ACADEMIC MEDICINE
Journal of the Association of American Medical Colleges

Author: Blanco Maria EdD; Prunuske Jacob MD, MSPH; DiCorcia Mark PhD, OTR; Learman Lee A. MD, PhD; Mutcheson Brock MEd, PhD; Huang Grace C. MD

Title: The DoCTRINE Guidelines: Defined Criteria To Report INnovations in Education

DOI: 10.1097/ACM.0000000000004634

Academic Medicine

DOI: 10.1097/ACM.0000000000004634

The DoCTRINE Guidelines: Defined Criteria To Report INnovations in Education

Maria Blanco, EdD, Jacob Prunuske, MD, MSPH, Mark DiCorcia, PhD, OTR, Lee A. Learman, MD, PhD, Brock Mutcheson, MEd, PhD, and Grace C. Huang, MD

M. Blanco is associate professor of psychiatry and associate dean for faculty development, Tufts University School of Medicine, Boston, Massachusetts; ORCID: http://orcid.org/0000-0002-7469-6050.

J. Prunuske is professor of family and community medicine and assistant dean for clinical learning, Medical College of Wisconsin–Central Wisconsin, Wausau, Wisconsin; ORCID: https://orcid.org/0000-0001-5638-5227.

M. DiCorcia is associate professor of integrated medical science and assistant dean for medical education, Charles E. Schmidt College of Medicine at Florida Atlantic University, Boca Raton, Florida; ORCID: https://orcid.org/0000-0002-4087-9668.

L. Learman is professor of obstetrics and gynecology and dean, Virginia Tech Carilion School of Medicine, Roanoke, Virginia; ORCID: https://orcid.org/0000-0001-8698-8918.

B. Mutcheson is assistant professor of basic science education and assistant dean of assessment and program evaluation, Virginia Tech Carilion School of Medicine, Roanoke, Virginia; ORCID: https://orcid.org/0000-0001-9952-9614.

G.C. Huang is dean for faculty affairs and associate professor of medicine, Harvard Medical School, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA; ORCID: http://orcid.org/0000-0003-2965-0341.

Correspondence should be addressed to Grace C. Huang, 25 Shattuck Street, Suite 206, Boston, MA 02115; telephone: 617-432-1106; email: ghuang@hms.harvard.edu; Twitter: @GraceHuangMD.

Supplemental digital content for this article is available at http://links.lww.com/ACADMED/B238.

Acknowledgments: The authors wish to thank the expert educators involved in the Delphi study and piloting of the DoCTRINE guidelines. They also wish to express gratitude to the educators involved in the interrater reliability testing: Dan Mayer, Jodi Abbott, Kathleen Kreutzer, Michael Cassara, Paul Edwards, and Richard Sabina.

Funding/Support: None reported.

Other disclosures: Four of the authors are editors for *MedEdPORTAL* (M.B., J.P., M.D., G.C.H.).

Ethical approval: Exempted by the Beth Israel Deaconess Medical Center Committee on Clinical Investigations.

Abstract

Purpose

Reporting guidelines assist authors in conducting and describing their research in alignment with evidence-based and expert-determined standards. However, published research-oriented guidelines do not capture all of the components that must be present in descriptions of educational innovations in health professions education. The authors aimed to create guidelines for educational innovations in curriculum development that would be easy for early-career educators to use, support reporting necessary details, and promote educational scholarship.

Method

Beginning in 2017, the authors systematically developed a reporting checklist for educational innovations in curriculum development, called Defined Criteria To Report INnovations in Education (DoCTRINE), and collected validity evidence for its use according to the 4 inferences of Kane's framework. They derived the items using a modified Delphi method, followed by pilot testing, cognitive interviewing, and interrater reliability testing. In May–November 2019, they implemented DoCTRINE for authors submitting to *MedEdPORTAL*, half of whom were randomized to receive the checklist (intervention group). The authors scored manuscripts using DoCTRINE while blinded to group assignment, and they collected data on final editorial decisions.

Results

The final DoCTRINE checklist consists of 19 items, categorized into 5 components: introduction, curriculum development, curriculum implementation, results, and discussion. The overall interrater agreement was 0.91. Among the 108 manuscripts submitted to *MedEdPORTAL* during the study period, the mean (SD) total score was higher for accepted than rejected submissions (16.9 [1.73] vs 15.7 [2.24], $P = .006$). There were no significant differences in

DoCTRINE scores between the intervention group, who received the checklist, and the control group, who did not.

Conclusions

The authors developed DoCTRINE using systematic approaches, for the scholarly reporting of educational innovations in curriculum development. This checklist may be a useful tool for supporting the publishing efforts of early-career faculty.

Reporting guidelines[1] have long been used by scholars to conduct research and prepare manuscripts that adhere to evidence-based and expert-determined standards. Without such guidelines, well-codified principles for carrying out specific study designs may be overlooked. Additionally, incomplete reporting can impede readers' abilities to replicate interventions and can stymie knowledge syntheses such as meta-analyses.

The need for reporting standards extends to disseminating "educational innovations,"[2] which we have defined for the purpose of our study as the implementation of activities considered novel due to the teaching method, setting, population of learners, or presentation of new content. Accordingly, these core elements of educational innovations—novelty and intervention—align well with principles of scholarship and empiricism. Further, the development, implementation, and evaluation of educational innovations collectively represent a form of scientific inquiry that is intended to advance understanding of the processes of teaching and learning while disseminating new practices—namely, educational scholarship.[2,3]

Existing research guidelines do not capture all of the components that must be present in the description of an educational intervention. Our search of the EQUATOR network,[4] a registry of reporting guidelines in the health sciences, identified a limited set of guidelines relevant to educational innovations in curriculum development across the health professions. The Standards for QUality Improvement Reporting Excellence in Education (SQUIRE-EDU) checklist[5] is an extension of the SQUIRE guidelines[6] to include educational improvement. Its focus is broad in scope given its origin in quality improvement, and it assumes the participation of interprofessional teams and the influence of the initiatives on a broad set of stakeholders, which may not be applicable to educational innovations, as they are narrow by nature. The Guideline for Reporting Evidence-based practice Educational interventions and Teaching (GREET)[7] was designed to improve the consistency and detail of reporting educational interventions for

evidence-based practice, dedicating 13 of its 17 items to important components of the intervention and its delivery (e.g., materials, incentives, environment). However, GREET does not address fundamental elements of scholarship, including articulation of the problem, outcomes of the intervention (beyond process measures like lessons learned or attendance), or interpretation of findings in relation to the literature.[8] Other health professions education checklists identified in our EQUATOR search relate to specific educational approaches, such as simulation,[9] team-based learning,[10] objective structured clinical examinations,[11] and standardized patients.[12]

Some guidelines, including those proposed in editorials,[2,13–15] emphasize particular aspects of writing and reporting about educational innovations in health professions education. Kanter[2] specifically focuses on the important constructs of generalizability and sustainability; however, the associated list of prompting questions that he proposes is expansive and may be challenging for novice scholars to translate into their writing. Other experts have summarized strategies for publishing educational innovations through the prisms of educational scholarship,[16] educational research,[17] or curriculum development.[18,19] All of these characterizations are highly relevant but they are not configured with sufficient specificity for authors conducting the work and drafting manuscripts. Finally, while some journals' descriptions of specific manuscript types (e.g., Innovation Reports in *Academic Medicine*[20]) serve as guidance for authors wishing to write about their educational innovations, journals' instructions to authors tend to focus on formatting, word limits, and definitions of scope, rather than providing granular lists of scholarly requirements. None of this literature relies on systematic approaches that demonstrate validity evidence.

Without clear guidelines and a checklist of expected elements for scholarly manuscripts that describe educational innovations, the process leading from curricular design to publication in health professions education journals is left more to chance than intention. Thus, we created reporting guidelines for educational innovations with a focus on curriculum development—Defined Criteria To Report INnovations in Education (DoCTRINE)—and collected validity evidence for DoCTRINE's use. A key principle was that our checklist should broadly meet the needs of health professions educators, who have widely variable research training and experience with scholarly writing. We aimed for a final product that would be easy for early-career educators to use, support the minimum reporting of details necessary for readers to replicate an educational innovation in curriculum development, and promote skills in educational scholarship, a competency that can lead to a stronger reputation, opportunities for collaboration and grant funding, and academic promotion.[21,22]

Method

Our study began in October 2017. We followed Moher et al's strategy for developing reporting guidelines[23] and Kane's framework for drafting a validity argument.[24] The steps we followed to develop the DoCTRINE guidelines are summarized in Supplemental Digital Appendix 1 at http://links.lww.com/ACADMED/B238.

Stage 1: Developing the guidelines

In developing (Stage 1) and piloting (Stage 2) the guidelines, we focused on accruing evidence related to the scoring inference of Kane's validity argument.[24] We performed a literature review for published tools that assessed the quality of descriptions of curriculum development, including implementation and evaluation. We used results of the search (see Supplemental Digital Appendix 2 at http://links.lww.com/ACADMED/B238) to inform the design of our guidelines. We drafted a preliminary list of items and modified it over 5 iterations.

We identified nationally recognized scholars in health professions education based on leadership and landmark contributions to the literature, and we invited them to participate in 2 rounds of a modified Delphi study[25] to ensure inclusion of all relevant concepts. All 14 experts whom we contacted agreed to participate in this study. In the first round, we asked them to rate the extent to which each item was necessary to assess completeness, reproducibility, and transparency in a report of an educational innovation in curriculum development. The experts rated items using a 5-point scale (never, rarely, sometimes, often, always). We invited their revisions and additions to the text of the items. We determined a priori that we would include items if the proportion of "often" and "always" responses exceeded 70%. In the second round, we asked participants to re-rate the items that we had revised based on their aggregated input.

Stage 2: Piloting the guidelines

Like items in other reporting guidelines, each DoCTRINE item is binary (i.e., scored as "present"= 1 or "absent"= 0). Each of us applied the checklist to 3 *MedEdPORTAL* publications describing educational innovations in curriculum development in medical education. We discussed our experiences, reexamined other reporting checklists, and performed a collaborative round of revisions. These revisions included formatting the checklist to align more closely with other reporting guidelines and simplifying the language of the item statements. Two of the authors (G.C.H., M.B.) then conducted cognitive pretesting[26] with 4 medical educators to explore their interpretation of the checklist items, and we revised the instrument based on this feedback.

Stage 3: Testing the guidelines

We sought to assess the generalization and extrapolation inferences of Kane's framework[24] by assessing reliability and exploring the performance of the checklist in practice. We asked 6 medical educators to apply the checklist to 6 *MedEdPORTAL* publications. Although *MedEdPORTAL* is unique in that it peer reviews and publishes health professions educational

resources through appendices, the manuscript accompanying the appendices features an educational innovation description, similar to descriptions of innovations published in other medical and health professions education journals. The 6 publications were selected to reflect the range of curricular innovations typically submitted to *MedEdPORTAL*.

We assessed estimates of interrater reliability by calculating average agreement. We elected this approach, rather than the kappa statistic, due to concerns about the "kappa paradox" for instruments with low variability.[27,28] Our use of 6 raters each scoring the 6 publications provided adequate statistical power (80%) to detect interrater reliability of 0.50 or greater, assuming a null hypothesis of nonagreement.[29]

Stage 4: Implementing the guidelines

To collect evidence supporting Kane's inference[24] of interpretation to real-world performance and implications for decision-making, we assessed whether completeness on the checklist was higher for accepted *MedEdPORTAL* submissions than for rejected submissions. We also assessed the impact of including a copy of DoCTRINE with *MedEdPORTAL*'s initial submission screening letters on the completeness of resubmitted manuscripts. We intervened at the point in the journal's routine process when authors typically receive screening letters after initial submission. The screening letter includes an individualized list of elements for the author to address in the submission, which they then resubmit for consideration for peer review, to ensure adherence to *MedEdPORTAL*'s submission standards. Elements requested of authors at this screening stage are procedural rather than substantive (e.g., ensuring that appendices are referenced within the manuscript, removing copyrighted materials).

In alternating months from May 2019 to November 2019, the *MedEdPORTAL* editorial staff included DoCTRINE with the screening letters along with instructions that asked the authors to consider using the checklist as part of their reformatting process (intervention group). No new

processes were implemented during this time that might have led to systematic bias in group assignment. Authors in the other months received the screening letter only (control group). Participation was voluntary; the letter's instructions were explicit that the checklist was for research purposes only and would not influence editorial decisions.

Each manuscript was independently scored by 1 of the 6 authors. We were blinded to group assignments (whether the submission had received DoCTRINE or not, whether the manuscript was at the prescreening stage or the postscreening stage [resubmitted in response to the screening letter]).

Analytic approach

We collected data on the editorial decisions for the scored manuscripts (namely, whether they were ultimately rejected or accepted for publication in *MedEdPORTAL*). We assessed overall completeness of the submissions by summing scores across the final 19 DoCTRINE items to generate interval-level data by creating a simple, equally weighted compensatory composite score for each submission at both the prescreening and postscreening stages.[30] Simply put, the DoCTRINE total score for a given manuscript was represented by the number of items present, for a maximum of 19 points. We used descriptive statistics, including frequencies and percentages of present and absent items, to summarize the data. Additionally, we examined the total mean DoCTRINE score distribution in consideration of the sample sizes. Because the independent sample *t*-test is asymptotically robust to the normality assumption under general conditions,[31] we used it to compare the mean postscreening scores for accepted and rejected submissions.

Next, we compared the prescreening and postscreening mean scores using paired 2-tailed *t*-tests. We also used *t*-tests to determine whether the inclusion of the DoCTRINE with the initial submission screening letter was associated with greater pre- to postscreening improvement than

the screening letter alone, and we correlated pre- to postscreening DoCTRINE score changes between the intervention and control groups. Because the Bonferroni method overcorrects for Type I error, we applied it only to the exploratory post-hoc analyses comparing accepted/rejected and intervention/control scores on the 19 individual DoCTRINE items.[32]

The study was determined to be exempt by the Beth Israel Deaconess Medical Center Institutional Review Board.

Results

Stages 1, 2, and 3: Developing, piloting, and testing the guidelines

The initial checklist we developed in stage 1 included 24 items. We organized these items into 5 components—introduction, curriculum development, curriculum implementation, results, and discussion—informed by the traditional manuscript structure and a proposal for reporting innovations.[18] After the 2 modified Delphi rounds, we kept the items that achieved our predetermined 70% threshold and made nominal edits to item wording. The modified Delphi process resulted in an average of 88% consensus on items in round 1 and 86% in round 2. By the end of stage 1, the checklist consisted of 20 items. We concluded that a round 3 would not provide additional information based on the degree of consensus in round 2.

Based on cognitive pretesting in stage 2 (piloting), we reworded several items for clarity. In Stage 3 (testing), we calculated overall interrater agreement to be 0.91, with item-level agreement ranging from 0.64 to 1.0 (Table 1). As a result of low agreement scores on 1 item, we changed "source of data collection instrument" to "origin of data collection instrument(s)." Also, because 2 items were perceived as similar, we removed 1 item, for a final total of 19 DoCTRINE items. We also changed 2 words to conform to commonly accepted pedagogical terms. The revisions to the checklist through each of these stages are summarized in Supplemental Digital Appendix 3 at http://links.lww.com/ACADMED/B238.

Stage 4: Implementing the guidelines

During the study period, 108 manuscripts were submitted to *MedEdPORTAL* (intervention group, n = 53; control group, n = 55). The total score distribution for all manuscripts had a mean (SD) of 16.4 (2.01) of 19 DoCTRINE items completed. Table 2 reports the frequencies and percentages of present and absent elements among the submitted manuscripts according to the final DoCTRINE checklist of 19 items, divided into 5 components.

The mean (SD) total score at the postscreening stage was higher for the 69 accepted submissions than for the 39 rejected submissions (16.9 [1.73] vs 15.7 [2.24], $P = .006$). Although the difference between these means did not appear to be large, we used Cohen's d to estimate an effect size of 0.615. This moderate effect size was influenced by smaller variances within both the accepted and the rejected submission distributions, and it highlighted the importance each DoCTRINE item played in the calculation of the total scores. Given the observed difference in total scores, we sought to identify trends in DoCTRINE components that were absent in the rejected submissions. Taken as a whole, the mean component scores for the results ($P = .02$) and discussion ($P = .002$) were significantly higher for accepted submissions than rejected submissions (Table 3).

We applied the Bonferroni correction to an exploratory post-hoc analysis comparing accepted and rejected submission total scores on the 19 individual DoCTRINE items. There were no statistically significantly higher item mean scores for accepted submissions compared with rejected submissions. Because each individual item was scored dichotomously (present or absent), the range and variance of scores on any single item were slightly restricted compared with the DoCTRINE total and component scores.

In sensitivity analyses comparing differences in total scores at the prescreening and postscreening stages, the means were equal (P = .49), suggesting no overall change in completeness from pre- to postscreening. Moreover, there were no significant differences between the DoCTRINE total or component scores associated with providing the checklist with the screening letter, nor any differences in pre- to postscreening changes in completeness between the intervention and control groups (see Table 4 for the latter finding).

Discussion

To address the need for reporting guidelines specific to descriptions of educational innovations in curriculum development in the health professions, we developed the DoCTRINE guidelines through a systematic iterative process that complied with guidelines for developing reporting checklists.[23] This involved a modified Delphi study, pilot testing, cognitive interviewing, interrater reliability assessment, and implementation in a real-world setting. Through this process, we were able to collect validity evidence supporting all 4 inferences in Kane's framework.[24]

We found high levels of interrater agreement at the item level and overall, demonstrating strong reliability of the DoCTRINE guidelines. We believe that our iterative process of soliciting input from both experts and users resulted in an instrument that would be relatively straightforward for other scholars to apply. Although experts may favor a more comprehensive list of elements to be included in curricular innovation reports,[33] we developed our checklist of clearly defined minimum elements to promote usability by early-career authors who may not be familiar with advanced concepts. This strategy appears to have been successful: Our interrater agreement may indicate ease of interpretation by future authors.

We suspected that submissions fulfilling more of the DoCTRINE items, thus providing sufficient detail, would have a greater likelihood of acceptance than submissions fulfilling fewer of the items. Indeed, we found that the mean scores of accepted submissions were significantly higher than rejected ones, suggesting 2 possibilities. First, completeness in reporting may have reflected an understanding of scholarly writing that became advantageous in the editorial decision-making process. Second, completeness may have been a marker of the quality of curricular design, which in itself may have portended success in the peer-review process. This source of validity evidence (i.e., differences in checklist scores between accepted and rejected submissions) was promising in that it reinforced the notion that authors benefit from using checklists that are aligned with sound educational practices and also convey the information that journals expect in descriptions of innovative curricula. Further analysis demonstrated that the results and discussion component mean scores were significantly higher for accepted submissions—which aligns with our experience that these sections tend to be more difficult for early-career scholars—whereas mean scores for the more-formulaic introduction and methods components were not. Also, this finding may reflect the importance of an evaluation component and linkage of findings to the literature as hallmarks of a scholarly approach that others can build upon. These components are not represented in the GREET checklist.[7]

Unfortunately, our findings do not suggest that giving DoCTRINE to authors as a resource to improve their manuscripts prior to resubmission improved the completeness of their reporting. In the intervention group, the checklist was intended as a general resource for authors to complete during their reformatting process and did not include any specific, actionable feedback. Authors may have ignored the checklist or skimmed it perfunctorily, since completing it was not required for resubmission. Future efforts to demonstrate DoCTRINE's value to authors could include highlighting the items that a given submission is missing or mentioning that checklist items

covering aspects of the results and discussion are the items most commonly missing from rejected submissions.

Comparing DoCTRINE with other reporting guidelines, there are similarities to Meinema et al's checklist for descriptions of curricular interventions.[34] These scholars modified the GREET checklist[7] to examine whether publications describing classroom teaching for postgraduate trainees met these criteria and found many lacking in all of the GREET elements. In this respect, they built on another reporting checklist and showed that many publications fall short. However, they did not create a new checklist based on a systematic approach. Our work is more comparable in focus to that of 2 other groups who created checklists for educational innovation reports. Hall et al[33] used a literature-based approach, and Van Hecke et al[35] based their checklist on expert consensus. We went a step further by involving multiple rounds of use by authors and journal editors. Additionally, there are many checklists and scoring schemas used to rate the quality of medical education research, which are best summarized by Hall et al.[33] However, their direct applicability to educational innovations is limited, as evaluation tends to be less robust in innovation descriptions. Also, they may not assist early-career faculty hoping to publish nonresearch work.

Limitations of this study included the focus on *MedEdPORTAL* submissions. All of this study's authors are familiar with *MedEdPORTAL*, and the checklist was tested on *MedEdPORTAL* submissions, which can be up to 4,000 words in length. That said, we believe the checklist is generalizable beyond *MedEdPORTAL*. This manuscript format is analogous to that of full-length reports on educational innovations in other health professions education journals. However, we have no evidence to support DoCTRINE's use for short-form submissions like research abstracts, which do not typically use checklists, or brief reports (e.g., less than 1,000 words), which may be too constrained by length requirements to accommodate all checklist items.

MedEdPORTAL's instructions to authors[36] reflect scholarly writing principles and thus may have attenuated the full impact of providing DoCTRINE to authors. DoCTRINE was explicitly not intended to measure quality, but with respect to our finding that accepted submissions had higher DoCTRINE overall completeness scores than rejected submissions, completeness may be associated with other aspects of quality that impact editorial decisions.

DoCTRINE has many potential applications. Innovators in health professions education seeking to disseminate their work should find the checklist helpful in supporting a scholarly approach that is informed by theory and research and contributes to the literature for other innovators to replicate, adapt, and extend. Future research should explore the transferability of DoCTRINE to the full range of educational innovations beyond curricular development, and to other journals publishing educational innovations. Journals could potentially incorporate DoCTRINE into their author instructions. Furthermore, the DoCTRINE checklist may have utility as a guide for peer reviewers and for editors to provide focused feedback to authors. Mentors and educators could use DoCTRINE to coach early-career faculty regarding the key components for designing, implementing, evaluating, and reporting their curricular innovations.

References

1. Altman DG, Simera I. Using reporting guidelines effectively to ensure good reporting of health research. In: Moher D, Altman DG, Schulz K, Simera I, Wager E, eds. Guidelines for Reporting Health Research: A User's Manual. 1st ed. Wiley and Sons, Ltd.; 2014:32-40.
2. Kanter SL. Toward better descriptions of innovations. Acad Med. 2008;83(8):703-704.
3. Crites GE, Gaines JK, Cottrell S, et al. Medical education scholarship: An introductory guide: AMEE Guide No. 89. Med Teach. 2014;36(8):657-674.
4. Enhancing the QUAlity and Transparency Of health Research. EQUATOR Network. https://www.equator-network.org/. Accessed September 24, 2021.
5. Ogrinc G, Armstrong GE, Dolansky MA, Singh MK, Davies L. SQUIRE-EDU (Standards for QUality Improvement Reporting Excellence in Education): Publication guidelines for educational improvement. Acad Med. 2019;94(10):1461-1470.
6. Ogrinc G, Mooney SE, Estrada C, et al. The SQUIRE (Standards for QUality Improvement Reporting Excellence) guidelines for quality improvement reporting: Explanation and elaboration. In: Qual Saf Health Care. 2008;17 (Suppl 1):i13-i32.
7. Phillips AC, Lewis LK, McEvoy MP, et al. Development and validation of the guideline for reporting evidence-based practice educational interventions and teaching (GREET). BMC Med Educ. 2016;16(1):237.
8. Glassick CE, Huber MR, Maeroff GI. Scholarship Assessed: Evaluation of the Professoriate. San Francisco, CA: Jossey-Bass; 1997.
9. Cheng A, Kessler D, Mackinnon R, et al. Reporting guidelines for health care simulation research. Simul Healthc. 2016;11(4):238-248.

10. Haidet P, Levine RE, Parmelee DX, et al. Perspective: Guidelines for reporting team-based learning activities in the medical and health sciences education literature. Acad Med. 2012;87(3):292-299.

11. Patrício M, Julião M, Fareleira F, Young M, Norman G, Vaz Carneiro A. A comprehensive checklist for reporting the use of OSCEs. Med Teach. 2009;31(2):112-124.

12. Howley L, Szauter K, Perkowski L, Clifton M, McNaughton N. Quality of standardised patient research reports in the medical education literature: Review and recommendations. Med Educ. 2008;42(4):350-358.

13. Sklar DP. Sharing new ideas and giving them wings: Introducing Innovation Reports. Acad Med. 2013;88(10):1401-1402.

14. Cook DA, Reed DA, Wayne DB, West CP. From the editors' desk: Renewing the call for innovations in medical education. J Gen Intern Med. 2010;25(9):887-888.

15. Sklar DP, Weinstein DF, Carline JD, Durning SJ. Developing programs that will change health professions education and practice: Principles of program evaluation scholarship. Acad Med. 2017;92(11):1503-1505.

16. Blanchard RD, Nagler A, Artino AR. Harvest the low-hanging fruit: Strategies for submitting educational innovations for publication. J Grad Med Educ. 2015;7(3):318-322.

17. O'Brien BC, West CP, Coverdale JH, Durning SJ, Roberts LW. On the use and value of reporting guidelines in health professions education research. Acad Med. 2020;95(11):1619-1622.

18. Reznich CB, Anderson WA. A suggested outline for writing curriculum development journal articles: The IDCRD Format. Teach Learn Med. 2009;13(1):4-8.

19. Thomas P, Kern DE, Hughes MT, Chen B. Curriculum Development for Medical Education: A Six-Step Approach. 3rd ed. Johns Hopkins University Press; 2016.

20. Durning SJ, O'Brien BC, West CP, Coverdale J, DeVilbiss MB, Roberts LW. Innovation Reports: Guidance from the editors. Acad Med. 2020;95(11):1623-1625.

21. Klingensmith ME, Anderson KD. Educational scholarship as a route to academic promotion: A depiction of surgical education scholars. Am J Surg. 2006;191(4):533-537.

22. Beck JB, DeVilbiss MB, Carline JD, McDaniel CE, Durning SJ. Innovation Reports: Successes and limitations for promoting innovation in medical education. Acad Med. 2020;95(11):1647-1651.

23. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. PLoS Med. 2010;7(2):e1000217.

24. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: A practical guide to Kane's framework. Med Educ. 2015;49(6):560-575.

25. Waggoner J, Carline JD, Durning SJ. Is there a consensus on consensus methodology? Descriptions and recommendations for future consensus research. Acad Med. 2016;91(5):663-668.

26. Willis G. Pretesting of health survey questionnaires: Cognitive interviewing, usability testing, and behavior coding. In: Johnson T, ed. Handbook of Health Survey Methods. John Wiley & Sons, Inc.; 2014: 217-242.

27. Zec S, Soriani N, Comoretto R, Baldi I. High agreement and high prevalence: The paradox of Cohen's Kappa. Open Nurs J. 2017;11(1):211-218.

28. Feinstein AR, Cicchetti D V. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol. 1990;43(6):543-549.

29. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Stat Med. 1998;17(1):101-110.

30. Organisation for Economic Co-operation and Development. Handbook on Constructing Composite Indicators: Methodology and User Guide. https://www.oecd.org/sdd/42495745.pdf. Published 2008. Accessed January 7, 2022.

31. Fay MP, Proschan MA. Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. Stat Surv. 2010;4:1-39.

32. Streiner DL, Norman GR. Correction for multiple testing: Is there a resolution? Chest. 2011;140(1):16-18.

33. Hall AK, Hagel C, Chan TM, Thoma B, Murnaghan A, Bhanji F. The writer's guide to education scholarship in emergency medicine: Education innovations (part 3). Can J Emerg Med. 2018;20(3):463-470.

34. Meinema JG, Buwalda N, van Etten-Jamaludin FS, Visser MRM, van Dijk N. Intervention descriptions in medical education: What can be improved? A systematic review and checklist. Acad Med. 2019;94(2):281-290.

35. Van Hecke A, Duprez V, Pype P, Beeckman D, Verhaeghe S. Criteria for describing and evaluating training interventions in healthcare professions – CRe-DEPTH. Nurse Educ Today. 2020;84:104254.

36. MedEdPORTAL. Author Center: Preparing and submitting. https://www.mededportal.org/author. Accessed February 2, 2022.

Table 1
Item-Level and Overall Interrater Reliability for the DoCTRINE Guidelines at the Testing Stage

Item	Interrater agreement (range)[a]
1. Need for the curriculum	0.98 (0.86-1.00)
2. Review of relevant literature, theories, models, and published curricula	0.95 (0.86-1.00)
3. Unique contribution of the curriculum to the literature	0.90 (0.71-1.00)
4. Purpose/goals of the curriculum	0.98 (0.86-1.00)
5. Outcome-based learning objectives	0.93 (0.71-1.00)
6. Target population of learners	0.98 (0.86-1.00)
7. Intended type of participants[b]	0.86 (0.71-1.00)
8. Educational setting for curriculum delivery[c]	0.90 (0.71-1.00)
9. Resources for implementing the curriculum	0.95 (0.86-1.00)
10. Description of instructional methods	0.95 (0.86-1.00)
11. Methods to evaluate achievement of outcome-based learning objectives	0.79 (0.57-0.86)
12. Origin of data collection instrument[c]	0.64 (0.57-0.86)
13. Number of learners participating in the curriculum	0.98 (0.86-1.00)
14. Number of participants included in the evaluation	1.00 (1.00-1.00)
15. Evidence of achievement of outcome-based learning objectives	0.83 (0.57-1.00)
16. Summary of findings	0.98 (0.86-1.00)
17. Interpretation of findings in relation to the existing literature	0.71 (0.57-0.86)
18. Lessons learned from the implementation of the curriculum	0.98 (0.86-1.00)
19. Limitations of the evaluation of the curriculum	0.93 (0.86-1.00)
20. Description of future implications of the curriculum	0.98 (0.86-1.00)
Overall	0.91 (0.57-1.00)

Abbreviation: DoCTRINE, Defined Criteria To Report INnovations in Education.
[a]Across 6 reviewers who reviewed 6 *MedEdPORTAL* publications selected to reflect the types of curricular innovations typically submitted to the journal.

[b]Removed from the final version of DoCTRINE; perceived to be the same as item 6.
[c]Wording changed in the final version of DoCTRINE to conform to commonly accepted pedagogical terms (see Table 2).

Table 2

DoCTRINE Elements Present and Absent in 108 Submissions to *MedEdPORTAL*, May-November 2019 Study Period[a]

Items by component	Present, no. (%)	Absent, no. (%)
Component 1: Introduction		
1: Need for curriculum	108 (100)	0 (0)
2: Review of relevant literature, theories, models, and published curricula	107 (99)	1 (1)
3: Unique contribution of the curriculum to the literature	93 (86)	15 (14)
Component 2: Curriculum Development		
4: Purpose/goals of the curriculum	106 (98)	2 (2)
5: Outcome-based learning objectives	107 (99)	1 (1)
6: Target population of learners	107 (99)	1 (1)
Component 3: Curriculum Implementation		
7: Instructional setting for curriculum delivery	103 (95)	5 (5)
8: Resources for implementing the curriculum	106 (98)	2 (5)
9: Description of instructional methods	104 (96)	4 (4)
10: Methods to evaluate achievement of outcome-based learning objectives	105 (97)	3 (3)
11: Origin of evaluation instrument(s)	51 (47)	57 (53)

Component 4: Results		
12: Number of learners participating in the curriculum	95 (88)	13 (12)
13: Number of participants included in the evaluation	80 (74)	28 (26)
14: Evidence of achievement of outcome-based learning objectives	90 (83)	18 (17)
Component 5: Discussion		
15: Summary of findings	96 (89)	12 (11)
16: Interpretation of findings in relation to the existing literature	49 (45)	59 (55)
17: Lessons learned from the implementation of the curriculum	97 (90)	11 (10)
18: Limitations of the evaluation of the curriculum	73 (68)	35 (32)
19: Description of future implications of the curriculum	97 (90)	11 (10)

Abbreviation: DoCTRINE, Defined Criteria To Report INnovations in Education.
[a]Each of the final 19 DoCTRINE items was scored dichotomously, using 1 = "present" or 0 = "absent."

Table 3

DoCTRINE Component Mean Scores: Comparisons Between Accepted and Rejected Submissions to *MedEdPORTAL*[a]

	Accepted (n = 69)		Rejected (n = 39)		95% CI		
Component[b]	Mean	(SD)	Mean	(SD)	Lower	Upper	*P* value
1. Introduction	2.9	(0.32)	2.8	(0.41)	-0.23	0.05	0.21
2. Curriculum Development	3.0	(0.24)	2.9	(0.22)	-0.12	0.07	0.64
3. Curriculum Implementation	4.4	(0.63)	4.2	(0.81)	-0.45	0.10	0.21
4. Results	2.6	(0.72)	2.2	(0.81)	-0.65	-0.51	0.02
5. Discussion	4.0	(0.76)	3.5	(1.07)	-0.90	-0.20	0.002

Abbreviation: DoCTRINE, Defined Criteria To Report INnovations in Education.

[a]The 108 submissions included in this study were submitted to *MedEdPORTAL* in May–November 2019.

[b]For the final version of DoCTRINE, with 19 items divided into 5 components, see Table 2.

Table 4

Differences in Completed DoCTRINE Items Between the Prescreening and Postscreening Stages in the Control Group (n = 53) Versus Intervention Group (n = 55)[a]

Component[b]	Intervention mean delta	Control mean delta	*P* value
1. Introduction (3 items)	0.02	0.00	.56
2. Curriculum Development (3 items)	-0.02	0.00	.56
3. Curriculum Implementation (5 items)	-0.04	0.05	.17
4. Results (3 items)	-0.06	0.00	.54
5. Discussion (5 items)	-0.08	-0.05	.87
Total (19 items)	-0.17	0.00	.48

Abbreviation: DoCTRINE, Defined Criteria To Report INnovations in Education.

[a]The intervention group received DoCTRINE with the screening letters editorial staff members sent to authors following initial submission. The instructions suggested authors consider using the checklist as part of their reformatting process, prior to resubmitting their manuscript for consideration for peer review, but were explicit that participation was voluntary. The control group received the screening letter only. Each manuscript was independently scored at the prescreening and postscreening stages by 1 of the 6 study authors, who were blinded to group assignments and stage of the submission.

[b]For the final version of DoCTRINE, with 19 items divided into 5 components, see Table 2.