Title

Development of a lecture evaluation tool rooted in cognitive load theory: A modified Delphi study.

Journal

AEM Education and Training, 7(1)

Authors

Haas, Mary
Hickam, Grace
Murray, Collyn
et al.

Publication Date

2023-02-01

DOI

10.1002/aet2.10839

Peer reviewed

Received: 16 September 2022 | Revised: 22 November 2022 | Accepted: 2 December 2022

DOI: 10.1002/aet2.10839

ORIGINAL CONTRIBUTION

Development of a lecture evaluation tool rooted in cognitive load theory: A modified Delphi study

Jaime Jordan MD, MAED[1] | Mary R. C. Haas MD, MHPE[2] | Grace Hickam MD[3] | Collyn T. Murray MC, MACM[4] | Jefferey Hill MD, MEd[5] | Stephen J. Cico MD, MEd[6] | Meg Wolff MD, MHPE[2] | David E. Manthey MD[7] | Jason C. Wagner MD[8] | Sally A. Santen PhD[2,3]

[1]Department of Emergency Medicine, David Geffen School of Medicine at UCLA, Los Angeles, California, USA

[2]Department of Emergency Medicine, University of Michigan Medical School, Ann Arbor, Michigan, USA

[3]Department of Emergency Medicine, Virginia Commonwealth University, Richmond, Virginia, USA

[4]Department of Emergency Medicine, University of North Carolina Chapel Hill, Chapel Hill, North Carolina, USA

[5]Department of Emergency Medicine, University of Cincinnati, Cincinatti, Ohio, USA

[6]Department of Emergency Medicine, University of Central Florida College of Medicine, Orlando, Florida, USA

[7]Department of Emergency Medicine, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA

[8]Department of Emergency Medicine, Washington University in St. Louis, St. Louis, Missouri, USA

Correspondence
Jaime Jordan, MD, MAEd, Department of Emergency Medicine, David Geffen School of Medicine at UCLA, UCLA Emergency Medicine, 924 Westwood Boulevard, Suite 300, Los Angeles, CA 90095, USA.
Email: jaimejordanmd@gmail.com

Supervising Editor: Dr. Lalena Yarris.

Abstract

Background: Didactics play a key role in medical education. There is no standardized didactic evaluation tool to assess quality and provide feedback to instructors. Cognitive load theory provides a framework for lecture evaluations. We sought to develop an evaluation tool, rooted in cognitive load theory, to assess quality of didactic lectures.

Methods: We used a modified Delphi method to achieve expert consensus for items in a lecture evaluation tool. Nine emergency medicine educators with expertise in cognitive load participated in three modified Delphi rounds. In the first two rounds, experts rated the importance of including each item in the evaluation rubric on a 1 to 9 Likert scale with 1 labeled as "not at all important" and 9 labeled as "extremely important." In the third round, experts were asked to make a binary choice of whether the item should be included in the final evaluation tool. In each round, the experts were invited to provide written comments, edits, and suggested additional items. Modifications were made between rounds based on item scores and expert feedback. We calculated descriptive statistics for item scores.

Results: We completed three Delphi rounds, each with 100% response rate. After Round 1, we removed one item, made major changes to two items, made minor wording changes to nine items, and modified the scale of one item. Following Round 2, we eliminated three items, made major wording changes to one item, and made minor wording changes to one item. After the third round, we made minor wording changes to two items. We also reordered and categorized items for ease of use. The final evaluation tool consisted of nine items.

Conclusions: We developed a lecture assessment tool rooted in cognitive load theory specific to medical education. This tool can be applied to assess quality of instruction and provide important feedback to speakers.

INTRODUCTION

Medical educators have employed the didactic lecture format since 1850.[1] Although utilization of educational methods that emphasize active learning is increasing, didactics continue to play a key role in medical education.[2,3] Feedback obtained through didactic lecture evaluation tools benefit instructors, learners, and curriculum planners alike by catalyzing the improvement of future iterations. Accordingly, the Accreditation Council for Continuing Medical Education (ACCME) requires the evaluation of continuing medical education activities for effectiveness, although it does not provide specifics about the optimal approach.[4] Further, evaluation of teaching excellence contributes to promotion.

Cognitive load theory is a learning theory that has been increasingly recognized in medical education and is particularly relevant given the workload and complexity of knowledge to be acquired by medical trainees.[5–7] Limited data suggest that optimizing cognitive load can positively influence learning outcomes.[6] Cognitive load theory provides a useful framework for centering lecture evaluation tools around features that directly impact learning of the material presented. Cognitive load theory explores the relationship between working memory and long-term memory.[5–7] Three types of cognitive load impact the transfer of information from working to long-term memory, including intrinsic, extrinsic, and germane load. Intrinsic cognitive load refers to the inherent complexity of a particular topic.[8] Instructors can minimize intrinsic load by modifying how they present material or choosing to limit the amount of material covered.[9] Extrinsic load represents the resources required to process material (including extraneous stimuli) and inhibits learning.[8] Instructors minimize it by reducing environmental distractions such as noise and extraneous talking, focusing content on learning objectives, and using visual aids that augment rather distract from key learning points.[9] Germane load facilitates learning through the process of organizing new data into schema to consolidate information.[8] Instructors can promote germane load by grouping information in meaningful ways.[9] Understanding the interplay between the different types of cognitive load can help instructors optimize the transfer of knowledge from working to long-term memory during educational experiences, such as didactic lectures.

Currently, there is no standardized didactic evaluation tool available for widespread use in medical education. While prior work has investigated lecture evaluation tools for emergency medicine (EM) residents, few available tools evaluate the effectiveness of instruction in the context of cognitive load.[10,11] A trial of one evaluation tool, adapted from Leppink et al,[12] demonstrated good internal validity for two components of cognitive load.[10,11] However, the original tool was created for evaluation of lectures given within the undergraduate university setting and adapted for use in graduate medical education.[10,11] Given the unique instructional setting of graduate medical education, with its advanced learners and complex knowledge to be acquired, a tool should be developed specifically for use in this setting and aimed at the goals of EM didactic instruction. Thoughtful design of didactic evaluation tools has been shown to enhance the quality and quantity of feedback obtained.[13] The development of a didactic lecture evaluation tool rooted in cognitive load theory can generate valuable feedback for lecturers of all levels, facilitate assessment and teaching of didactic skills, and enhance learning outcomes by encouraging effective presentation of content. The objective of this study was to develop an evaluation tool, rooted in cognitive load theory, to assess the quality of didactic lectures in graduate medical education including content, presentation, and delivery.

METHODS

Study design

We utilized a modified Delphi technique to achieve consensus on items for an evaluation tool, based on cognitive load theory, to assess the quality of didactic lectures. The modified Delphi technique is a systematic group consensus strategy designed to increase content validity.[14] This study was reviewed by the institutional review board of the David Geffen School of Medicine at UCLA and determined to be exempt.

Study setting and participants

The first author invited nine EM educators with expertise in cognitive load from diverse regions across the United States who have been working on evaluating cognitive load to participate in the Delphi panel.[9,11] All had published research incorporating cognitive load theory. Previous studies have recommended that six to 10 experts is an appropriate number for obtaining stable results in the modified Delphi method.[15–17] All invited panelists agreed to participate. The panel consisted of three professors, three associate professors, and three assistant professors. Six members held advanced degrees in education related fields. We collected data between February and May 2022.

Study protocol

The first author, who was not a member of the Delphi panel, drafted initial items for the evaluation tool after review of the literature and other tools utilizing cognitive load theory to optimize content validity. We utilized an electronic survey platform (SurveyMonkey) to administer and collect data from the Delphi surveys.[18] In the first two rounds of the modified Delphi process, expert panelists rated the importance of including each item in the evaluation rubric on a 1 to 9 Likert scale with 1 labeled as "not at all important" and 9 labeled as "extremely important." We determined a priori that items with a mean score of 7 or greater advanced to the next round and items

with a mean score of 3 or below were eliminated. The first author, who served as the Delphi panel moderator, applied discretion for items with mean scores between 4 and 6, with the aim of both adhering to the opinions of the experts and creating a comprehensive evaluation tool. Each item consisted of a stem and anchored choices with associated point-value assignments. In each round, the experts were invited to provide additional written comments, edits, and suggestions for each item and given an opportunity to suggest additional items that were not included initially. The moderator made modifications between rounds based on scores and feedback from panelists. After each round, the moderator provided panelists with summary mean item scores, written comments, and an edited version of the items derived from the results of the previous round. The panelists were then asked to rate the revised items and provide additional edits or suggestions. In the third round, panelists were asked to make a binary choice of whether the item should be included in the final evaluation tool. For this round, we determined an inclusion threshold of 75% "yes" a priori. Similar to prior rounds, in the third round, we also invited experts to provide written comments, edits, and suggestions. After consensus was achieved, we created a final evaluation tool.

Data analysis

We calculated and reported descriptive statistics for item scoring during Delphi rounds.

RESULTS

We completed three Delphi rounds (Table 1). The response rate was 100% for each. After Round 1, we removed one item, made major changes to two items, made minor wording changes to nine items, and modified the scale of one item. No additional items were added. Following Round 2, we eliminated three items, made major wording changes to one item, and made minor wording changes to one item. No additional items were added. After the third round, we made minor wording changes to two items. We also reordered and categorized items for ease of use. All items met our a priori threshold for inclusion and no additional items were suggested. Thus, we determined that we had achieved consensus. All panelists approved the final evaluation tool which contained nine items and is displayed in Figure 1.

DISCUSSION

Use of a robust instrument to measure cognitive load can assist faculty and residents in optimizing didactic sessions to enhance learning outcomes. This study provides initial content validity evidence for an instrument tailored to graduate medical education. The instrument incorporates all three forms of cognitive load (intrinsic, extrinsic, and germane). The instrument is intended to evaluate different facets of a didactic session and provide meaningful, actionable feedback to the presenter using the evidence principles of cognitive load education theory.

Measures of cognitive load include self-reported instruments, measures of psychological parameters, and tools for learners.[12,19,20] There are several published cognitive load instruments.[12,19,20] Our group previously collected validity evidence on the instrument by Leppink et al.[11,12] in the graduate medical virtual didactic setting. We found the 10-item cognitive load instrument demonstrated reasonable validity evidence.[11] While the tool of Leppink et al.[12] and others measure only cognitive load, this current instrument includes additional measures of excellence in didactics. The instrument highlights specific behaviors that can be modified. The instrument was developed to assess features of each type of cognitive load (intrinsic, extrinsic, germane) and to look for behaviors that are commonly used to combat cognitive load to enhance learning. For example, the speaker may slow down the pace of their speech when discussing a particularly challenging concept.

Intrinsic cognitive load is inherent to both what is being taught and to whom it is being taught. Teaching advanced calculus is intrinsically difficult if taught to students who are versed in calculus but tremendously more difficult to those who have never been introduced to calculus. The speaker can decrease intrinsic load by tailoring their content according to the audience's knowledge and utilizing strategies such as limiting the amount of new, complex information; presenting content in various manners; and reactivating prior knowledge. Our instrument specifically includes items on content difficulty, volume of material, organization, and activation of prior learning. A speaker can readily modify extrinsic cognitive load through control of the learning environment by minimizing disruption and presenting information in a way that harmonizes visual and auditory information. Our instrument incorporates extrinsic load by assessing content relevance, audiovisual presentation format, and speaker delivery. Presenters should increase, rather than limit, germane load, which facilitates learning through strategies such as helping learners form schemas or organized ways of thinking about the topic. Our instrument assesses multiple aspects of facilitation of learning.

This lecture evaluation tool has multiple potential applications in an educational system. First, educators can readily apply it across didactic programming and curricula to assess quality of instruction. Second, it can serve as an effective feedback tool for instructors. It diverges from typical affective domain items used commonly in evaluations that focus on entertainment and enjoyability in favor of those rooted in theory about effective instruction. Beyond content, results gathered from this evaluation can focus attention on resources, design, and personal factors. Improved feedback to educators on their didactics should enhance the quality of future instruction.[21] Speakers can also utilize this tool as they plan their presentations as a design rubric that allows them to proactively incorporate cognitive load theory. Training programs can use it as part of an evaluation system for speakers, providing consistency in the evaluation process and

TABLE 1 Results of modified Delphi process

Scoring rubric items	Delphi process		
	Round 1[a]	Round 2[a]	Round 3[b]
Item 1			
Content relevance 0 = Content is not relevant to resident learning of EM 1 = Content is somewhat relevant to resident learning of EM 2 = Content is relevant to resident learning of EM	6.78		
Content relevance 0 = Content is not relevant to learners in the audience 1 = Content is somewhat relevant to learners in the audience 2 = Content is relevant to learners in the audience		7.33	
Content relevance 0 = Content is not relevant to the majority of learners in the audience 1 = Content is somewhat relevant to the majority of learners in the audience 2 = Content is relevant to the majority of learners in the audience			8
Item 2			
Content alignment with objectives 0 = Content is not aligned with learning objectives 1 = Content is somewhat aligned with the learning objectives 2 = Content is aligned with the learning objectives	5.44		
Item 3			
Management of environment 0 = There are many environmental distractions (poor lighting, dysfunctional audio/visual technology, interrupting alerts, etc.) during the didactic session 1 = There are occasional environmental distractions during the didactic session 2 = The lecturer takes great care to minimize environmental distractions during the didactic session	7.00		
Management of environment 0 = There were many environmental distractions not mitigated by the speaker 1 = There were some environmental distractions not mitigated by the speaker 2 = The speaker minimized nearly all environmental distractions		6.11	
Item 4			
Slides and audiovisual design 0 = Slides/audiovisual aids distract from the didactic session (not aligned with session goals and objectives, excessive text, font that's too small to read, etc.) 1 = Slides/audiovisual aids are adequate, but could be improved 2 = Slides/audiovisual aids augment the didactic session (aligned with objectives, focus on images rather than text, no distracting extraneous pictures/GIFs, etc.)	8.44		
Slides and audiovisual design 0 = Slides/audiovisual aids distract from the didactic session (material is not aligned with topic, excessive text, font that is too small to read, etc.) 1 = Slides/audiovisual aids neither augment nor distract from the didactic session 2 = Slides/audiovisual aids augment the didactic session (aligned with topic, focus on images rather than text, no distracting extraneous pictures/GIFs, etc.)		8.78	
Slides and audiovisual design 0 = Slides/audiovisual aids distract from the didactic session (material is not aligned with topic, excessive text, font that is too small to read, etc.) 1 = Slides/audiovisual aids neither augment nor distract from the didactic session 2 = Slides/audiovisual aids augment the didactic session (aligned with topic, focus on images rather than text, no distracting extraneous pictures/GIFs, etc.)			9
Item 5			
Lecturer preparedness 0 = Lecturer is unprepared 1 = Lecturer is somewhat prepared and demonstrates basic knowledge of topic 2 = Lecturer is well prepared and demonstrates command knowledge of topic	6.89		

TABLE 1 (Continued)

Scoring rubric items	Delphi process		
	Round 1[a]	Round 2[a]	Round 3[b]
Speaker preparedness 0 = Speaker appears unprepared 1 = Speaker appears somewhat prepared and demonstrates basic knowledge of topic 2 = Speaker appears well prepared and demonstrates command knowledge of topic		6.67	
Item 6			
Lecturer verbal presentation 0 = The lecture's verbal presentation is ineffective (too fast or too slow, inappropriate volume, often uses filler words like "um," etc.) 1 = The lecture's verbal presentation is somewhat effective 2 = The lecturer's verbal presentation is effective (good pace of speech, volume, etc.)	7.89		
Speaker verbal presentation 0 = The speaker's verbal presentation is detrimental to learning (too fast or too slow, inappropriate volume, often uses filler words like "um," etc.) 1 = The speaker's verbal presentation is somewhat effective for learning, but with significant room for improvement 2 = The speaker's verbal presentation is effective for learning (good pace of speech, volume, etc.)		7.67	
Speaker verbal presentation 0 = The speaker's verbal presentation is detrimental to learning (too fast or too slow, inappropriate volume, often uses filler words like "um," etc.) 1 = The speaker's verbal presentation is somewhat effective for learning, but with significant room for improvement 2 = The speaker's verbal presentation is effective for learning (good pace of speech, volume, etc.)			9
Item 7			
Content difficulty 0 = Content is either too easy or too difficult for level of learner 1 = Content is mostly appropriate for level of learner but at times may be too easy or too difficult 2 = Content is consistently appropriate difficulty for level of learner	7.56		
Content difficulty 0 = Speaker frequently provides content that is either too easy or too difficult for the levels of learners present 1 = Content is mostly appropriate for level of learner but at times may be too easy or too difficult (or if the audience consists of multiple learner levels, speaker only provides content of appropriate difficulty to a single level of learner) 2 = Speaker consistently provides content of appropriate difficulty for all level of learners present		7.89	
Content difficulty 0 = Speaker frequently provides content that is either too easy or too difficult for the levels of learners present 1 = Content is mostly appropriate for level of learner but at times may be too easy or too difficult (or if the audience consists of multiple learner levels, speaker only provides content of appropriate difficulty to a single level of learner) 2 = Speaker consistently provides content of appropriate difficulty for all level of learners present			9
Item 8			
Volume of material 0 = Volume of material covered in lecture is inappropriate for the time allotted (too much or too little) 1 = Volume of material covered in lecture is somewhat appropriate for the time allotted 2 = Volume of material covered in lecture is very appropriate for the time allotted	7.44		
Volume of material 0 = Volume of material covered in the didactic session is inappropriate for the time allotted (too much or too little) 1 = Volume of material covered in the didactic session is somewhat appropriate for the time allotted 2 = Volume of material covered in lecture is optimized for the time allotted		8.44	

(Continues)

TABLE 1 (Continued)

Scoring rubric items	Delphi process		
	Round 1[a]	Round 2[a]	Round 3[b]
Volume of material 0 = Volume of material covered in the didactic session is inappropriate for the time allotted (too much or too little) 1 = Volume of material covered in the didactic session is somewhat appropriate for the time allotted 2 = Volume of material covered in lecture is optimized for the time allotted			9
Item 9			
Organization of delivery 0 = Presentation is disorganized 1 = Material is presented in a linear fashion, but could be improved 2 = Material is optimally organized and presented from simple to complex to facilitate learning	8.33		
Organization of delivery 0 = Presentation is disorganized, hindering learning 1 = Material is presented in a linear fashion, but could be improved 2 = Material is optimally organized to facilitate learning		8.22	
Organization of delivery 0 = Presentation is disorganized, hindering learning 1 = Material is presented in a linear fashion, but could be improved 2 = Material is optimally organized to facilitate learning			9
Item 10			
Activation of learner knowledge 0 = The lecturer does not activate learner prior knowledge (i.e., questioning, review of important content) 1 = The lecturer occasionally activates learner prior knowledge (i.e., questioning, review of important content) 2 = The lecturer often activates learner's prior knowledge (i.e., questioning, review of important content)	7.78		
Activation of learner prior knowledge 0 = The speaker does not activate learner prior knowledge (i.e., questioning, review of important content) 1 = The speaker occasionally activates learner prior knowledge (i.e., questioning, review of important content) 2 = The speaker often activates learner's prior knowledge (i.e., questioning, review of important content)		7.78	
Activation of learner prior knowledge 0 = The speaker does not activate learner prior knowledge (i.e., questioning, review of important content) 1 = The speaker occasionally activates learner prior knowledge (i.e., questioning, review of important content) 2 = The speaker often activates learner's prior knowledge (i.e., questioning, review of important content)			8
Item 11			
Audience engagement 0 = The lecturer does not engage the audience 1 = The lecturer occasionally engages the audience (i.e., audience response system, interactive questions, small group discussions or activities, worked examples) 2 = The lecturer often engages the audience (i.e., audience response system, interactive questions, small group discussions or activities, worked examples)	7.00		
Audience engagement 0 = The speaker does not engage the audience 1 = The speaker occasionally engages the audience (i.e., shares a story, asks questions that require a verbal response, creates small group discussions or activities, constructs worked examples) 2 = The speaker often engages the audience to facilitate learning (i.e., shares a story, asks questions that require a verbal response, creates small group discussions or activities, constructs worked examples)		7.11	

TABLE 1 (Continued)

Scoring rubric items	Delphi process		
	Round 1[a]	Round 2[a]	Round 3[b]
Facilitation of learning 0 = The speaker's presentation of material does not facilitate learning 1 = The speaker's presentation occasionally facilitates learning (i.e., shares a story, encourages reflection, groups materials in meaningful ways, asks questions that require a verbal response, creates small group discussions or activities, constructs worked examples, repetition) 2 = The speaker often engages the audience to facilitate learning (i.e., shares a story, encourages reflection, groups material in meaningful ways, asks questions that require a verbal response, creates small group discussions or activities, constructs worked examples, repetition)			8
Item 12			
Facilitation of learning 0 = The lecturer never "chunks" or groups information in meaningful ways 1 = The lecturer occasionally "chunks" or groups information in meaningful ways 2 = The lecturer often "chunks" or groups information in meaningful ways	7.78		
Facilitation of learning (consider potential techniques to facilitate learning such as grouping material in meaningful ways, interleaving, repetition, application of information, etc.) 0 = The speaker's presentation of material did not facilitate my learning at all 1 = The speaker's presentation style somewhat facilitated my learning 2 = The speaker's presentation style greatly facilitated my learning		6.78	
Item 13			
Overall quality of lecture 0 = Poor 1 = Average 2 = Outstanding	7.11		
Overall quality of lecture 0 = Poor 1 = Fair 2 = Good 3 = Excellent 4 = Outstanding		8.44	
Overall quality of lecture 0 = Poor 1 = Fair 2 = Good 3 = Excellent 4 = Outstanding			9

[a]Panelists rated items on 1 to 9 scale of importance to include (1 = not at all important include, 9 = extremely important to include). Results reported: mean score.

[b]Panelists voted yes/no on whether an item should be included in the final evaluation tool. Results reported: frequency of "yes" response (total $n = 9$).

meeting standards set forth by institutional requirements and regulatory bodies. When applied systematically, this added consistency and quality in evaluation can enhance the future delivery of education across training programs. Finally, it can provide evidence for excellence in teaching for educator's portfolios under consideration for promotion. This tool was designed to be practical and easy to use. We have demonstrated content validity evidence, but have not determined how this instrument will perform in practice. Further research is required to provide additional validity evidence to support its use. While we anticipate that this tool will provide higher quality feedback that will lead to enhanced future instruction, future studies evaluating changes in speaker behavior and assessing the impact of its broader utilization can examine this hypothesis.

LIMITATIONS

The study has limitations. Choice of Delphi panel experts may not be representative of the larger pool of medical educators. It is possible that a larger or differently composed panel may have yielded different results. Additionally, as this method was implemented electronically, there may be limited discussion and elaboration. With three performance assessment responses for each item, it is possible that the evaluations will lack discriminatory power particularly for moderately or highly experienced lecturers. Also, the lack of a narrative feedback section on the evaluation could limit an evaluator's opportunity to provide specific feedback suggesting actionable improvements for future lectures.

Lecture Evaluation

Content

1) Content relevance

0 = Content is not relevant to the majority of learners in the audience
1 = Content is somewhat relevant to the majority of learners in the audience
2 = Content is relevant to the majority of learners in the audience

Score: _____

2) Content difficulty

0 = Speaker frequently provides content that is either too easy or too difficult for the levels of learners present
1 = Content is mostly appropriate for level of learner but at times may be too easy or too difficult *(or if the audience consists of multiple learner levels, speaker only provides content of appropriate difficulty to a single level of learner)*
2 = Speaker consistently provides content of appropriate difficulty for all level of learners present

Score: _____

3) Volume of material

0 = volume of material covered in the didactic session is inappropriate for the time allotted *(too much or too little)*
1 = volume of material covered in the didactic session is somewhat appropriate for the time allotted
2 = Volume of material covered in lecture is optimized for the time allotted

Score: _____

Presentation & Delivery

4) Slides and audio-visual design

0 = Slides/audio-visual aids distract from the didactic session *(i.e. material is not aligned with topic, excessive text, font that's too small to read, etc.)*
1 = Slides/audio-visual aids neither augment nor distract from the didactic session
2 = Slides/audio-visual aids augment the didactic session *(i.e. aligned with topic, focus on images rather than text, no distracting extraneous pictures/gifs, etc.)*

Score: _____

5) Speaker verbal presentation

0 = The speaker's verbal presentation is detrimental to learning *(i.e. too fast or too slow, inappropriate volume, often uses filler words like "um", etc.)*
1 = The speaker's verbal presentation is somewhat effective for learning, but with significant room for improvement
2 = The speaker's verbal presentation is effective for learning *(i.e. good pace of speech, volume, etc.)*

Score: _____

FIGURE 1 Final lecture evaluation tool

6) Organization of presentation

0 = Presentation is disorganized, hindering learning
1 = Material is presented in a linear fashion, but could be improved
2 = Material is optimally organized to facilitate learning

Score: _____

Instructional Techniques

7) Activation of learner prior knowledge

0 = The speaker does not activate learner prior knowledge *(e.g. questioning, review of important content, etc.)*
1 = The speaker occasionally activates learner prior knowledge *(e.g. questioning, review of important content, etc.)*
2 = The speaker often activates learner's prior knowledge *(e.g. questioning, review of important content, etc.)*

Score: _____

8) Facilitation of learning *(Example techniques that facilitate learning include: story-telling, encouraging reflection, grouping material in meaningful ways, asking questions that require a verbal response, creating small group discussions or activities, use of worked examples, repetition, etc.)*

0 = The speaker's presentation of material does not facilitate learning
1 = The speaker's presentation occasionally facilitates learning
2 = The speaker often engages the audience to facilitate learning

Score: _____

Summary

9) Overall quality of lecture

0 = Poor
1 = Fair
2= Good
3 = Excellent
4= Outstanding

Score: _____

Total Score: _____/20

FIGURE 1 Continued

CONCLUSIONS

We developed a lecture assessment tool rooted in cognitive load theory specific to graduate medical education. Using a modified Delphi consensus building process, we derived a final assessment tool with nine items clustered into three domains: content, presentation and delivery, and instructional techniques. This tool can be applied to assess quality of instruction and provide important feedback to speakers.

AUTHOR CONTRIBUTIONS

Jaime Jordan, Mary R. C. Haas, Grace Hickam, Collyn T. Murray, Jefferey Hill, Stephen J. Cico, Meg Wolff, David E. Manthey, Jason C. Wagner, and Sally A. Santen conceived and designed the study. Jaime Jordan acquired and interpreted the data. Jaime Jordan, Mary R. C. Haas, Grace Hickam, Collyn T. Murray, Jefferey Hill, Stephen J. Cico, Meg Wolff, David E. Manthey, Jason C. Wagner, and Sally A. Santen contributed to drafting and critical revision of the manuscript.

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

ORCID

Jaime Jordan https://orcid.org/0000-0002-6573-7041
Mary R. C. Haas https://orcid.org/0000-0002-9506-5928
Grace Hickam https://orcid.org/0000-0002-6247-4530
Meg Wolff https://orcid.org/0000-0002-3637-2653
Jason C. Wagner https://orcid.org/0000-0001-8702-0706
Sally A. Santen https://orcid.org/0000-0002-8327-8002

REFERENCES

1. Matheson C. The educational value and effectiveness of lectures. *Clin Teach*. 2008;5:218-221.
2. Olson CA, Tooman TR. Didactic CME and practice change: don't throw that baby out quite yet. *Adv Health Sci Educ Theory Pract*. 2012;17:441-451.
3. Deiorio NM, Fitch MT, Jung J, et al. Evaluating educational interventions in emergency medicine. *Acad Emerg Med*. 2012;19(12):1442-1453.
4. The ACGME's essential areas and their elements. Accreditation Council for Graduate Medical Education. Accessed September 16, 2022. https://www.uhms.org/images/CME-Sponsorship/essential_areas_and_elements.pdf
5. Wissman AW. Cognitive load theory: applications in medical education. Technical Communication Capstone Course. 25. Minnesota State University, Mankato. 2018. Accessed September 16, 2022. https://cornerstone.lib.mnsu.edu/cgi/viewcontent.cgi?article=1024&context=eng_tech_comm_capstone_course
6. Venkat MV, O'Sullivan PS, Young JQ, Sewell JL. Using cognitive load theory to improve teaching in the clinical workplace. *MedEdPORTAL*. 2020;16:10983.
7. Young JQ, Van Merrienboer J, Durning S, Ten Cate O. Cognitive load theory: implications for medical education: AMEE guide no. 86. *Med Teach*. 2014;36(5):371-384.
8. Chandler P, Sweller J. Cognitive load theory and the format of instruction. *Cogn Instr*. 1991;8(4):293-332.
9. Jordan J, Wagner J, Manthey D, Wolff M, Santen S, Cico S. Optimizing lectures from a cognitive load perspective. *AEM Educ Train*. 2019;4(3):306-312.
10. Hill J, Stull M, Stettler M, Paulsen R, Hart K, McDonough E. Development and validation of a lecture assessment tool for emergency medicine residents. *AEM Educ Train*. 2018;2(4):310-316.
11. Hickam G, Jordan J, Haas MRC, et al. Validity evidence for an instrument for cognitive load for virtual didactic sessions. *AEM Educ Train*. 2022;6(1):e10718.
12. Leppink J, Paas F, Van der Vleuten CPM, Van Gog T, Van Merriënboer JJG. Development of an instrument for measuring different types of cognitive load. *Behav Res Method*. 2013;45(4):1058-1072.
13. Nelson AM, Traba CM. Redesigning evaluation forms improves didactic conference assessment. *Med Educ*. 2019;53(5):505-506.
14. Hasson F, Keeney S, Mckenna HP. Research guidelines for the Delphi survey technique research guidelines for the Delphi survey. *J Adv Nurs*. 2000;32(4):1008-1015.
15. Thangaratinam SK, Redman CW. The Delphi Technique. *Obstet Gynaecol*. 2005;7:120-125.
16. Pines JM, Alfaraj S, Batra S, et al. Factors important to top clinical performance in emergency medicine residency: results of an ide-ation survey and Delphi panel. *AEM Educ Train*. 2018;2(4):269-276.
17. Eubank BH, Mohtadi NG, Lafave MR, et al. Using the modified Delphi method to establish clinical consensus for the diagnosis and treatment of patients with rotator cuff pathology. *BMC Med Res Methodol*. 2016;16:56.
18. SurveyMonkey. 2021. Accessed September 9, 2022. https://www.surveymonkey.com/
19. Cierniak G, Scheiter K, Gerjets P. The scientific value of cognitive load theory: a research agenda based on the structuralist view of theories. *Educ Psychol Rev*. 2009;21:43-54.
20. DeLeeuw KE, Mayer RE. A comparison of three measures of cognitive load: evidence for separable measures of intrinsic, extraneous, and germane load. *J Educ Psychol*. 2008;100(1):223-234.
21. Nevins SR, Floden RE. Intensive mentoring as a way to help beginning teachers develop balanced instruction. *J Teacher Educ*. 2009;60(2):112-122.

How to cite this article: Jordan J, Haas MRC, Hickam G, et al. Development of a lecture evaluation tool rooted in cognitive load theory: A modified Delphi study. *AEM Educ Train*. 2023;7:e10839. doi:10.1002/aet2.10839