

# Proposed Standards for Medical Education Submissions to the Journal of General Internal Medicine

David A. Cook, MD, MHPE<sup>1</sup>, Judith L. Bowen, MD<sup>2</sup>, Martha S. Gerrity, MD, PhD<sup>3</sup>, Adina L. Kalet, MD, MPH<sup>4</sup>, Jennifer R. Kogan, MD<sup>5</sup>, Anderson Spickard, MD, MS<sup>6</sup>, and Diane B. Wayne, MD<sup>7</sup>

<sup>1</sup>Office of Education Research and Division of General Internal Medicine, Mayo Clinic College of Medicine, Rochester, MN, USA; <sup>2</sup>Division of General Internal Medicine & Geriatrics, Department of Medicine, Oregon Health and Science University, Portland, OR, USA; <sup>3</sup>Portland VA Medical Center and Division of General Internal Medicine and Geriatrics, Department of Medicine, Oregon Health and Science University, Portland, OR, USA; <sup>4</sup>Division of Educational Informatics and Division of General Internal Medicine, Department of Medicine, NYU School of Medicine, New York, NY, USA; <sup>5</sup>Division of General Internal Medicine, University of Pennsylvania Health System, Philadelphia, PA, USA; <sup>6</sup>Department of Medicine and Biomedical Informatics, Vanderbilt School of Medicine, Nashville, TN, USA; <sup>7</sup>Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA.

To help authors design rigorous studies and prepare clear and informative manuscripts, improve the transparency of editorial decisions, and raise the bar on educational scholarship, the Deputy Editors of the Journal of General Internal Medicine articulate standards for medical education submissions to the Journal. General standards include: (1) quality questions, (2) quality methods to match the questions, (3) insightful interpretation of findings, (4) transparent, unbiased reporting, and (5) attention to human subjects' protection and ethical research conduct. Additional standards for specific study types are described. We hope these proposed standards will generate discussion that will foster their continued evolution.

**KEY WORDS:** medical education; scholarship; research design; research methods; writing.

J Gen Intern Med 23(7):908-13

DOI: 10.1007/s11606-008-0676-z

© Society of General Internal Medicine 2008

As part of its mission to serve the needs of generalist physicians, the *Journal of General Internal Medicine (JGIM)* publishes a substantial number of medical education articles. In 2007, *JGIM* published 58 articles related to medical education, and the 2008 medical education special issue alone contains over 40 articles and editorials.

Journal editors are responsible for selecting manuscripts most relevant to their readership and of the highest quality. *JGIM* has embraced efforts to evaluate and improve the quality of its medical education publications.<sup>1</sup> In this issue, Reed et al.<sup>2</sup> report a study evaluating the quality of all submissions to this medical education special issue. Their study noted a diversity of methodological quality among the submissions, yet also found that the highest quality manuscripts were ultimately selected for publi-

cation. This study has generated much discussion among the *Journal* editors, and we anticipate that it will likewise stimulate dialogue in the general community.

Moving forward, we wish to articulate the standards by which medical education submissions to *JGIM* are currently judged. The guidelines that follow reflect other published guidelines,<sup>3-6</sup> with an emphasis on the types of studies commonly submitted to *JGIM* and the issues they raise. We developed and refined these guidelines as we reviewed submissions for this special issue and made editorial decisions. By articulating these guidelines we hope to: (1) help scholars design rigorous studies and prepare clear and informative manuscripts, (2) assist manuscript reviewers in providing high quality critiques, (3) improve the transparency of editorial decisions, and (4) raise the bar on educational scholarship published in *JGIM*. A concise summary of these guidelines is available on the *JGIM* website and in the Appendix. We recognize these guidelines represent our views, but hope they will stimulate discussion among the broader community of education researchers and journal editors.

## WHAT CONSTITUTES QUALITY?

*JGIM* embraces the broad concept of scholarship outlined by Boyer<sup>7</sup> and will consider manuscripts demonstrating high-quality scholarship of any type. We also endorse the six standards described by Glassick<sup>8</sup> for assessing the quality of scholarship: clear goals, adequate preparation, appropriate methods, significant results, effective presentation, and reflective critique. Quality is multifaceted and includes not only rigorous research methods,<sup>9</sup> but also starting with an important question or goal,<sup>10-13</sup> interpreting results objectively and making valid inferences,<sup>4,14,15</sup> and reporting all of these clearly.<sup>4,16</sup> In the discussion that follows, we review these principles concisely and use them to frame recommendations for medical education manuscripts submitted to *JGIM*.

## Quality Questions

The research question is arguably the most important part of any scholarly activity.<sup>10,17,18</sup> The research question can be framed in many ways (purpose, objective, goal, aim, or hypothesis), but

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s11606-008-0676-z) contains supplementary material, which is available to authorized users.

should illustrate the relationship between the variables being studied (population, independent, and dependent variables). A focused question dictates appropriate methods and frames the interpretation of results.

*JGIM* emphasizes research relevant to the needs of generalist physicians, which includes both applied and theoretical education research. Education researchers can ask questions ranging<sup>19</sup> from concrete and practical (“How can we effectively teach students to perform medication reconciliation?”) to more abstract (“Why do faculty feel certain student behaviors are appropriate, while others are not?”).

Asking good questions requires a firm grasp of what has been previously done.<sup>13,17,20</sup> Adequate preparation is a hallmark of scholarship,<sup>8</sup> and is typified by a thorough and critical literature review that culminates in a “problem statement”<sup>11</sup> highlighting the gap in understanding that the present study seeks to fill. Yet medical education research studies frequently lack a critical literature review.<sup>16</sup> Without demonstration of adequate preparation, it is impossible to judge how a scholarly effort advances the field. Authors should present a concise but thorough examination of relevant literature, including strengths and weaknesses of previous studies.

Some questions are more important than others based on their implications for practice and research. Many questions are important by virtue of their relevance to pressing issues such as work hours or health disparities. However, importance is best supported through the use of a conceptual framework.<sup>11,13,21</sup> A conceptual framework “situates the research question, intervention methods, or study design within a model or theoretical framework that facilitates meaningful interpretation of the methods and results”<sup>16</sup> and subsequent application to new settings and future research. While frameworks may take the form of formal theories,<sup>22</sup> more often they are models for how things work (Glassick’s criteria<sup>8</sup> provide a framework for thinking about quality of scholarship) or systematic approaches to a problem (for example, an approach to the study of computer-based learning<sup>23</sup>). Questions that incorporate conceptual frameworks, and seek to clarify educational processes,<sup>24</sup> will be most relevant to other educators and researchers. Unfortunately, conceptual frameworks are frequently absent from reported education research.<sup>16</sup>

## Methods to Match the Question

Authors should select methods appropriate for the question. Guidelines relevant to specific study types are outlined below and in other sources.<sup>25–30</sup> In general, authors should explain critical methodological decisions, particularly when decisions lead to unusual or suboptimal methods. Justification can be logistical (“it was not feasible to randomize”), logical (“after careful consideration of various options, we decided to ... because ...”), or supported by literature.

Outcomes studied should also match the study goals. Studies that aim to improve knowledge should measure knowledge, and studies designed to improve skills should measure skills. Unfortunately, we often see this principle violated. Given concerns about the accuracy of learners’ subjective (i.e., self-assessed) ratings of knowledge or skills,<sup>31</sup> objective assessments are preferred over subjective measures. Although “higher-order” outcomes (behaviors in practice or patient outcomes<sup>32</sup>) are desirable,<sup>33,34</sup> they are not currently the standard<sup>35</sup> and may inadvertently weaken a study if measurements are questionable or outcomes do not align with objectives.

Authors should use appropriate statistical tests. Among the errors we commonly see are failure to adjust for multiple independent comparisons,<sup>36</sup> use of statistical tests of inference without verifying underlying assumptions,<sup>37</sup> and ambiguity about the statistical tests used. Investigators should consider collaboration and/or consultation with a statistician beginning in the planning stages (when the study design can still be adjusted and strengthened).

## Insightful Interpretation

Glassick’s “significant results” refer not to statistical significance, but rather to the impact of results on the field—in this case, the needs of generalist educators and researchers. A good question aligned with rigorous methods will facilitate relevance and defensible interpretations, but meaningful inferences also require objective analysis and reflective critique.<sup>38</sup> In addition to reporting percentages and statistical test results, a reflective scholar will identify strengths and shortcomings, situate the work in the context of prior studies, and identify immediate applications (often few) and directions for future research.<sup>4</sup>

No study is perfect, and even modestly flawed studies can support valid inferences. Authors should carefully consider how best to convey the study findings and integrate these with prior work without overstating the scope or importance of a study (over-generalization) or understating either the limitations or the implications of their work. Finding this delicate balance constitutes the art of reflective critique.<sup>8</sup>

## Transparent Reporting

Even the best study will fail to have an impact without effective communication of findings. Yet we know that medical education research reporting has much room for improvement.<sup>16,39,40</sup> Authors should consult appropriate guidelines,<sup>3,4,37</sup> follow *JGIM*’s “Instructions to Authors,” and obtain assistance if needed to clearly communicate their results. In addition to using accepted or prescribed organizing headings, authors must use clear, concise language and avoid jargon (i.e., locally developed or specialty-specific terminology).

In quantitative research, authors should report means and standard deviations for continuous variables, numerator and denominator (not just percentages) for categorical variables, and in all cases emphasize confidence intervals and effect sizes rather than p values alone.<sup>37,41,42</sup> Qualitative research should report specific themes along with supporting quotations and excerpts. Abstracts should be as “informative” as possible.<sup>40,43</sup>

## ETHICAL ISSUES IN EDUCATION RESEARCH

Education studies pose risks to human subjects,<sup>44</sup> yet many reported studies fail to comment on human subjects’ protections.<sup>16</sup> The power differential between the teacher and trainee, similar to that between the physician and patient, characterizes the trainee as a “vulnerable subject.” Furthermore, educational outcomes, both measured (e.g., grades) and unmeasured (e.g., acquired knowledge), are important to learners and can have lasting effects. Institutional review board (IRB) and informed consent requirements for education research vary across institutions and study designs. Investigators should obtain IRB review (with approval, exemption, or waiver as appropriate)

and then follow *JGIM*'s "Instructions to Authors" to "include a statement about informed consent and institutional review board approval in the methods section." Regardless of local requirements, investigators should diligently protect human subject rights.

Scholars should also adhere to standards of scientific integrity. Authors must be able to take public responsibility for the full content of an article to justify authorship, meaning they have not only read it, but have contributed meaningfully to the ideas presented. *JGIM* strongly opposes ghostwriting,<sup>45</sup> honorary authorship, undisclosed conflicts of interest,<sup>46</sup> duplicate publication, and plagiarism. Reporting a study's findings as a series of separate articles in order to maximize the number of publications is inappropriate.<sup>47</sup>

## STANDARDS FOR SELECTED STUDY TYPES

Below we highlight key or often-neglected quality elements for specific study "types." These types comprise a mixture of study purposes, designs, methods, and manuscript categories frequently found among *JGIM* medical education submissions, and are neither mutually exclusive nor all inclusive. Many studies will use a combination of these types.

This is not a comprehensive list of standards, even for a given type. Absence from discussion below does not mean a quality element is unimportant, but might simply mean we perceive it as less frequently problematic. Authors should continue to refer to relevant sources to guide the systematic design, conduct, and reporting of research.<sup>26-29,48-50</sup> Likewise, the absence of a specific study type does not indicate that we do not value such scholarship. We do not address systematic reviews here, but would welcome rigorous reviews on important education questions<sup>51</sup> and refer authors to published guidelines.<sup>52-54</sup> Similarly, we accept and encourage theory-building and programmatic research.<sup>21,22,24</sup>

## Educational Innovations

*JGIM*'s "Educational Innovations" are "succinct descriptions of innovative approaches to improving medical education" and often represent the product of scholarship of teaching.<sup>7</sup> The *JGIM* "Instructions for Authors" contain detailed specifications. The most important part of an "Innovation" study is demonstration that it is indeed innovative. This necessitates documentation of a thorough literature search. Evaluations of activities that have already been described might merit publication as "Original Research," but they are not appropriate as Innovations. Yet even when an idea has never been previously described, a diligent search will invariably identify previous work (empiric and theoretical) to support the approach followed. Scholarly innovations do not appear from thin air; they build on prior work.

Authors must describe the innovation, including both the educational objectives and the innovation itself, sufficiently well that a reader could implement or adapt the innovation at his/her own institution. As most of these articles represent the scholarship of teaching rather than research, a rigorous evaluation of the innovation is not mandatory. However, only the most innovative and best prepared ideas will merit publication without adequate evaluation. As the degree of innovation goes down, the evaluation rigor must go up. Even then, the key to a successful Educational Innovation publication will be a novel, well-described idea that

addresses an important need, has an adequate theoretical/empirical foundation, and builds on prior work.

Authors must demonstrate reflective critique by discussing what went well, what did not work as planned, how and why results vary from other studies, and areas for improvement and future research. Honesty and candor are not penalized. Indeed, an innovation with neutral or unexpectedly negative effects may have as much or more importance in publication as an innovation with statistically significant positive effects. However, the usual caveats of sample size, sensitivity of outcome measures, and strength of intervention apply in studies showing no effect.

## Survey Research

Much medical education research relies on a survey as a means of collecting data. Although this is strictly a method rather than a study type, its ubiquity justifies a brief discussion.

Surveys are subject to various sources of bias. They are susceptible to researcher bias in the wording of questionnaire items and the sample selection. Low response rates also introduce possible bias. Surveys often generate large amounts of data, which introduce the danger of bias from conducting multiple statistical analyses and then reporting only the statistically significant findings. Lengthy surveys can also breed long Results sections in which key points are lost amidst excessive data. We propose the following as a starting point for studies using surveys (in addition to the general standards of scholarship noted above) and refer authors to other sources<sup>26,55</sup> for details.

The research question should be clearly stated and justified. This will focus authors' collection, analysis, and presentation of data, and also ensure that the survey addresses an important issue.

Based on the research question, a study sample should be selected to reflect the population to which results will be generalized.

The questionnaire must have evidence to support the validity of its scores for answering this research question (see guidelines for the [development and evaluation of assessment tools](#) detailed below). If the study uses a previously published instrument, validity evidence should be concisely summarized and referenced. If the instrument is new, the study at a minimum should report evidence of content (breadth and depth of coverage of topic, systematic development process, qualifications of item writers, expert review, and pilot testing) and score reliability (from pilot testing, actual administration, or both).

Authors should report information on the format of survey administration (mail, Web, phone, other) and describe methods used to encourage response. Although there is no universal definition of adequate response rate, authors should keep this in mind while interpreting results.

It is virtually impossible to avoid investigator bias in studies that conduct multiple analyses and then report only those that are significant or interesting. The best defense against such problems is to develop a focused research question and plan all analyses in advance. When reporting, authors should describe all analyses conducted (including those whose results are not reported). Authors should account for independent comparisons using methods such as omnibus tests of statistical significance or Bonferroni's adjustment.<sup>36</sup> The Results and Discussion should highlight key points that support a clear message.

Authors should generally report verbatim the survey questions, along with any scoring rubrics, either in a table (reporting

questions and results in the same table) or as an appendix. It is rarely necessary to publish the actual instrument and saves space to report only the questions. If all questions are not reported, authors should report at least a few examples of typical questions.

## Needs Analysis Studies

“Needs analyses” are intended to identify the current state of a specific medical education issue. These frequently address potential deficiencies in content knowledge, but can also explore other educational “gaps,” such as work hour violations or inequities in academic promotion. Most studies evaluating educational interventions will have at least a rudimentary needs analysis, but studies designed as needs analyses face a higher bar.

Needs analyses can employ a variety of methods, including surveys and tests, focus groups, chart audits, task analyses, and literature reviews, but all pose challenges. First, such studies are particularly susceptible to researcher biases and special interests. If we looked hard enough, we would likely conclude that every issue in medical education has unmet needs—at least through the eyes of a person with a particular interest in that topic. Second, the results of a needs analysis depend greatly on the participants sampled and the instrument used. Unfortunately, we frequently see needs analyses employing poorly designed measures administered to convenience samples (e.g., a locally developed and administered anatomy exam). Finally, needs analysis studies often collect far more information than can reasonably be (or needs to be) reported and are susceptible to data analysis problems discussed under Survey Research.

Thus, we propose that needs analysis manuscripts meet four minimum requirements (in addition to other relevant standards). First, the research question should be clearly stated and justified. Second, the study sample must reasonably represent the target population (typically a national scope). Since a deficiency at one institution rarely indicates a national need, single-institution needs analyses—though important to an institution—will generally meet with skepticism. Third, the outcome measures must have evidence to support the “plausible” validity of scores. Authors should generally quote verbatim at least a subset of the items including the scoring rubric (i.e., in a table or as an appendix). Fourth, the Results and Discussion should highlight a clear, concise take-home message.

## Development and Evaluation of Assessment Tools

Studies describing the development, evaluation, or revision of assessment tools employ a variety of designs, but in all cases the investigators seek to support the validity of an instrument’s scores for making specific inferences.<sup>56,57</sup> Rather than try to address all possible study designs, we will discuss a framework or approach to validity that will facilitate high-quality studies.

The current conceptualization of validity unifies all different “types” of validity (content, criterion, construct, etc.) as “construct validity.”<sup>6,58–61</sup> Instruments are intended to generate scores reflecting some underlying construct, and validity is the degree to which scores truly reflect that construct.<sup>6</sup> We emphasize that validity is a property of scores, not instruments. Instruments are not inherently valid, but rather scores are valid for a particular purpose.

Validity is best viewed as an hypothesis supported by evidence from various sources.<sup>61</sup> As with any hypothesis, validity cannot be proven. Rather, investigators should create a validity *argu-*

*ment*<sup>62</sup> by first stating an initial hypothesis about what construct the scores should reflect; second, collecting evidence (see below) to support or refute that hypothesis (ideally testing the *weakest* assumption first); third, revising the hypothesis (either the instrument, the construct, or the context of application) if needed; and fourth, repeating the second and third steps until sufficient evidence has been collected to support (or reject) the validity argument. The sufficiency of evidence will vary depending on the application (a high-stakes Boards exam will require more evidence than a medical school second-year midterm). The evidence should answer the question: is it *plausible* that the scores reflect the intended construct?

There are five currently-accepted sources of validity evidence<sup>6</sup>: content (how well does the instrument match the intended construct domain?), response process (how do idiosyncrasies of the actual responses affect scores?), internal structure (typically psychometric data such as reliability or factor analysis), relations to other variables (how do scores relate to other variables that purport to measure a similar or different construct?), and consequences (do the scores make a difference?<sup>63</sup>). A publishable validity study will present data from *several* (but rarely all) complementary sources of evidence,<sup>64</sup> and ideally address the most critical or questionable aspects of the validity argument. Instruments intended for broad use often warrant a series of studies. Other sources contain further details and examples.<sup>56,58,59,61</sup> We discourage use of the term face validity<sup>61,65</sup> and note that this term is frequently misused to allude to content evidence.

Investigators can employ a similar approach when evaluating or adapting instruments for use in a particular study. When reporting the use of a previously described instrument, authors should briefly summarize the evidence supporting its scores for this application. For example, authors might write, “Felder and Solomon developed the Index of Learning Styles (ILS) to assess the ... learning style dimensions defined by Felder and Silverman. ... [Studies] have used internal consistency, test-retest reliability, and factor analysis to support the internal structure of ILS scores. ILS scores have also been shown to discriminate college students with different majors and college students from faculty.”<sup>66</sup>

## Evaluation Studies

Although much education research evaluates the outcomes of specific interventions, we touch only lightly on this study type because other sources<sup>25–28,67,68</sup> provide adequate guidance for authors. Guidelines developed for behavioral interventions in clinical medicine and public health, such as the TREND guidelines<sup>69</sup>, STROBE statement,<sup>70</sup> and CONSORT extension,<sup>71</sup> are also relevant. Authors should highlight an empirical or theoretical grounding for the intervention, focus on a gap in theory or educational practice, and conduct an appropriate evaluation using outcomes aligned with both the educational intervention and the study goals. Conceptual frameworks are useful for both applied and theory-building work.<sup>24</sup> Randomized designs are not required, but authors must carefully consider relevant validity threats.<sup>38,67</sup>

## Qualitative Research

Qualitative research will continue to proliferate as researchers recognize the limitations of quantitative methods in answering many important questions and gain necessary skills.<sup>72</sup> JGIM has long supported such studies.<sup>73</sup> However, such studies must

adhere to rigorous standards.<sup>30,74–79</sup> Key standards include a focused research question; appropriate sampling and data collection methodologies; inductive analytic methods that promote trustworthiness, credibility, dependability, and transferability (duplicate coding, triangulation, member checks, saturation, peer review, etc.); results that demonstrate a clear logic of inquiry and present appropriate data (i.e., themes and supporting quotations); and a synthesis with clear conclusions. We encourage use of accepted qualitative paradigms or approaches (grounded theory, ethnography, discourse analysis, etc). Mixed methods approaches (using both quantitative and qualitative methods) can often answer questions better than either approach alone.

## CONCLUSION

We anticipate that these standards will generate discussion and that they will continue to evolve with input from the education research community. In the meantime, *JGIM* editors will use these guidelines as part of the evaluation process for manuscripts received. We hope that medical education scholars will welcome our attempt to continue to raise the bar, and respond by submitting high-quality work to this journal and thereby advance scholarship in medical education.

---

**Conflict of Interest:** None disclosed.

**Funding source:** None.

**Corresponding Author:** David A. Cook, MD, MHPE; Office of Education Research and Division of General Internal Medicine, Mayo Clinic College of Medicine, Baldwin 4-A, 200 First Street SW, Rochester, MN 55905, USA (e-mail: cook.david33@mayo.edu).

## REFERENCES

1. Veet LL, Shea JA, Ende J. Our continuing interest in manuscripts about education. *J Gen Intern Med.* 1997;12:583–5.
2. Reed DA, Beckman TJ, Wright SM, Levine RB, Kern DE, Cook DA. Predictive validity evidence for medical education research study quality instrument scores: quality of submissions to *JGIM*'s medical education special issue. *J Gen Intern Med.* 2008;23:xxx–xx.
3. Bordage G, Caelleigh AS, Steinecke A, et al. Review criteria for research manuscripts. *Acad Med.* 2001;76:897–978.
4. American Educational Research Association. Standards for reporting on empirical social science research in AERA publications. *Educ Res.* 2006;35(6):33–40.
5. Education Group for Guidelines on Evaluation. Guidelines for evaluating papers on educational interventions. *BMJ.* 1999;318:1265–7.
6. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association; 1999.
7. Boyer EL. Scholarship reconsidered: Priorities of the professoriate. Princeton, NJ: Carnegie Foundation for the Advancement of Teaching; 1990.
8. Glassick CE, Huber MT, Maeroff G. Scholarship assessed: Evaluation of the professoriate. San Francisco, CA: Jossey-Bass; 1997.
9. Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM. Association between funding and quality of published medical education research. *JAMA.* 2007;298:1002–9.
10. Morrison J. Developing research questions in medical education: the science and the art. *Med Educ.* 2002;36:596–7.
11. McGaghie WC, Bordage G, Shea JA. Problem statement, conceptual framework, and research question. *Acad Med.* 2001;76:923–4.
12. Shea JA, Arnold L, Mann KV. A RIME perspective on the quality and relevance of current and future medical education research. *Acad Med.* 2004;79:931–8.
13. Prideaux D, Bligh J. Research in medical education: asking the right questions. *Med Educ.* 2002;36:1114–5.
14. Regehr G. Reporting of statistical analyses. *Acad Med.* 2001;76:938–9.
15. Regehr G. Presentation of results. *Acad Med.* 2001;76:940–2.
16. Cook DA, Beckman TJ, Bordage G. Quality of reporting of experimental studies in medical education: A systematic review. *Med Educ.* 2007;41:737–45.
17. Bordage G. Reasons reviewers reject and accept manuscripts: The strengths and weaknesses in medical education reports. *Acad Med.* 2001;76:889–96.
18. Bordage G, Dawson B. Experimental study design and grant writing in eight steps and 28 questions. *Med Educ.* 2003;37:376–85.
19. Albert M, Hodges B, Regehr G. Research in medical education: Balancing service and science. *Adv Health Sci Educ Theory Pract.* 2007;12:103–15.
20. Crandall SJ, Caelleigh AS, Steinecke A. Reference to the literature and documentation. *Acad Med.* 2001;76:925–7.
21. Bordage G. Moving the field forward: going beyond quantitative-qualitative. *Acad Med.* 2007;82(10 suppl):S126–8.
22. Gerrity MS. Medical education and theory-driven research. *J Gen Intern Med.* 1994;9:354–5.
23. Cook DA. The research we still are not doing: an agenda for the study of computer-based learning. *Acad Med.* 2005;80:541–8.
24. Cook DA, Bordage G, Schmidt HG. Description, justification, and clarification: a framework for classifying the purposes of research in medical education. *Med Educ.* 2008;42:128–33.
25. Green JL, Camilli G, Elmore PB. Handbook of complementary methods in education research. Mahway, NJ: Lawrence Erlbaum; 2006.
26. Fraenkel JR, Wallen NE. How to design and evaluate research in education. New York, NY: McGraw-Hill; 2003.
27. Cook TD, Campbell DT. Quasi-Experimentation: Design and Analysis Issues for Field Settings. Boston: Houghton Mifflin; 1979.
28. Cronbach LJ. Designing Evaluations of Educational and Social Problems. San Francisco: Jossey-Bass; 1982.
29. Norman GR, Streiner DL. Biostatistics: The Bare Essentials. 3rd ed. Hamilton: BC Decker; 2007.
30. Miles MB, Huberman AM. Qualitative Data Analysis: An Expanded Sourcebook. Thousand Oaks, CA: Sage; 1994.
31. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA.* 2006;296: 1094–102.
32. Kirkpatrick D. Revisiting kirkpatrick's four-level model. *Train Dev.* 1996;50(1):54–9.
33. Prystowsky JB, Bordage G. An outcomes research perspective on medical education: the predominance of trainee assessment and satisfaction. *Med Educ.* 2001;35:331–6.
34. Chen FM, Bauchner H, Burstin H. A call for outcomes research in medical education. *Acad Med.* 2004;79:955–60.
35. Shea JA. Mind the gap: some reasons why medical education research is different from health services research. *Med Educ.* 2001;35:319–20.
36. Bland JM, Altman DG. Statistics notes: Multiple significance tests: the Bonferroni method. *BMJ.* 1995;310:170.
37. Wilkinson L, Task Force on Statistical Inference. Statistical methods in psychology journals: guidelines and explanations. *Am Psychol.* 1999;54:594–604.
38. Colliver JA, McGaghie WC. The reputation of medical education research: quasi-experimentation and unresolved threats to validity. *Teach Learn Med.* 2008;20:101–3.
39. Price EG, Beach MC, Gary TL, et al. A systematic review of the methodological rigor of studies evaluating cultural competence training of health professionals. *Acad Med.* 2005;80:578–86.
40. Cook DA, Beckman TJ, Bordage G. A systematic review of titles and abstracts of experimental studies in medical education: Many informative elements missing. *Med Educ.* 2007;41:1074–81.
41. Thompson B. Research synthesis: effect sizes. In: Green JL, Camilli G, Elmore PB, eds. Handbook of Complementary Methods in Education Research. Mahway, NJ: Lawrence Erlbaum; 2006:583–603.
42. Hojat M, Xu G. A visitor's guide to effect sizes. *Adv Health Sci Educ Theory Pract.* 2004;9:241–9.
43. Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. More informative abstracts revisited. *Ann Intern Med.* 1990;113:69–76.
44. Tomkowiak JM, Gunderson AJ. To IRB or Not to IRB? *Acad Med.* 2004;79:628–32.

45. **Ross JS, Hill KP, Eglman DS, Krumholz HM.** Guest authorship and ghostwriting in publications related to rofecoxib: a case study of industry documents from rofecoxib litigation. *JAMA*. 2008;299:1800–12.
46. **Schwartz RS, Curfman GD, Morrissey S, Drazen JM.** Full disclosure and the funding of biomedical research. *N Engl J Med*. 2008;358:1850–1.
47. **Mojon-Azzi SM, Mojon DS.** Scientific misconduct: from salami slicing to data fabrication. *Ophthalmologica*. 2004;218:1–3.
48. **Beckman TJ, Cook DA.** Developing scholarly projects in education: a primer for medical teachers. *Med Teach*. 2007;29:210–8.
49. **Carney PA, Nierenberg DW, Pipas CF, Brooks WB, Stukel TA, Keller AM.** Educational epidemiology: applying population-based design and analytic approaches to study medical education. *JAMA*. 2004;292:1044–50.
50. **Bordage G.** Considerations on preparing a paper for publication. *Teach Learn Med*. 1989;1:47–52.
51. **Wolf FM, Shea JA, Albanese MA.** Toward setting a research agenda for systematic reviews of evidence of the effects of medical education. *Teach Learn Med*. 2001;13:54–60.
52. **Reed D, Price EG, Windish DM, et al.** Challenges in systematic reviews of educational intervention studies. *Ann Intern Med*. 2005;142:1080–9.
53. **Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF.** Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Quality of Reporting of Meta-analyses*. *Lancet*. 1999;354:1896–900.
54. **Stroup DF, Berlin JA, Morton SC, et al.** Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA*. 2000;283:2008–12.
55. **Fink A.** How to conduct surveys: a step-by-step guide. Thousand Oaks, CA: Sage; 2005.
56. **Kane MT.** Validation. In: Brennan RL, ed. *Educational measurement*, 4th ed. Westport: Praeger; 2006:17–64.
57. **DeVellis RF.** Scale development: theory and applications. 2Thousand Oaks, CA: Sage Publications; 2003.
58. **Downing SM.** Validity: on the meaningful interpretation of assessment data. *Med Educ*. 2003;37:830–7.
59. **Downing SM, Haladyna TM.** Validity threats: overcoming interference with proposed interpretations of assessment data. *Med Educ*. 2004;38:327–33.
60. **Downing SM.** Reliability: on the reproducibility of assessment data. *Med Educ*. 2004;38:1006–12.
61. **Cook DA, Beckman TJ.** Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med*. 2006;119:166.e7–16.
62. **Kane MT.** An argument-based approach to validity. *Psychol Bull*. 1992;112:527–35.
63. **Lord SJ, Irwig L, Simes RJ.** When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? *Ann Intern Med*. 2006;144:850–5.
64. **Beckman TJ, Cook DA, Mandrekar JN.** What is the validity evidence for assessments of clinical teaching? *J Gen Intern Med*. 2005;20:1159–64.
65. **Downing SM.** Face validity of assessments: faith-based interpretations or evidence-based science? *Med Educ*. 2006;40:7–8.
66. **Cook DA.** Reliability and validity of scores from the Index of Learning Styles. *Acad Med*. 2005;80(10 suppl):S97–101.
67. **Cook DA, Beckman TJ.** Reflections on experimental research in medical education. *Adv Health Sci Educ Theory Pract*. 2008; Epub ahead of print 22 April 2008; DOI [10.1007/s10459-008-9117-3](https://doi.org/10.1007/s10459-008-9117-3).
68. **Wilkes M, Bligh J.** Evaluating educational interventions. *BMJ*. 1999;318:1269–72.
69. **Des Jarlais DC, Lyles C, Crepaz N.** Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health*. 2004;94:361–6.
70. **von Elm E, Altman DG, Egger M, et al.** The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med*. 2007;147:573–7.
71. **Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P,** for the CONSORT Group. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med*. 2008;148:295–309.
72. **Harris I.** Qualitative methods. In: Norman G, Van der Vleuten C, Newble D, eds. *International Handbook of Research in Medical Education*. Dordrecht: Kluwer Academic Publishers; 2002:97–126.
73. **Inui TS, Frankel RM.** Evaluating the quality of qualitative research: A proposal pro tem. *J Gen Intern Med*. 1991;6:485–6.
74. **Devers KJ.** How will we know “good” qualitative research when we see it? Beginning the dialogue in health services research. *Health Serv Res*. 1999;34(5 part 2):1153–88.
75. **Elliott R, Fischer CT, Rennie DL.** Evolving guidelines for publication of qualitative research studies in psychology and related fields. *Br J Clin Psychol*. 1999;38:215–29.
76. **Malterud K.** Qualitative research: standards, challenges, guidelines. *Lancet*. 2001;358:483–8.
77. **Côté L, Turgeon J.** Appraising qualitative research articles in medicine and medical education. *Med Teach*. 2005;27:71–5.
78. **Giacomini MK, Cook DJ,** for the Evidence-Based Medicine Working Group. Users' guides to the medical literature: XXIII. qualitative research in health care B. what are the results and how do they help me care for my patients? *JAMA*. 2000;284:478–82.
79. **Giacomini MK, Cook DJ,** for the Evidence-Based Medicine Working Group. Users' guides to the medical literature: XXIII. qualitative research in health care A. are the results of the study valid? *JAMA*. 2000;284:357–62.