JGME-D-19-00748.1>

experienced during a morning report, including its impact on learning outcomes, and to gather validity evidence to support the interpretation of the results derived from this tool.

Methods

We used a systematic approach to survey creation to develop our group learning environment assessment tool and Messick's framework to guide the accrual of validity evidence to support the interpretation of assessment results.^{11–13} Content and response process evidence was obtained during tool development. Internal structure evidence was obtained through psychometric evaluation.

Tool Development and Theoretical Framework

First, we affirmed the need for assessing psychological safety within the Baylor College of Medicine neonatology section through a comprehensive review of the literature,^{1–3,5,6,9,14–17} informal interviews with fellows and personal communication with faculty and experts on medical education and teamwork within and outside of the institution, and a review of our Accreditation Council for Graduate Medical Education survey results from 2016 to 2018. Through an iterative process by investigators and experts, we developed a theoretical framework (FIGURE 1) based on synthesis of the literature using a “Teaming” theory (team psychological safety and team learning behaviors models³) as well as the addition of Webb's Depth of Knowledge model.¹⁸ Edmondson defines team psychological safety as “a shared belief that the team is safe for interpersonal risk taking” and team learning behaviors as “an ongoing process of reflection and action, characterized by asking questions, seeking feedback, experimenting, reflecting on results, and discussing errors or unexpected outcomes of actions.”³ Together, the team psychological safety and team learning behaviors delineate desirable learning attributes of interest. The depth of knowledge model offers a framework for learning outcomes and requires learners to demonstrate their level of understanding and their ability to transfer knowledge between various contexts (ie, morning report to the bedside).

This theoretical framework guided the tool development. Item creation was based on Edmondson's prior psychological safety survey,³ which resulted in 18 items covering 3 domains: team psychological safety, team learning behaviors, and depth of knowledge. Each item was written as a short phrase asking participants to respond on a 5-point Likert scale. To ascertain content validity, the tool went through an iterative process using expert consensus of faculty in the pediatric sections of neonatology, critical care,

What was known and gap

The literature demonstrates psychological safety is important to resident perception of the work environment and can enhance resident satisfaction survey scores. However, there is no evidence in medical education literature specifically addressing relationships between psychological safety and learning behaviors or its impact on learning outcomes.

What is new

A group learning environment assessment tool using Edmondson's Teaming Theory as a framework.

Limitations

Single center study which limits generalizability. Survey relied on self-reported data and may have also been influenced by social desirability and acquiescence biases.

Bottom line

The tool is a reliable instrument to assess psychological safety, learning behaviors, and learning outcomes during group learning sessions.

hospital medicine, and medical education. In total, we reworded 3 items and added 3 items. Cognitive interviews with end users to assess response process validity were performed. Ten faculty members (neonatology, hospital medicine, critical care) and 10 fellows (neonatology, cardiology and critical care) offered feedback on clarity, understanding, and interpretation of tool items and format. As a result, examples were added to each item in the depth of knowledge domain. The tool was then piloted in December 2018 and revised before administration via paper form.

Setting and Participants

After development, the tool was administered to the neonatology section morning report participants using purposive sampling of morning report on day 5 and 26 of each 4-week service block as well as 2 to 3 additional randomized days during each block. The study period began in December 2018 and concluded in July 2019. Morning report is hosted daily from 8:00 to 8:45 AM where the post-call fellow presents overnight admissions and any active or clinically challenging patients. It serves as an overnight handoff to the day team and offers opportunities to review imaging and discuss diagnostic and clinical challenges. Participants include on-service fellows, faculty, radiologists, and other learners (off-service fellows and faculty, medical students, and neonatal nurse practitioners). At the conclusion of morning report, participants were invited to complete the tool reflecting on their experiences during the session.

Psychometric Evaluation

We conducted exploratory factor analysis (EFA) to examine the factor structure. Confirmatory factor

analysis (CFA) was then performed to assess how well our final model fits the data, followed by measurement of internal consistency (Cronbach's alpha, inter-factor correlations, and item-total correlations) to determine validity evidence for internal structure. Sample size was determined using the items-to-participants criterion due to lack of a priori knowledge of communalities. We aimed for 180 minimum responses for EFA (10 participants per item) and 200 minimum responses for CFA.^{19,20} Kaiser-Meyer-Olkin measure was used to determine sample adequacy.²¹

Exploratory factor analysis determined whether the tool items intended to assess the proposed 3 domains would load on their respective factors. Estimation of factors was done with weighted least squares means and variance adjusted estimator and rotated with an Oblimin (Geomin) rotation that provided the best-defined factor structure. We planned a priori to interpret an item as loading on a factor if the rotated factor pattern loading was ≥ 0.50 for that factor.²²

Confirmatory factor analysis was then conducted to assess how well the model fits another set of data. Multiple fit indices, measuring the degree to which the factor model reproduced the empirical covariance matrix, were computed based on the recommended cutoffs. Cronbach's α coefficient was calculated to determine internal consistency of items within a factor (≥ 0.7 satisfactory).²³ Homogeneity of factors was examined using item-total correlations (≥ 0.40 indicate all item are positively correlated).²⁴ In addition, all corrected item-total correlations were assessed, with scores above 0.2 indicating an acceptable correlation between each item and the overall score. Statistical significance was set at $P \leq .05$.

We performed factor analyses on Mplus version 8.3 (Muthén & Muthén, Los Angeles, CA) and used SPSS 25.0 (IBM Corp, Armonk, NY) for assessing internal consistency and Kaiser-Meyer-Olkin Measure of Sample Adequacy.

The Institutional Review Board of Baylor College of Medicine approved the study.

Results

The tool comprises 3 domains: team psychological safety, team learning behaviors, and depth of knowledge (FIGURE 2). We administered 450 surveys and received 393 completed surveys (87% response rate) from 25 morning report sessions. The Kaiser-Meyer-Olkin of 0.87 (> 0.6) indicated sample adequacy for EFA, and EFA on the first 190 responses suggested 3 factors based on Eigenvalue (≥ 1) and scree plot. Given very high cross-loadings on team psychological safety items 6 and 7, both were deleted to reduce

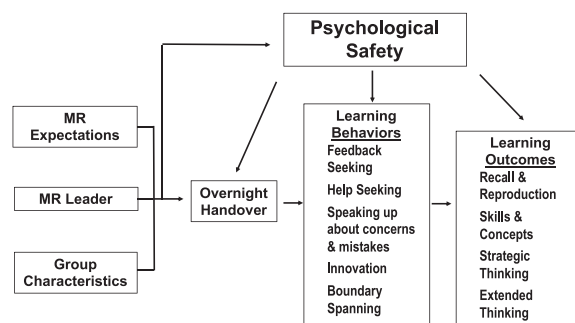


FIGURE 1
Theoretical Framework of Group Learning Environment Assessment Tool

Note: Theoretical framework derived from Edmonson's work³ proposes a link between psychological safety, learning behaviors, and learning outcomes. In this framework, learning is optimized by promoting 5 core learning behaviors among learners: feedback seeking, help seeking, speaking up about concerns and mistakes, innovation, and boundary spanning.³ Given that there is no consensus regarding what learning outcomes are expected for morning report in the literature, we chose to use Webb's depth of knowledge levels¹⁴ as a framework to operationally define differing levels of knowledge gained during a morning report. The levels vary sequentially and in increasing knowledge depth from recall and reproduction (What is the knowledge?); skills and concepts (How can the knowledge be used?); strategic thinking (Why can the knowledge be used?); and extended thinking (How else can the knowledge be used?).¹⁴

redundancy among the items. Due to very high correlation coefficient (0.9) between depth of knowledge items 3 and 4, item 4 was removed as we learned that many respondents had difficulty answering it. The rotated factor loadings are presented in TABLE 1. The high loadings (> 0.5) for all factors indicated a pivotal relationship between the factor and variable.²⁵ All items were clustered with like items consistent with the proposed 3 domains.

Confirmatory factor analysis on the remaining 203 observations showed acceptable to excellent fit indices of the 3-factor model: standardized root mean square residual of 0.034 (< 0.055 = ideal²⁵), root mean square error of approximation of 0.088 (0.08–1.0 = mediocre), and comparative fit index of 0.987 (> 0.96 = excellent²⁵). These findings indicate that the overall structure of the model fits the data. Structural equation modeling provided standardized path coefficients and significance levels for parameter estimates of the associations as shown in FIGURES 3a and 3b. The results indicate significant effect of team psychological safety on team learning behaviors ($\beta = 0.75$, $P < .0001$) and significant effect of team learning behaviors on depth of knowledge ($\beta = 0.66$, $P < .0001$). Team psychological safety has no significant effect on depth of knowledge ($\beta = 0.143$, $P = .09$), so that path was excluded from the final model. All path coefficients were high (0.66 to 0.97) with favorable t values (> 2.58) supporting the 3-factor model.²⁵ Almost all standardized residual

Please rate whether the following statements are accurate or inaccurate descriptions of the social climate of today's morning report.	Strongly Disagree	←————→				Strongly Agree
When someone makes a mistake, it is often held against him or her.	1	2	3	4	5	
It is easy to discuss difficult issues and problems.	1	2	3	4	5	
Participants are sometimes rejected for having different opinions.	1	2	3	4	5	
It is completely safe to take a risk in discussion.	1	2	3	4	5	
It is difficult to ask other participants for help.	1	2	3	4	5	
No participant deliberately undermines another participant.	1	2	3	4	5	
Participants value and respect each other's contributions.	1	2	3	4	5	
How often did you observe the following behaviors demonstrated during today's morning report?	Rarely	←————→				Very Frequently
Participants seek feedback from one another through discussion.	1	2	3	4	5	
Differences of opinion are handled as teaching opportunities.	1	2	3	4	5	
Problems/errors are communicated appropriately so that ongoing issues are addressed.	1	2	3	4	5	
Participants seek new information that leads to important changes in our plans/work processes.	1	2	3	4	5	
Participants talk about mistakes or misconceptions and their potential solutions.	1	2	3	4	5	
Participants raise concerns they have about plans or decisions.	1	2	3	4	5	
Multidisciplinary views are presented and discussed.	1	2	3	4	5	
At the end of today's morning report, how confident are you in completing the following tasks related to the cases presented?	Not at all Confident	←————→				Very Confident
State medical facts, principles or rules pertinent to the clinical scenario(s) presented. <i>Example: time constant = compliance x resistance</i>	1	2	3	4	5	
Explain how concepts and skills can be used to solve the clinical scenario(s) presented. <i>Example: Babies with RDS have a short time constant, this is because they have decreased lung compliance.</i>	1	2	3	4	5	
Draw conclusions from observations/reasoning/evidence discussed, and apply them to variations of the same clinical scenario(s). <i>Example: Last time I treated RDS, I used a fast rate with a short iT because they have decreased lung compliance. This new baby has RDS, let's use a fast rate and short iT.</i>	1	2	3	4	5	
Formulate generalized knowledge and strategies, and apply them to solve different clinical scenario(s). <i>Example: Fast rates and short iT work for RDS because they have a short time constant and decreased lung compliance. If infants with BPD have longer time constants and increased resistance, then maybe a slower rate with a longer iT will be a good ventilation strategy to try.</i>	1	2	3	4	5	

FIGURE 2

Group Learning Environment Assessment Tool for Morning Reports

Note: The team psychological safety (7-item, 5-point agreement scale) assesses the extent to which individuals view the learning environment as conducive to interpersonal risk; the team learning behavior (7-item, 5-point frequency scale) assesses the extent to which individuals engage in desirable learning behaviors according to team psychological safety; and the depth of knowledge (4-item, 5-point confident scale) assesses the products of learning.

TABLE 1

Rotated Factor Loadings From Exploratory Factor Analysis for Final 15 Items in the Tool^a

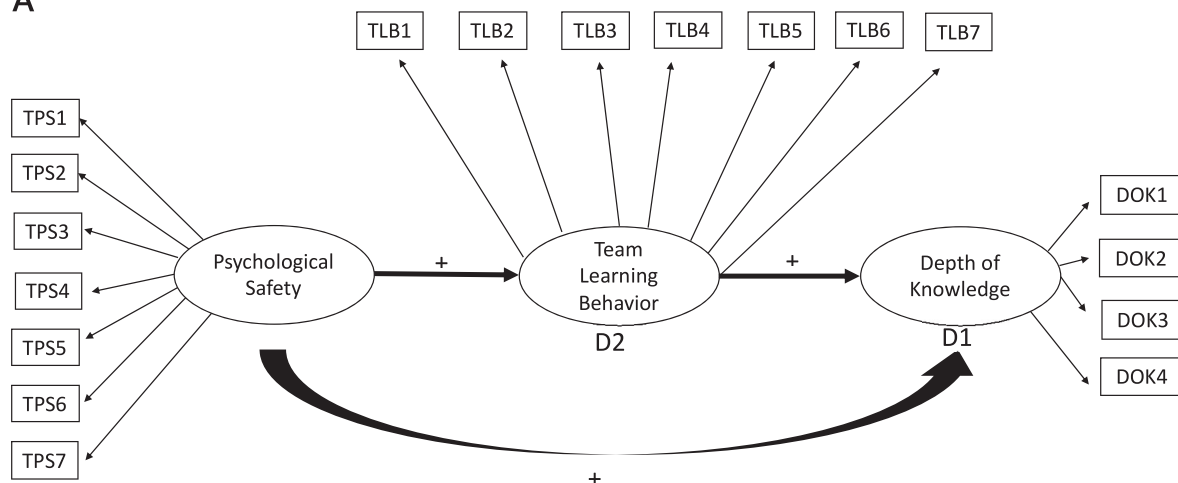
Abbreviated Tool Item	Factor 1: Team Psychological Safety	Factor 2: Team Learning Behavior	Factor 3: Depth of Knowledge
Mistake held against individual ^b	0.758^c	-0.026	0.157
Easy to discuss issues and problems	0.540	0.361	0.038
Rejected for having different opinions ^b	0.935	0.020	0.034
Completely safe to take risk in discussion	0.601	0.453	-0.022
Difficult to ask other participants for help ^b	0.655	0.178	-0.029
Participants seek feedback from one another	0.025	0.720	0.033
Difference of opinion held as teaching opportunity	0.177	0.769	0.016
Problems/errors are communicated appropriately	0.037	0.872	-0.008
Participants seek new information	0.038	0.792	0.016
Participants talk about mistakes or misconceptions	-0.165	0.858	0.010
Participants raise concern about plans or decisions	-0.103	0.978	-0.132
Multidisciplinary views are presented	0.011	0.893	0.008
State medical facts, principles, or rules	-0.010	0.071	0.920
Explain how concepts and skills can be used	-0.010	-0.023	1.004
Draw conclusions from observations/reasoning/evidence discussed	-0.005	0.073	0.921

^a Based on the responses of 190 Baylor College of Medicine neonatology faculty, trainees, and medical students.

^b Items 1, 3, and 5 were reverse coded.

^c Bold fonts indicate the items in each factor that had factor coefficients greater than 0.50.

A



B

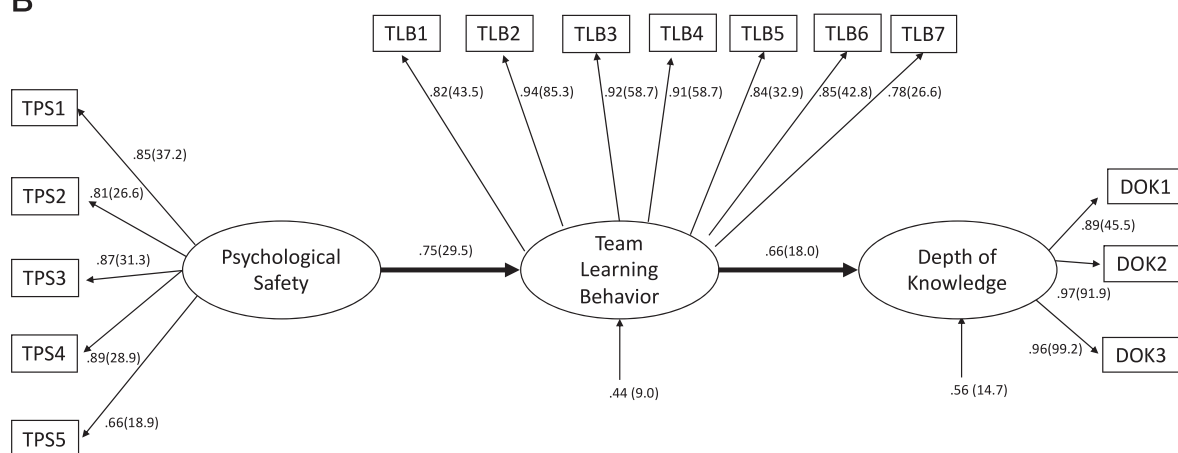


FIGURE 3A

The Hypothesized Model of Group Learning Environment Assessment Tool

Note: Hypothesized model for the tool. The 3 domains—team psychological safety, team learning behaviors, and depth of knowledge—assess the extent to which individuals view the learning environment as conducive to interpersonal risk, the extent to which individuals engage in desirable learning behaviors, and the products of learning respectively.

FIGURE 3B

The Best Fit Model of Group Learning Environment Assessment Tool

Note: Final fitted model had 5 manifest variables for team psychological safety, 7 for team learning behavior, and 3 for depth of knowledge. Parameters expressed as maximum likelihood estimates (standardized path coefficients). Parenthetical numbers indicate associated t values for standardized path coefficients (All t values are significant at the .01 level because their absolute value exceeds 2.58). Standardized path coefficient can be interpreted as following. For example, the standardized path coefficient for the effect on team learning behavior on depth of knowledge is 0.66 ($t = 18$, $P < .01$). This means there is an increase of 0.66 standard deviation in depth of knowledge for an increase of one standard deviation in team learning behavior, while holding constant the effect of the other independent variable.

correlations were small (< 0.10),²⁶ indicating a good fit of the model.

Factor descriptives, Cronbach's alpha, and inter-factor correlations are given in TABLE 2. Cronbach's alpha values for all 3 factors indicated good to excellent internal consistency reliability (≥ 0.8). The correlations among the 3 factors was high, ranging from 0.66 to 0.70. For each of the 15 retained items, item-total score correlations were positive and significant ($P < .001$), affirming significant contribution of each item to the total score of the tool.

Discussion

We developed a group learning environment assessment tool using “Teaming” theory as a theoretical framework. The tool content was guided by this framework based on 3 existing domains: team psychological safety,³ team learning behaviors,³ and depth of knowledge.¹⁸ Through the development process and psychometric evaluation, we gathered validity evidence pertaining to content, response process, and internal structure.¹³ To our knowledge,

TABLE 2

Factor Descriptives and Internal Consistency for Final 15 Items of the Tool^a

Factor (No. of Items)	Mean (Range)	Cronbach's Alpha	Inter-Factor Correlations ^b 1 2 3	Range of Corrected Item-Total Correlations ^c
Team psychological safety (5)	4.05 (1–5)	0.87	1	0.62–0.77
Team learning behavior (7)	4.08 (1–5)	0.92	0.70 1	0.65–0.85
Depth of knowledge (3)	3.97 (1–5)	0.93	0.66 0.70 1	0.80–0.90

^a Based on the responses of 203 Baylor College of Medicine neonatology faculty, trainees, and medical students.^b Inter-factor correlations are reported along the major diagonal.^c None of the corrected item-total correlations were less than 0.50, indicating that all items correlate very well with the subscales.

this is the first study offering empirical evidence in medical education addressing psychological safety and learning behaviors and outcomes in group learning sessions. Although the setting of this study is a morning report session, the systematic approach we described can be used to guide an adaptation of the instrument for other forms of group learning sessions (case conferences, morbidity and mortality conferences, bedside teaching, etc) in medical training that require an educational dialogue²⁷ to facilitate optimal learning. We propose this tool can be administered to faculty and trainees to assess the quality of group learning conferences and inform program improvement efforts. We suggest administering the tool 2 to 4 times per year or as needed to assess psychological safety within these sessions.

Psychological safety is a vital component to group learning and team behaviors.¹⁴ It allows learners to take interpersonal risks, exhibit curiosity, ask questions, and show their vulnerabilities of knowledge deficits.⁶ In a recent qualitative study in medical education, safe learning environments not only built trust and set clear expectations for learners, but also encouraged critical thinking.²⁸ Psychologically safe environments can have profound effects on learning, growth potential, and physician burnout, and may help distinguish Socratic teaching from “pimping.”^{8,9,17,28}

In addition to bringing attention to psychological safety domains for assessing education process within medical training, this is the first study to evaluate psychological safety at an institutional level and examine the relationships between psychological safety, learning behaviors, and learning outcomes (which have been shown to be critical relationships in organizational work teams).^{2,3,15,16} Although we did not find an association between team psychological safety and depth of knowledge, this may be due to rating errors. Upon reviewing the raw data, there appears to be acquiescence bias in the depth of knowledge domain (ie, giving the same rating to each depth of knowledge item). A relative ranking scale for this domain may have provided more accurate data

than the Likert scale we used as this would have encouraged participants to identify how much of each level they perceived during morning report.

This study has limitations. Self-reporting of the team learning behaviors and depth of knowledge poses threats to validity. Social desirability bias, acquiescence bias, and survey fatigue could have influenced the results. This is a single center study, which limits generalizability. Future studies should investigate the impacts of psychological safety on observed learning behaviors and performance measures. Finally, we did not include validity evidence pertaining to relations to other variables and consequences.

Conclusions

In this study, we developed a theory informed group learning environment assessment tool. We demonstrated acceptable validity evidence to support the use of the tool as a means of assessing psychological safety and its consequences—learning behaviors and learning outcomes—during group learning sessions such as morning report.

References

1. Toralba KD, Loo LK, Byrne JM, Baz S, Cannon GW, Keitz SA, et al. Does psychological safety impact the clinical learning environment for resident physicians? Results from the VA's Learner's Perceptions Survey. *J Grad Med Educ*. 2016;8(5):699–707. doi:10.4300/JGME-D-15-00719.1.
2. Edmondson AC. Managing the risk of learning: psychological safety in work teams. https://www.hbs.edu/faculty/Publication%20Files/02-062_0b5726a8-443d-4629-9e75-736679b870fc.pdf. Accessed July 2, 2020.
3. Edmondson AC. Psychological safety and learning behavior in work teams. *ASQ*. 1999;44(2):350–383.
4. Parrino TA, AG Villanueva. The principles and practice of morning report. *JAMA*. 1986;256(6):730–733.
5. Appelbaum NP, Dow A, Mazmanian PE, Jundt DK, Appelbaum EN. The effects of power, leadership and

- psychological safety on resident event reporting. *Med Educ.* 2016;50(3):343–350. doi:10.1111/medu.12947.
6. Edmondson AC. *The Fearless Organization: Creating Psychological Safety in the Workplace for Learning, Innovation, and Growth*. Hoboken, NJ: John Wiley & Sons, Inc; 2019.
 7. Appelbaum NP, Santen SA, Aboff BM, Vega R, Munoz JL, Hemphill RR. Psychological safety and support: assessing resident perceptions of the clinical learning environment. *J Grad Med Educ.* 2018;10(6):651–656. doi:10.4300/JGME-D-18-00286.1.
 8. Ismail M, Johnson SL, Weaver SJ, Wu AW, Gielen AC. Factors influencing burn-out among resident physicians and the solutions they recommend. *Postgrad Med J.* 2018;94(1115):540–542. doi:10.1136/postgradmedj-2018-135777.
 9. Stoddard HA, O'Dell DV. Would Socrates have actually used the “Socratic method” for clinical teaching? *J Gen Intern Med.* 2016;31(9):1092–1096. doi:10.1007/s11606-016-3722-2.
 10. Varpio L, Paradis E, Uijtdehaage S, Young M. The distinctions between theory, theoretical framework and conceptual framework [published online ahead of print November 12, 2019]. *Acad Med.* doi:10.1097/ACM.0000000000003075.
 11. Artino AR Jr, La Rochelle JS, Dezee KJ, Gehlback H. Developing questionnaires for educational research: AMEE Guide No. 87. *Med Teach.* 2014;36(6):463–474. doi:10.3109/0142159X.2014.889814.
 12. Rickards G, Magee C, Artino AR Jr. You can't fix by analysis what you've spoiled by design: developing survey instruments and collecting validity evidence. *J Grad Med Educ.* 2012;4(4):407–410. doi:10.4300/JGME-D-12-00239.1.
 13. Downing SM. Validity: on meaningful interpretation of assessment data. *Med Educ.* 2003;37(9):830–837. doi:10.1046/j.1365-2923.2003.01594.x.
 14. Edmondson A, Lei Z. Psychological safety: the history, renaissance, and future of an interpersonal construct. *Ann Re Orgl Psychol Organl Behav.* 2014;1:23–43.
 15. Edmondson A. Psychological safety, trust, and learning in organizations: a group-level lens. In: Kramer R, Cook K, eds. *Trust and Distrust in Organizations: Dilemmas and Approaches*. New York, NY: Russell Sage Foundation; 2004:239–272.
 16. Edmondson AC. The competitive imperative of learning. *Harvard Business Review.* <https://hbr.org/2008/07/the-competitive-imperative-of-learning>. Accessed July 2, 2020.
 17. Bynum WE, Haque TM. Risky business: psychological safety and the risks of learning medicine. *J Grad Med Educ.* 2016;8(5):780–782. doi:10.4300/JGME-D-16-00549.1.
 18. Webb NL. Depth-of-knowledge levels for four content areas. <https://www.acpsd.net/cms/lib/SC02209457/Centricity/Domain/74/DOK%20Four%20Content%20Areas%20Webb.pdf>. Accessed July 2, 2020.
 19. Floyd FJ, Widman KF. Factor analysis in the development and refinement of clinical assessment instruments. *Psychol Assess.* 1995;7(3):286–299. doi:10.1037/1040-3590.7.3.286.
 20. Asparouhov T, Muthén BO. SRMR in Mplus. <https://www.semanticscholar.org/paper/SRMR-in-Mplus-Asparouhov-Muth%C3%A9n/7d1ade06295314ea4fce371d62577c9057096280>. Accessed July 2, 2020.
 21. Kaiser HF. An index of factor simplicity. *Psychometrika.* 1974;39(1):31–36.
 22. Tabachnick BG, Fidell LS. *Using Multivariate Statistics*. 6th ed. Boston, MA: Pearson; 2013.
 23. Nunnally J. *Psychometric Theory*. New York, NY: McGraw-Hill; 1978.
 24. Lounsbury JW, Gibson LW, Saudardas RA. Scale development. In: Leong FTL, Austin JT, eds. *The Psychology Research Handbook: A Guide for Graduate Students and Research Assistants*. Thousand Oaks, CA: Sage; 1996.
 25. O'Rourke N, Hatcher L. *A Step-by-Step Approach to Using SAS for Factor Analysis and Structural Equation Modeling*. 2nd ed. Cary, NC: SAS Institute Inc; 2013.
 26. Kline RB. *Principles and Practice of Structural Equation Modeling*. 4th ed. New York, NY: Guilford Press; 2016.
 27. Walter T, Eppich W, Cheng A, Miller S, Teunissen PW, Watling CJ, et al. Learning conversations: an analysis of their theoretical roots and their manifestations of feedback and debriefing in medical education [published online ahead of print July 30, 2019]. *Acad Med.* doi:10.1097/ACM.0000000000002932.
 28. Jaffe LE, Lindell D, Sullivan AM, Huang GC. Clear skies ahead: optimizing the learning environment for critical thinking from a qualitative analysis of interviews with expert teachers. *Perspect Med Educ.* 2019;8(5):289–297. doi:10.1007/s40037-019-00536-5.



Cristina E. Welch, MD, is a Perinatal-Neonatal Fellow, Baylor College of Medicine, Texas Children's Hospital; **Melissa M. Carbajal, MD**, is Assistant Professor, Neonatal-Perinatal Medicine Fellowship Program Director, Department of Pediatrics, Baylor College of Medicine, Texas Children's Hospital; **Shelley Kumar, MS, MSc**, is Instructor, Center for Research, Innovation and Scholarship in Medical Education, Department of Pediatrics, Texas Children's Hospital; and **Satid Thammasitboon, MD, MHPE**, is Deputy Director, Best Evidence Medical Education International Collaborating Center, Associate Professor, Chair, Resident Scholarship Program, Department of Pediatrics, and Director, Center for Research, Innovation and Scholarship, Baylor College of Medicine, Texas Children's Hospital.

ORIGINAL RESEARCH

Funding: This project was funded by the Evangelina “Evie” Whitlock Fund, Texas Children’s Hospital Educational Grant.

Conflict of interest: The authors declare they have no competing interests.

This work was previously presented at the Pediatric Academic Societies Annual Meeting, Baltimore, MD, April 24–May 1, 2019.

The authors would like to thank Drs. K. Suresh Gautham, B. Gandhi, J. Garcia-Prats, S. Parmekar, and G. Singhal (Baylor College of Medicine) for their valuable and constructive

suggestions during the planning and development of this research work. Their willingness to give their time so generously has been very much appreciated.

Corresponding author: Cristina E. Welch, MD, Texas Children’s West Tower, 6621 Fannin Street, Houston, TX 77030, 859.913.0385, cristi.welch@gmail.com

Received October 24, 2019; revisions received March 1, 2020, and May 14, 2020; accepted May 29, 2020.