Author Manuscript Post-Print
Originally published in *Annals of Emergency Medicine*
http://dx.doi.org/10.1016/j.annemergmed.2016.02.018

Examining Reliability and Validity of an Online Score (ALiEM AIR) for Rating Free Open Access Medical Education Resources

Teresa Man-Yee Chan, BEd, MD[1,2]; Andrew Grock, MD[3]; Michael Paddock, DO, MS[4]; Kulamakan Kulasegaram, PhD[5]; Lalena M. Yarris, MD, MCR[6]; Michelle Lin, MD[2,7]

Study objective: Since 2014, Academic Life in Emergency Medicine (ALiEM) has used the Approved Instructional Resources (AIR) score to critically appraise online content. The primary goals of this study are to determine the interrater reliability (IRR) of the ALiEM AIR rating score and determine its correlation with expert educator gestalt. We also determine the minimum number of educator-raters needed to achieve acceptable reliability.
Methods: Eight educators each rated 83 online educational posts with the ALiEM AIR scale. Items include accuracy, usage of evidence-based medicine, referencing, utility, and the Best Evidence in Emergency Medicine rating score. A generalizability study was conducted to determine IRR and rating variance contributions of facets such as rater, blogs, posts, and topic. A randomized selection of 40 blog posts previously rated through ALiEM AIR was then rated again by a blinded group of expert medical educators according to their gestalt. Their gestalt impression was subsequently correlated with the ALiEM AIR score.
Results: The IRR for the ALiEM AIR rating scale was 0.81 during the 6-month pilot period. Decision studies showed that at least 9 raters were required to achieve this reliability. Spearman correlations between mean AIR score and the mean expert gestalt ratings were 0.40 for recommendation for learners and 0.35 for their colleagues.
Conclusion: The ALiEM AIR scale is a moderately to highly reliable, 5-question tool when used by medical educators for rating online resources. The score displays a fair correlation with expert educator gestalt in regard to the quality of the resources. The score displays a fair correlation with educator gestalt. [Ann Emerg Med. 2016; Online First, March 30, 2016)

Author Affiliations:

1. Division of Emergency Medicine, Department of Medicie, Hamilton, ON, Canada
2. MedEdLIFE Research Collaborative
3. Department of Emergency Medicine, University of California, Olive View, CA
4. Regions Hospital and the Department of Emergency Medicine, University of Minnesota
5. Department of Family and Community Medicine, Faculty of Medicine, University of Toronto, and the Wilson Centre for Health Professions Education, University Health Network, Toronto, ON, Canada
6. Department of Emergency Medicine, Oregon Health & Science University, Portland, OR
7. Department of Emergency Medicine, University of California, San Francisco, CA

Original Citation:

Chan, T. M.-Y., Grock, A., Paddock, M., Kulasegaram, K., Yarris, L. M., & Lin, M. (2016). Examining Reliability and Validity of an Online Score (ALiEM AIR) for Rating Free Open Access Medical Education Resources. Annals of Emergency Medicine, 1–7. Online First. http://dx.doi.org/10.1016/j.annemergmed.2016.02.018

MedEdLIFE
research collaborative

Introduction

Background

Medical educators aim to instill a habit of lifelong learning in their trainees. In the past, educators encouraged learners to read selected scientific articles and textbooks and then guided them in discussions in departmental conferences or institutional grand rounds. Today, 2 major factors have reshaped the landscape of lifelong learning toward a more open and globally crowd-sourced approach. The first factor has been the cultural adoption of evidence-based medicine, which is predicated on the dissemination of clinically relevant and methodologically sound scientific studies to clinical end users who might best apply this new knowledge in a process known as knowledge translation.[1] However, the half-life of scientific facts has been recently called into question by recent scholars, leading to a call for educators to provide learners with the ability to critically appraise content in addition to fostering a habit of continuous learning.[2] Developing these evaluative skills is key to allow learners to discern the intrinsic value and influence of what is reported and disseminated. Second, there has been a shift in publication away from publishing houses and increasingly into the hands of anyone with Internet access. Following the theoretical framework

Author Manuscript Post-Print
Originally published in *Annals of Emergency Medicine*

Table 1 | The Academic Life in Emergency Medicine's Approved Instructional Resources scoring instrument (Best Evidence in Emergency Medicine, BEEM; emergency physician, EP)

Score	BEEM Rater Scale	Content accuracy	Educational Utility	Evidence Based Medicine	Referencing
	Assuming that the results of this article are valid, how much does this article impact on EM clinical practice?	Do you have any concerns about the accuracy of the data presented or conclusions of this article?	Are there useful educational pearls in this article for residents?	Does this article reflect evidence based medicine (EBM) and thus lack bias?	Are the authors and literature clearly cited?
1	Useless information	Yes, many concerns from many inaccuracies	Low value: No valuable pearls	Not EBM based, only expert opinion	No
2	Not really interesting, not really new, changes nothing	-	-	-	-
3	Interesting and new, but doesn't change practice	Yes, a major concern about few inaccuracies	Yes, but there are only a few (1-2) valuable or multiple (≥3) less-valuable educational pearls	Minimally EBM based	-
4	Interesting and new, has the potential to change practice	-	-	-	Yes, authors and general references are listed (but no in-line references)
5	New and important: this would probably change practice for some EPs	Minimal concerns over minor inaccuracies	Yes, there are several (≥3) valuable educational pearls, or a few (1-2) KEY educational pearls that every resident should know before graduating	Mostly EBM based	-
6	New and Important: this would change practice for most EPs	-	-	-	-
7	This is a "must know" for EPs	No concerns over inaccuracies	Yes, there are multiple KEY educational pearls that residents should know before graduating	Yes exclusively EBM based (unbiased)	Yes, authors and in-line references are provided

BEEM, Best Evidence in Emergency Medicine; EBM, evidence-based medicine.

outlined by Christensen's disruptive innovation model, blogs, podcasts, and other social media have greatly democratized publishing.[3,4] This has led to a veritable explosion in the number of online medical education resources in the past decade.[5] As of 2013, there were 141 educational blogs and 42 podcasts in the fields of emergency medicine and critical care alone.[5] This growth has been mirrored in the scholarly literature, reporting innovative digital educational resources,[6] their popularity for self-study,[7-10] and integration of these resources into graduate medical education programs.[11,12] A case study of this phenomenon is the free open access medical education movement in the specialties of emergency medicine and critical care.[13,14] In this ever-changing environment, learners are avidly seeking online medical education resources and using them for cocurricular or extracurricular learning, despite a lack of instruction on the appropriate review of these secondary literature resources.[7-12] In an attempt to address the perceived lack of quality in online medical education resources, some Web sites are beginning to use an open, expert peer review process.15 However, blogs and podcasts rarely describe their quality assurance processes or disclose how content is selected. Others are using popularity (eg, social media followership) as a surrogate marker for quality, hoping that others may be following resources they deem of high quality.[16] Others may rely on educator recommendations, although this may be insufficient because often learners are using these online resources much more prevalently than the educators.[7] A scoring system to help health care providers determine the quality of online medical education resources for patients (eg, the DISCERN score)[17] has been described, but none have been described for clinical trainees and providers. Ultimately, adequate guidance in determining content quality is lacking.

Academic Life in Emergency Medicine's (ALiEM's) Approved Instructional Resources (AIR) series presents a novel review system that assigns a certification of quality to blog posts and podcasts.[18] In developing this certification, a novel scoring system was created and used. The primary objective of this study was to determine the interrater reliability and validity evidence for the ALiEM AIR scoring system. A secondary objective was to determine the minimum number of educator-raters needed to achieve acceptable reliability with the ALiEM AIR scoring system.

Methods

This project received institutional review exemption from the Hamilton Integrated Research Ethics Board chairperson.

From July to December 2014, 8 volunteer emergency medicine educators each rated 83 online educational posts with the AIR scoring instrument. Items included the Best Evidence in Emergency Medicine rating score (a previously validated scoring tool used to rate the clinical influence of primary scientific literature19,20), content accuracy, educational utility, evidence-based medicine, and referencing. The AIR scoring instrument is featured in Table 1. Although this included some content from ALiEM.com, we included content from many other sites as well. A complete listing of the ALiEM AIR-rated blog posts can be found in Appendix A. At the study, a small number of previously reviewed Web sites had gone offline. Therefore, only a total of 71 of the original 78 (91%) ALiEM AIR-reviewed educational resources were eligible for review during our present study.

Previous research has used generalizability studies to perform reliability analyses of multifaceted critical appraisal scoring systems.[19] Calculated generalizability coefficients function like other measures of reliability such as intraclass correlation but confer more ability to determine facets that might contribute to the variability of the scoring in multifaceted complex systems such as the one used in the ALiEM scoring tool. This technique takes into account that increasing the number of raters will increase reliability and that the contribution of variance of multiple facets may affect the rating system. In this study, the facets that may cause error variance in measurement of reliability are raters, curricular topic, and the nesting of blog or podcast posts within blogs or podcasts, or topic areas. Accounting for all of these facets in a reliability analysis allows a more accurate calculation of interrater reliability, both absolute (ϕ coefficient) and relative (ε coefficient) interrater reliability compared with a traditional intraclass correlation.[19,21]

A limitation of this approach is that confidence intervals cannot be calculated with precision because of the multiple sources of variance interacting with one another. An acceptable generalizability coefficient (ie, reliability score) of 0.7 has been used in previous literature for other scoring systems,[15] as well as higher-stakes decisions. However, the range for this coefficient depends on the context and use for the score that is generated.

Decision study analysis uses the reliability calculations
from a preceding generalizability study, taking into account each facet's variance, and allows a reliable estimate to minimize error when a multifaceted system such as the ALiEM AIR scoring tool is created.[19,21] Decision study analysis was used to predict the number of raters required for reliable outcomes of the ALiEM AIR scoring tool.[19,21]

Generalizability coefficients and decision studies were calculated with G-String IV (Hamilton, Ontario, Canada). For our validity studies, we chose to use Kane's validity framework to structure our evaluation of the ALiEM AIR scoring tool. The framework creates 4 inferences: score to observation, observation to universe score, universe score to target domain, and target domain to construct. In this study, the universe score was determined to be that of peer or educator recommendation. A novel educator recommendation gestalt rating tool was created to fulfill the requirement for determining a universe score. [22,23] The ALiEM AIR series rates blog posts from a given topic by manual search of the top 50 sites, as determined by the Social Media Index. [16] A complete tabular listing of all the posts and their AIR ratings can be found in Appendix B, along with a scatter plot that depicts the AIR ratings compared with our reference standards. Of these accessible blog posts, a random selection of 40 posts from within the ALiEM AIR-scored posts were separately reviewed and rated by a group of volunteer expert medical educators; all ALiEM AIR raters were excluded from being nominated into this pool. These expert medical educators were recruited through a peer nomination technique by the coinvestigators of this study. The complete listing of the posts rated by the expert educators can be found in Appendix B. The expert educators rated each item on 2 novel anchored 7-point Likert scales, which harnessed the educationalists' gestalt for good-quality educational resources. Because there were no previously derived scores for rating educational quality of medical professional-oriented materials, we chose to use an anchored scale to determine the educator's general gestalt for high-quality materials (Table 2).

Table 2 | Gestalt rating scale on blogs and podcasts for continuing medical education

Would you recommend this to a learner?	Unsure	1. No, this is an inappropriate resource for this audience	2	3	4. This may be useful to this audience	5	6	7. Yes this is a great resource for this audience
Would you recommend this to a staff to engage in continuing medical education?	Unsure	1. No, this is an inappropriate resource for this audience	2	3	4. This may be useful to this audience	5	6	7 Yes this is a great resource for this audience

Table 3: Demographic information on the ALiEM AIR Raters and Expert Raters

	ALiEM AIR Raters (n=8)		Expert Raters (n=20)	
% Male	62.5% male		90% male	
% by Country of Origin	100% United States of America		75% United States of America 25% Canada	
Average years in practice	11.9 years (SD 7.5)		10.3 years (SD 10.2)	
% with advanced degrees other than MD or DO degree	25%		60%	
% With academic title at institution	87.5% total		95% total	
	Breakdown		*Breakdown*	
	Full Professor	25%	Full Professor	10%
	Associate Professor	12.5%	Associate Professor	15%
	Assistant Professor	50%	Assistant Professor	65%
	Clinical Appointment	0%	Clinical Appointment	10%
	None	12.5%	None	5%
% with current or past official medical education position within institution*	87.5% total		90% total	
	Breakdown		*Breakdown*	
	Dean / Chair	12.5%	Dean / Chair	15%
	Residency PD	12.5%	Residency PD	40%
	Residency APD	50%	Residency APD	45%
	Other GME role	25%	Other GME role	30%
	Clerkship Director / UGME role	37.5%	Clerkship Director / UGME role	30%
	Research/Quality Improvement role	0%	Research/Quality Improvement role	20%

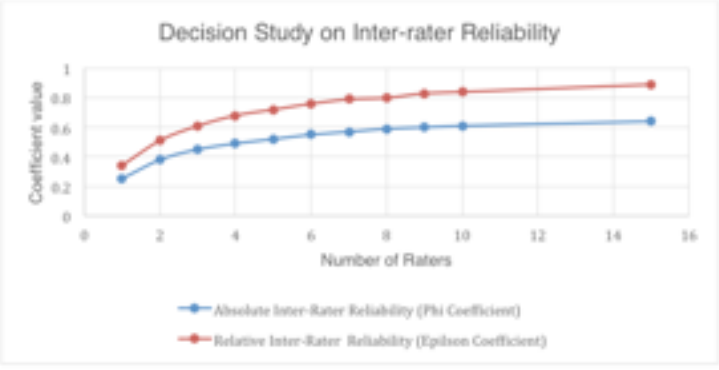
Program Director, PD; Associate or Assistant Program Director, APD; Graduate Medical Education, GME; Undergraduate Medical Education, UGME

Table 4 | Intra-class correlations between Expert Educator Gestalt Score and the ALiEM AIR Score

	Mean Expert Gestalt score for recommendation to colleague (another attending) for CME	Mean BEEM Rating Scale Score (subscore analysis)	AIR Mean Rating Score
Mean Educator Gestalt score for recommendation to Learner	0.71**	0.24*	0.40*
Mean Educator Gestalt score for recommendation to Colleague (Another attending) for CME	N/A	0.57**	0.35**

** Correlation is significant at the 0.05 level (2-tailed)*
*** Correlation is significant at the 0.01 level (2-tailed)*
Academic Life in Emergency Medicine's Approved Instructional Resources, ALiEM AIR; Continuing Medical Education, CME; Best Evidence in Emergency Medicine, BEEM

Figure 1 | The D-Study Results



The data were examined with histograms to check for parametricity. The Spearman correlation for absolute agreement between the mean AIR score and the mean expert gestalt ratings was calculated to assist with comparing expert educator gestalt with the AIR and emergency medicine scores. All calculations for the validity studies were completed with SPSS (version 23; IBM Corp).

Results

Two groups of raters, those who rated blogs or podcasts in the AIR series and those who provided expert gestalt ratings, were compared in this study. Their demographics are described in Table 3. A total of 78 blog posts were rated during a 6-month period. On average, the length of posts rated during the study period was 1,002 words (SD 715 words). When the nesting of articles by subject topic area was taken into account, the absolute interrater reliability across 9 raters for the ALiEM AIR rating scale was $\phi = 0.81$ during the 6-month pilot period. The decision study for the AIR score was conducted according to the complete ratings of all items within the initial 6-month period, which was a total of 80 posts (blogs or podcast-related posts). Decision studies showed that a minimum of 9 raters was required to achieve this reliability. Figure 1 reveals the decision study results. The Spearman correlation for absolute agreement between the mean AIR score and the mean expert gestalt ratings was 0.40 when an online medical education resource for learners was recommended and was 0.35 for their colleagues (Table 4). Overall, there was a moderate correlation (r=0.71) between the 2 gestalt ratings. The AIR performed similarly compared with the validated emergency medicine rating scale but had a great correlation with learners compared with fellow colleague gestalt scores, adding to its construct validity.[19,20]

Limitations

One interesting finding of the generalizability study was that there was a fairly large effect in the interactions between the subscore components and the items and blogs. This suggests that specific attributes of the blog Web sites themselves may have had a significant effect on the ratings, possibly as a result of a halo bias based on the specific Web sites. This may also have to do with inherent quality attributes of the Web sites' layout or other internal processes(eg, peer review processes).[27] Further study is clearly needed to determine what variables contribute most to quality within medical education blogs and podcasts.

Discussion

To keep up with the increasing body of medical

literature, Sir William Osler originally hosted a journal club to help update a community of practice about new articles, aiming to increase new knowledge uptake by clinicians.[24] Rapid dissemination, however, is now augmented by open-access, online media, resulting in phenomena such as online journal clubs[12,25-33] and other vehicles for more scientist-to-practitioner direct communication and conversation.[34] Online media—more specifically, online medical education resources—additionally reduce traditional educational barriers such as knowledge translation, awareness, and timeliness. Anecdotally, however, skepticism exists among educators and scholars, especially given that most online medical education resources lack the traditional peer-review process and that one might easily publish material online, using disruptive technologies such as blogs or podcasts. Although there are efforts to determine impact[16] and quality indicators for learner-oriented blogs and podcasts,[4,35,36] to our knowledge there has not yet been a scoring instrument developed to help guide educators in selecting high-quality blogs and podcasts for medical education. Recent efforts have led to the derivation of various lists of quality indicators for blogs and podcasts.[35,36] However, much in the vein of the JAMA Users' Guides to the Medical Literature series,[37] the limitations of these resources are that they are largely qualitative and lack robust psychometric analyses to support their use.

To our knowledge, the ALiEM AIR Series is the first nationwide attempt to evaluate online medical education resources, confirm their accuracy, and provide a postpublication scoring system for recommending quality resources to learners. It was initially created to address the Accreditation Council for Graduate Medical Education's Individualized Interactive Instructional curricular initiative to grant a portion of their conference attendance credit if they learned from asynchronous online medical education resources. There was a need to identify transparently vetted, quality resources that residency programs could trust and recommend to residents for such credit.

Although the ALiEM AIR 5-domain scoring instrument is not perfect, the results of our generalizability study and decision study suggest that it is reliable when there are at least 9 raters participating in the rating activity. We acknowledge that it may not be convenient to find 9 raters in most educational programs and as such recommend the use of this score by larger groups, perhaps connected through educational organizations such as the Council of Residency Directors of Emergency Medicine. Moreover, moderate to high correlation with educator recommendations for their peers for continuing education suggests that the resources recommended by the ALiEM AIR are high-quality medical education products that both learners and clinicians might use to augment traditional learning and encourage a culture of continuous education. Our study demonstrates that the ALiEM AIR series' scoring instrument is moderately to highly reliable in identifying quality online medical education resources, especially since it aligns well with the Accreditation Council for Graduate Medical Education's Individualized Interactive Instructional program goals.

The ALiEM AIR score is a moderately to highly reliable, 5-question scoring instrument when used by medical educators for rating online educational resources. Sources of variance in ratings may be a result of Web site interfaces. The score also correlates well with expert educator gestalt in regard to the quality of the resources. More research is needed to determine the contribution of Web site–specific elements toward perceptions of quality.

Acknowledgements: The authors acknowledge these colleagues for reviewing and rating the blog posts for this study: James Ahn, MD, University of Chicago; Matt Astin, MD, Mercer University School of Medicine; Robert Cooney, MD, Geisinger Health Systems; Sandy Dong, MD, University of Alberta; Justin Hensley, MD, Texas A&M/Christus Spohn; David Jones, MD, Oregon Health & Sciences University; Ian Julie, MD, University of California at Davis; Gloria Kuhn, DO, PhD, Wayne State University School of Medicine; Sean Moore, MD, University of Ottawa; Erik Nordquist, MD, Cook County Health and Hospital System; Dimitri Papanagnou, MD, MPH, Sidney Kimmel Medical College at Thomas Jefferson University; Alim Pardhan, MD, MBA, McMaster University; Hans Rosenberg, the University of Ottawa, MD; Christopher Ross, MD, Cook County Health and Hospital System; Mike Schindlbeck, MD, Cook County Health and Hospital System; Mark Silverberg, MD, Kings County; Nelson Wong, MD, Mount Sinai, NY; Andrew Worster, MD, McMaster University; Brian Wright, Stony Brook, MD. The authors also acknowledge the ALiEM AIR raters: Felix Ankel, MD, HealthPartners Institute; Jeremy Branzetti, MD, University of Washington; Nikita Joshi, MD, Stanford University; Eric Morley, MD, Stony Brook University; Anand Swaminathan, MD, New York University; Taku Taira, MD, University of Southern California; and Keeth Krishnan (University of Toronto medical student), who was invaluable as a research associate for this work.

References

1. Lang ES, Wyer PC, Haynes RB. Knowledge Translation: Closing the Evidence-to-Practice Gap. *Ann Emerg Med.* 2007;49(3):355-363. doi: 10.1016/j.annemergmed.2006.08.022.
2. Arbesman S. *The Half-Life of Facts: Why Everything We Know Has an Expiration Date.* First. New York, NY, USA: Penguin Group; 2012.
3. Christensen C, Raynor ME. *The Innovator's Solution: Creating and Sustaining Successful Growth.* Boston, MA: Harvard Business School Press; 2003.
4. Paterson QS, Thoma B, Milne WK, Lin M, Chan TM. A Systematic Review and Qualitative Analysis to Determine Quality Indicators for Health Professions Education Blogs and Podcasts. Journal of graduate medical education. 2015 Dec;7(4):549-54.
5. Cadogan M, Thoma B, Chan TM, Lin M. Free Open Access Meducation

(FOAM): the rise of emergency medicine and critical care blogs and podcasts (2002-2013). *Emerg Med J*. 2014:1-2. doi:10.1136/emermed-2013-203502.
6. Thoma B, Chan T, Benitez J, Lin M. Educational scholarship in the digital age: a scoping review and analysis of scholarly products. *The Winnower*. 2014;1(e141827.77297):1-13. doi:10.15200/winn.141827.77297.
7. Purdy E, Thoma B, Bednarczyk J, Migneault D, Sherbino J. The use of free online educational resources by Canadian emergency medicine residents and program directors. *Cjem Cjem Jcmu*. 2015;1717(22):101-106. doi:10.1017/cem.2014.73.
8. Mallin M, Schlein S, Doctor S, Stroud S, Dawson M, Fix M. A survey of the current utilization of asynchronous education among emergency medicine residents in the United States. *Acad Med J Assoc Am Med Coll*. 2014;89(4):598-601. doi:10.1097/ACM.0000000000000170.
9. Pearson D, Bond M, Kegg J, et al. Evaluation of Social Media Use by Emergency Medicine Residents and Faculty. *West J Emerg Med*. 2015;16(5):715-720. doi:10.5811/westjem.2015.7.26128.
10. Thurtle N, Banks C, Cox M, Pain T, Furyk J. Free Open Access Medical Education resource knowledge and utilisation amongst Emergency Medicine trainees: A survey in four countries. *African J Emerg Med*. 2015;(December). doi:http://dx.doi.org/10.1016/j.afjem.2015.10.005.
11. Scott KR, Hsu CH, Johnson NJ, Mamtani M, Conlon LW, DeRoos FJ. Integration of social media in emergency medicine residency curriculum. *Ann Emerg Med*. 2014;64(4):396-404. doi:10.1016/j.annemergmed.2014.05.030.
12. Hayes BD, Kobner S, Trueger NS, Yiu S, Lin M. Social Media in the Emergency Medicine Residency Curriculum: Social Media Responses to the Residents' Perspective Article. *Ann Emerg Med*. 2015;65(5):573-583. doi:10.1016/j.annemergmed.2015.01.018.
13. Cadogan M, Nickson C. Free Open Access Medical Education. Life in the Fast Lane. http://lifeinthefastlane.com/foam/. Published 2014. Accessed January 9, 2016.
14. Weingart SD, Thoma B. The online hierarchy of needs: A beginner's guide to medical social media and FOAM. *Emerg Med Australas*. 2015;27(1):5. doi:10.1111/1742-6723.12361.
15. Thoma B, Chan T, Desouza N, Lin M. Implementing peer review at an emergency medicine blog: bridging the gap between educators and clinical experts. *CJEM*. 2015;17(2):188-191. doi:10.2310/8000.2014.141393.
16. Thoma B, Sanders JL, Lin M, Paterson QS, Steeg J, Chan TM. The social media index: measuring the impact of emergency medicine and critical care websites. *West J Emerg Med*. 2015;16(2):242-249. doi:10.5811/westjem.2015.1.24860.
17. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. *J Epidemiol Community Health*. 1999;53(2):105-111. doi:10.1136/jech.53.2.105.
18. Lin M, Joshi N, Grock A, et al. Approved Instructional Resources (AIR) Series: A national initiative to identify quality emergency medicine blog and podcast content for resident education. *J Grad Med Educ*. 2016;8(2):In press.
19. Worster A, Kulasegaram K, Carpenter CR, et al. Consensus conference follow-up: inter-rater reliability assessment of the Best Evidence in Emergency Medicine (BEEM) rater scale, a medical literature rating tool for emergency physicians. *Acad Emerg Med*. 2011;18(11):1193-1200. doi:10.1111/j.1553-2712.2011.01214.x.
20. Carpenter CR, Sarli CC, Fowler S a, et al. Best Evidence in Emergency Medicine (BEEM) rater scores correlate with publications' future citations. *Acad Emerg Med*. 2013;20(10):1004-1012. doi:10.1111/acem.12235.
21. Webb NM, Shavelson RJ. Generalizability Theory: Overview. In: *Wiley StatsRef: Statistics Reference Online*. John Wiley & Sons, Ltd; 2014. doi:10.1002/9781118445112.stat06729.
22. Downing SM. Validity: on the meaningful interpretation of assessment data. *Med Educ*. 2003;37(9):830-837. doi:10.1046/j.1365-2923.2003.01594.x.
23. Schuwirth LWT, van der Vleuten CPM. Programmatic assessment and Kane's validity perspective. *Med Educ*. 2012;46(1):38-48. doi:10.1111/j.1365-2923.2011.04098.x.
24. Linzer M. The journal club and medical education: over one hundred years of unrecorded history. *Postgrad Med J*. 1987;63:475-478. doi:10.1136/pgmj.63.740.475.
25. Thangasamy I a, Leveridge M, Davies BJ, Finelli A, Stork B, Woo HH. International Urology Journal Club via Twitter: 12-Month Experience. *Eur Urol*. 2014;66(1):112-117. doi:10.1016/j.eururo.2014.01.034.
26. Choo EK, Ranney ML, Chan TM, et al. Twitter as a tool for communication and knowledge exchange in academic medicine: A guide for skeptics and novices. *Med Teach*. 2015;37(5):411-416. doi:10.3109/0142159X.2014.993371.
27. Melvin L (University of T, Chan TM (McMaster U. Using Twitter in Clinical Education and Practice. *J Grad Med Educ*. 2014;6(3)(September):581-582. doi:10.4300/JGME-D-14-00342.1.
28. Chan TM, Rosenberg H, Lin M. Global emergency medicine journal club: social media responses to the January 2014 online emergency medicine journal club on subarachnoid hemorrhage. *Ann Emerg Med*. 2014;64(1):88-94. doi:10.1016/j.annemergmed.2014.05.015.
29. Thoma B, Rolston D, Lin M. Global Emergency Medicine Journal Club: Social media responses to the March 2014 Annals of Emergency Medicine Journal Club on targeted Temperature Management. *Ann Emerg Med*. 2014;64(2):207-212. doi:10.1016/j.annemergmed.2014.06.003.
30. Radecki RP, Rezaie SR, Lin M. Global emergency medicine Journal Club: Social media responses to the November 2013 Annals of Emergency Medicine Journal Club. *Ann Emerg Med*. 2014;63(4):490-494. doi:10.1016/j.annemergmed.2014.01.033.
31. Rezaie SR, Swaminathan A, Chan T, Shaikh S, Lin M. Global Emergency Medicine Journal Club: A Social Media Discussion About the Age-Adjusted D-Dimer Cutoff Levels to Rule Out Pulmonary Embolism Trial. *Ann Emerg Med*. 2015;65(5):604-613. doi:10.1016/j.annemergmed.2015.02.024.
32. Westafer L, Hensley J, Shaikh S, Lin M. Global Emergency Medicine Journal Club: A Social Media Discussion About the Lack of Association Between Press Ganey Scores and Emergency Department Analgesia. *Ann Emerg Med*. 2015;66(4):1-7. doi:10.1016/j.annemergmed.2015.04.012.
33. Chan TM, Thoma B, Radecki R, et al. Ten Steps for Setting Up an Online Journal Club. *J Contin Educ Health Prof*. 2015;35(2):148-154. doi:10.1002/chp.21275.
34. Chan T, Joshi N, Lin M, Mehta N. Using Google Hangouts on Air for Medical Education: A Disruptive Way to Leverage and Facilitate Remote Communication and Collaboration. *J Grad Med Educ*. 2015;7(2):171-173. doi:10.4300/JGME-D-14-00545.1.
35. Lin M, Thoma B, Trueger NS, Ankel F, Sherbino J, Chan T. Quality indicators for blogs and podcasts used in medical education: modified Delphi consensus recommendations by an international cohort of health professions educators. *Postgrad Med J*. 2015;91(1080):546-550. doi:10.1136/postgradmedj-2014-133230.
36. Thoma B, Chan TM, Paterson QS, Milne WK, Sanders JL, Lin M. Emergency Medicine and Critical Care Blogs and Podcasts: Establishing an International Consensus on Quality. *Ann Emerg Med*. 2015. doi:10.1016/j.annemergmed.2015.03.002.
37. Guyatt G, Meade MO, Rennie D, Cook DJ. *User's Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice, 3rd. Edition*. 3rd ed. McGraw-Hill Education; 2015.