



Short Communication

Analyzing instructor facilitation patterns in debriefings using AI-based speech metrics

Soledad Armijo-Rivera, (Mg) MD, MEd^{a,b,*}, Scarlett Vicencio-Clarke, (Mg) MD, MEd^c,
Hernán Caamaño^d, Pía Díaz^e, Carla Pino^e, Francesca Caldo^d, Daniel Herrera^f,
Carlos Hinrichsen, (Mg)^g

^a Facultad de Medicina, Universidad del Desarrollo, Las Condes, Santiago, Chile

^b Universidad de Barcelona, Barcelona, España

^c Unidad de Simulación e Innovación en Salud, Universidad San Sebastián, Providencia, Santiago, Chile

^d Escuela de Medicina, Universidad San Sebastián, Providencia, Santiago, Chile

^e Escuela de Obstetricia, Universidad San Sebastián, Providencia, Santiago, Chile

^f I+D+i, Link-USS, Unidad de Simulación e Innovación en Salud, Universidad San Sebastián, Recoleta, Santiago, Chile.

^g Escuela de Diseño Digital e Industrias Creativas, Facultad de Arquitectura, Arte y Diseño, Universidad San Sebastián, Recoleta, Santiago, Chile

ARTICLE INFO

Keywords:

Artificial intelligence
Debriefing facilitation
GailBot API
Novice instructors
Reflective learning
Simulation-based education

ABSTRACT

Background: Facilitation of debriefings is critical for fostering reflection and behavioral change. However, little is known about how novice instructors distribute their speaking time and how this impacts perceived debriefing quality. Artificial intelligence (AI)-driven conversational analysis offers an objective method to characterize facilitation patterns.

Methods: Participants were fifteen novice healthcare instructors with less than one year of experience who recently completed an online simulation course. A cross-sectional study was conducted using GailBot API for automated transcription and speech analysis of remote interprofessional debriefings. Standardized scenarios and consistent student participants controlled external variables. Instructor facilitation patterns were compared to DASH evaluations from students, instructors, and an expert observer.

Results: Instructors spoke for an average of 56.09% of debriefing time. Higher instructor talk time correlated negatively with DASH self-ratings (Spearman's Rho = -0.627 , $p = .012$).

Conclusion: AI-based analysis provides valuable insights into facilitation patterns, supporting multidisciplinary faculty development.

© 2025 International Nursing Association for Clinical Simulation and Learning. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Background

Debriefing is a fundamental practice in simulation-based education, guided by standards of practice, some of which explicitly invite fostering critical reflection and behavioral change in participants (INACSL Standards Committee et al., 2025). Reflectivity has been identified as the primary determinant of such changes in debriefing participants, with multiple instructor-dependent factors playing a role in this experience (Kainth et al., 2024). In aviation, debriefing practice was described using three levels of facilitation: high facilitation levels imply minimal instructor involvement, intermediate facilitation levels require guided discussion with more participation of the facilitator, and low facilitation level occurs

with directive guidance and active participation of the facilitator (McDonnell et al., 1997).

While early debriefing advocates in healthcare recommended high levels of facilitation with minimal instructor intervention and three levels of facilitation, there is no clear definition of specific time considered to limit these levels (Dismukes et al., 2006). Additionally, substantial differences persist in debriefing practices, underscoring the need for a more precise evaluation of how debriefings are implemented across various geographic contexts (Lomuscio et al., 2025). Only 21% of studies have documented how time is allocated in debriefings (Duff et al., 2024), and there is insufficient information on how novice instructors distribute their interventions and whether these affect the perceived quality of debriefing.

Simulation instructor debriefing training has evolved over the past decades, with programs ranging from basic to advanced

* Corresponding author.

E-mail address: s.armijo@udd.cl (S. Armijo-Rivera).

courses, although many programs do not deeply address the progression of facilitation levels (Cheng et al., 2015). The lack of standardization in training programs limits instructors' abilities to adopt effective strategies across clinical simulation settings (Ing et al., 2022). In this context, implementing Artificial Intelligence (AI) tools for conversational analysis in debriefing (Brutschi et al., 2024) is an emerging strategy that could provide objective data on facilitation. The GailBot API is an AI-based customizable system designed to transcribe and analyze verbal interactions (Umair et al., 2022). This system could be useful in designing local solutions for debriefing analysis that allow the determination of the intervention time of participants and the identification of facilitation patterns. This study aimed to bridge the knowledge gap regarding facilitation strategies used by novice simulation instructors through AI analysis of debriefing interventions using a customized Gailbot API. The findings could improve instructor training and optimize the quality of learning in clinical simulation.

Methods

A cross-sectional, relational analytic study was conducted to explore the levels of facilitation achieved by novice instructors during debriefing sessions in the context of remote interprofessional simulations at a Chilean university. The sample consisted of 15 voluntary novice simulation instructors from four cities in Chile. A novice instructor was defined as a healthcare professional with less than one year of experience as a high-fidelity simulation instructor and no prior formal training in simulation. They completed an online course in clinical simulation-based education in January 2024, which included online synchronous high-fidelity simulations. The instructors were predominantly female (86.67%). In terms of age distribution, most participants were between 30 and 34 years (33.33%), followed by those aged 35–39 and 40–44 years (each group representing 26.67%), and a smaller proportion between 45 and 49 years (13.33%). The professional backgrounds of the instructors were diverse, with the largest group being physical therapists (26.67%), followed by nurses (20.00%), midwives and occupational therapists (each 13.33%), dietitians (13.33%), and smaller representations from speech therapist and psychologists (each 6.67%). This reflects a multidisciplinary group with varied clinical expertise and professional experiences.

Automated conversational analysis using the GailBot API was used to assess interactions during debriefing. GailBot was selected for its open-source availability, adaptability to research needs, and capacity to process large volumes of audio data efficiently. The setup process of the Gailbot tool included the following phases, functions and milestones.

1. *Data preparation:* Audio files were stored in a designated input folder in Google Drive. An output folder was specified to store the results of the analysis.
2. *Gailbot initialization:* The necessary libraries (Gailbot, whisperx, python-dateutil, pydantic-settings) were installed using PIP. An IBM Watson speech-to-text (STT) engine was configured with API credentials and configuration parameters.
3. *Transcription:* Audio files were added as sources to Gailbot, and a profile for transcription and analysis was applied using XMLPlugin, ChatPlugin, TextPlugin, CSVPlugin, SyllableRatePlugin, etc. The "transcribe()" method was called to generate transcripts. A numeric label was automatically assigned to each detected speaker during audio processing. However, the correspondence between each tag and its specific speaker was adjusted manually.
4. *Speak time differences function:* A function (calculate_time_differences) was defined to extract the duration

of speakers' turns from the generated conversation.csv files. It calculated each speaker's total duration and number of turns in debriefings. The results were summarised in a durations_summary.csv file.

5. *Speak percentage function:* The function analyzes_durations_summary calculates each speaker's speaking time percentage and stores it in analyzed_durations_summary.csv. This function was further refined to calculate the duration and count percentages and store the final results in durations_analysis.csv.

The first iteration of the analysis of a debriefing was contrasted with a formal qualitative analysis performed by an experienced researcher, which yielded 100% agreement on the units of instructor intervention, instructor intervention time, and the identification of pauses longer than three seconds. To ensure the validity of the automated analysis using Gailbot, transcriptions were compared to audios by a researcher, and two main variables were controlled: the scenario and the students' performance.

The scenario corresponded to a 5-minute standardized video of an obstetric and neonatal care scenario involving two students: one medical student, responsible for the delivery care role, and one nurse-midwife student, responsible for the initial neonatal care. The scenario video was recorded with the same students who participated in debriefings. They were final-semester undergraduate students with two years of experience in high-fidelity simulation and debriefing. They demonstrated motivation, strong communication skills, and were trained in the topics to be analyzed. The online training for standardized students includes instruction that allows them to reflect on their performance if given the opportunity. As members of the research team, the standardized students played a specific role in debriefings and were therefore exempt from signing an informed consent form.

Each instructor participated in a 20-minute debriefing session that did not include the simulation scenario itself. They conducted debriefings freely, without a rigid script. The debriefing sessions were audio-recorded and transcribed verbatim by a researcher. After that, we used Gailbot to generate a report of time spent and interventions performed by the instructors and participants. To categorize instructor participation patterns, we divided the total time dedicated to the task into four equal parts based on time proportions. This approach allowed us to group instructors according to the proportion of time they spent speaking during the debriefing, facilitating comparative analysis of their level of intervention and its relationship to perceived debriefing quality. At the end of the debriefing, the students, the instructor, and a certified DASH researcher applied the short version of the DASH instrument in its Spanish version and applied for remote simulations (Olvera-Cortés et al., 2022) and a socio-demographic and professional survey through the SurveyMonkey platform.

The analysis plan included descriptive statistics (frequency, percentage, mean, median, minimum, and maximum) and relational analyses using Pearson's r or Spearman's Rho . Data normality was assessed with the Shapiro-Wilk test. Statistical analyses were performed using JASP 0.19.1, considering a significance level of $p < .05$.

This study was approved by the institutional ethics committee (Act 169-23). An analysis of the benefits and potential risks to participants was conducted and included in the application for review submitted to the committee. Instructors participated voluntarily after signing an informed consent form, conducted previously to data collection. Data collected were anonymized before analysis, and blinded observers were included.

Table 1
Instructors Speaking Time and Interventions During Interprofessional Remote Debriefings.

ID	Instructor Time (second)	Total Time (second)	% Instructor Time	N° of Instructor Interventions	Total No. of Interventions	% Instructor Interventions	DASH Mean Score
1	866.82	1363.14	63.59	61	121	50.41	4.83
2	959.85	1215.88	78.94	27	56	48.21	5.33
3	613.75	1000.01	61.37	39	78	50.00	5.00
4	706.69	1348.98	52.39	34	125	27.20	6.33
5	473.15	683.39	69.24	27	55	49.09	5.33
6	456.34	940.59	48.52	82	196	41.84	5.17
7	394.43	785.12	50.24	42	84	50.00	3.67
8	327.90	638.84	51.33	71	141	50.35	5.67
9	326.62	798.72	40.89	45	90	50.00	5.17
10	394.31	918.93	42.91	35	69	50.72	4.00
11	546.58	1084.84	50.38	29	58	50.00	6.00
12	594.93	855.42	69.55	93	185	50.27	5.50
13	868.52	1295.71	67.03	54	107	50.47	4.67
14	207.47	564.38	36.76	24	59	40.68	6.00
15	417.84	717.29	58.25	70	139	50.36	4.83

n = 15; sec = seconds; % = Percentage.

Table 2
Correlations Table

			Pearson		Spearman		Assumption checks	
							Shapiro-Wilk Test for Bivariate Normality	
			r	p	rho	p	Shapiro-Wilk	p
ResDASH	-	InstDASH	0.470	0.077	0.437	0.103	0.922	0.172
ResDASH	-	Std2DASH	0.442	0.099	0.343	0.211	0.942	0.431
ResDASH	-	Std1DASH	-0.268	0.334	-0.256	0.356	0.960	0.807
ResDASH	-	% Educator Time	-0.180	0.520	-0.167	0.552	0.961	0.813
ResDASH	-	% Educator Interventions	-0.364	0.182	-0.166	0.555	0.778	< .001**
InstDASH	-	Std2DASH	0.066	0.816	-0.130	0.644	0.949	0.568
InstDASH	-	Std1DASH	0.004	0.989	0.010	0.972	0.972	0.963
InstDASH	-	% Educator Time	-0.022	0.937	0.002	0.995	0.967	0.919
InstDASH	-	% Educator Interventions	-0.557	0.031*	-0.627	0.012*	0.768	< .001**
Std2DASH	-	Std1DASH	0.073	0.796	0.087	0.757	0.958	0.762
Std2DASH	-	% Educator Time	0.188	0.503	0.167	0.553	0.951	0.610
Std2DASH	-	% Educator Interventions	-0.142	0.614	0.177	0.528	0.752	< .001**
Std1DASH	-	% Educator Time	0.313	0.256	0.329	0.231	0.972	0.968
Std1DASH	-	% Educator Interventions	0.167	0.551	0.263	0.344	0.776	< .001**
% Educator Time	-	% Educator Interventions	0.256	0.357	0.146	0.604	0.815	< .001**

ResDASH = Researcher Mean Value DASH; InstDASH = Instructor Mean Value DASH; Std1DASH = Student 1 Mean DASH; Std2DASH = Student 2 Mean DASH

* p<0.05

** p<0.001

Results

In the 15 debriefings analyzed, instructors occupied 56.09% of the total debriefing time, while students intervened an average of 43.91% of the time. The individual values showed significant variability in the proportion of instructor intervention time, ranging from 36.76% to 78.94% of the total debriefing time. In only four debriefings, the students spoke for more than 50% of the time (51.48%, 57.09%, 59.11% and 63.24%, respectively), while in most cases, the instructor maintained higher participation than the students. Additionally, the number of interventions made by facilitators averaged 49.3% of the total number of interventions per debriefing, with values ranging from 27.2% to 50.72% (Table 1).

We found that six instructors (26.7%) were in the second participation pattern (26%-50% of the time), ten instructors (66.7%) were in the third (51%-70%), and one (6.7%) in the fourth (>70%) participation pattern.

The average DASH scale evaluation given by the students was 5.71 (range 4.00-6.83), the instructors rated themselves at 5.17 (range 3.67-6.33), and the principal investigator gave the highest average score at 5.74 (range 4.50-6.83). A strong and statistically significant correlation was found between the instructor's average DASH self-assessment and the percentage of interventions the participants made during the debriefing session (Table 2). Specif-

ically, the lower the percentage of interventions, the higher the perceived quality of the debriefing (Spearman's Rho = -0.627, p-value = .012).

Based on instructor speaking time and DASH self-assessments, we identified three patterns that could align with Dismukes' facilitation levels: low facilitation corresponds to >70%, intermediate facilitation corresponds to 51%-70%, and high facilitation corresponds to 26%-50% of total debriefing time.

Conclusion

This study contributes novel insights into instructor-participant interaction during debriefing by applying AI-driven conversational analysis as an innovative methodology for evaluating debriefing facilitation within a multidisciplinary simulation context. Unlike traditional scale-based (Olvera et al., 2022) or qualitative debriefing evaluations (Kihlgren et al., 2015), this approach provides an objective and replicable method for measuring facilitation time and participation patterns. It contributes to the existing literature by proposing a new analytical lens for understanding how novice instructors structure debriefings and manage their interactions with learners, complementary to technology-supported conversational analysis (Brutschi et al., 2024). Additionally, this study demonstrates that a free, adaptable, customized, and scalable tool, such

as GailBot API (Umair et al., 2022) can be easily used to transcribe and analyze speech and subsequently characterize facilitation patterns during debriefing.

A key strength of this study is its ability to isolate instructor facilitation behaviors by controlling for participant variability, a limitation that has been present in previous research. Novice instructors participating in this study showed a high average use of time during debriefing and tend to be in the third and second patterns of time used, dominating the conversation and limiting students' opportunities for participation. This finding is consistent with previous studies that highlight novice instructors' difficulty in balancing their debriefing role (Kihlgren et al., 2015). The findings suggest that instructor training should focus on strategies that promote greater student autonomy and reduce instructor directiveness.

The inverse correlation between instructors' self-assessed DASH scores and their frequency of interventions suggests a lack of confidence in promoting learner-centered dialogue or a potential disconnect between self-perceived competence and actual facilitation behaviors. These findings highlight the need for structured training with data-driven feedback and supervised practice. Tools like the GailBot API may also support real-time monitoring and refinement of facilitation behaviors, enhancing instructor development.

This study proposes using time proportions as a practical method to classify facilitation levels. Such an approach could be broad and easily implemented, without excessive costs or the need for experienced facilitators' time, making it both scalable and equitable. However, this classification remains preliminary and requires further validation and testing in broader contexts.

Among the study's limitations is the small, homogeneous sample of novice instructors, which may affect the generalizability of the findings. Additionally, while the GailBot API provided valuable objective data on facilitation levels, future studies could explore the evolution of facilitation levels in instructors over time and the influence of contextual factors, such as the complexity of the scenario and the learner's prior training. Combining automated conversational analysis with qualitative methodologies could provide a more complete picture of the most effective facilitation strategies.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used Grammarly to improve the readability of the manuscript, because the authors are not native English speakers. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Funding

This research receives an Academic Innovation Grant from the Universidad San Sebastián (041-2023).

Conflicts of interest

None.

CRedit authorship contribution statement

Soledad Armijo-Rivera: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Scarlett Vicencio-Clarke:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Hernán Caamaño:** Writing – original draft, Project administration, Investigation, Data curation. **Pía Díaz:** Writing – original draft, Resources, Data curation. **Carla Pino:** Writing – original draft, Investigation, Data curation. **Francesca Caldo:** Writing – original draft, Investigation, Data curation. **Daniel Herrera:** Writing – original draft, Software, Methodology. **Carlos Hinrichsen:** Writing – review & editing, Methodology.

References

- Brutschi, R., Wang, R., Kolbe, M., Weiss, K., Lohmeyer, Q., & Meboldt, M. (2024). Speech recognition technology for assessing team debriefing communication and interaction patterns: An algorithmic toolkit for healthcare simulation education. *Advances in Simulation* (London, England), 9(1), 42. <https://doi.org/10.1186/s41077-024-00315-1>.
- Cheng, A., Grant, V., Dieckmann, P., Arora, S., Robinson, T., & Eppich, W. (2015). Faculty development for simulation programs: Five issues for the future of debriefing training. *Simulation in Healthcare: Journal of the Society for Simulation in Healthcare*, 10(4), 217–222. <https://doi.org/10.1097/SIH.0000000000000090>.
- Dismukes, R. K., Gaba, D. M., & Howard, S. K. (2006). So many roads: Facilitated debriefing in healthcare. *Simulation in Healthcare: Journal of the Society for Simulation in Healthcare*, 1(1), 23–25. <https://doi.org/10.1097/01266021-200600110-00001>.
- Duff, J. P., Morse, K. J., Seelandt, J., Gross, I. T., Lydston, M., Sargeant, J., Dieckmann, P., Allen, J. A., Rudolph, J. W., & Kolbe, M. (2024). Debriefing methods for simulation in Healthcare: A systematic review. *Simulation in Healthcare: Journal of the Society for Simulation in Healthcare*, 19(15), S112–S121. <https://doi.org/10.1097/SIH.0000000000000765>.
- INACSL Standards Committee, Decker, S., Sapp, A., Bibin, L., Chidume, T., Crawford, S. B., ... Szydlowski, J. (2025). Healthcare simulation standards of best practice@ the debriefing process. *Clinical Simulation in Nursing*, 105, 101775. <https://doi.org/10.1016/j.ecns.2025.101775>.
- Ing, L., Cheng, A., & Lin, Y. (2022). Debriefing for simulation-based medical education: A survey from the International network of simulation-based Pediatric innovation, research and education. *Simulation in Healthcare: Journal of the Society for Simulation in Healthcare*, 17(1), 1–6. <https://doi.org/10.1097/SIH.0000000000000541>.
- Kainth, R., & Reedy, G. (2024). Transforming professional identity in simulation debriefing: A systematic metaethnographic synthesis of the simulation literature. *Simulation in Healthcare: Journal of the Society for Simulation in Healthcare*, 19(2), 90–104. <https://doi.org/10.1097/SIH.0000000000000734>.
- Kihlgren, P., Spanager, L., & Dieckmann, P. (2015). Investigating novice doctors' reflections in debriefings after simulation scenarios. *Medical Teacher*, 37(5), 437–443. <https://doi.org/10.3109/0142159X.2014.956054>.
- Lomuscio, S., Capogna, E., Sironi, S., Sguanci, M., Morales Palomares, S., Cangelosi, G., Ferrara, G., Mancin, S., Amodeo, A., Destrebecq, A., Parozzi, M., & Dal Bello, S. (2025). Debriefing methodologies in nursing simulation: An exploratory study of the Italian settings. *Nursing Reports*, 15(1), 7. <https://doi.org/10.3390/nursrep15010007>.
- McDonnell, L. K., Jobe, K. K., & Dismukes, R. K. (1997). *Facilitating LOS debriefings: A training manual* (NASA-TM-112192; DOT/FAA/AR-97/6). NASA Ames Research Center <https://ntrs.nasa.gov/citations/19970015346>.
- Olvera-Cortés, H. E., Hernández-Gutiérrez, L. S., Scherer-Castaneda, E., Ramírez-Arias, J. D., Argueta Muñoz, F. D., & Gutiérrez-Barreto, S. E. (2022). Who can debriefing assessment for simulation in healthcare? Different perspectives on teledebriefing assessments. *Simulation in Healthcare: Journal of the Society for Simulation in Healthcare*, 17(5), 308–312. <https://doi.org/10.1097/SIH.0000000000000642>.
- Umair, M., Mertens, J., Albert, S., & de Ruiter, J. P. (2022). GailBot: An automatic transcription system for conversation analysis. *Dialogue & Discourse*, 13(1), 63–95. <https://doi.org/10.5210/dad.2022.103>.