AAMC

Tomorrow's Doctors, Tomorrow's Cures®

# Review Criteria for Research Manuscripts

## 2nd Edition

Learn

Serve

Lead

Association of
American Medical Colleges

# Review Criteria for Research Manuscripts

## 2nd Edition

### A joint project of *Academic Medicine* and the Group on Educational Affairs' Medical Education Scholarship Research and Evaluation Section

**Editors:**
**Steven J. Durning, MD, PhD**
**Jan D. Carline, PhD**

**August 2015**

Suggested citation:
Durning SJ, Carline JD, eds. Review Criteria for Research Manuscripts, 2nd ed. Washington, DC: Association of American Medical Colleges; 2015.

Suggested citation (example) for individual chapters:
Dine CJ, Caelleigh AS, Shea JA. Chapter 1: Review process. In Durning SJ, Carline JD, eds. Review Criteria for Research Manuscripts, 2nd ed. Washington, DC: Association of American Medical Colleges; 2015: pp 3–6.

---

This is a publication of the Association of American Medical Colleges.
The AAMC serves and leads the academic medicine community to improve the health of all. www.aamc.org.

## Table of Contents

# Foreword
# Improving Scholarship: The Role of Peer Review

**David Sklar, MD**

**D. Sklar** is editor-in-chief, *Academic Medicine*, and associate dean emeritus and distinguished professor emeritus, Graduate Medical Education, University of New Mexico Health Science Center, Department of Emergency Medicine, Albuquerque, New Mexico.

In September 2001, *Academic Medicine* published "Review Criteria for Research Manuscripts." Now, 14 years later, we have updated the resource and added new material as well. We are very fortunate that many of the authors from the first edition have been willing to help us with this second edition.

The staff and editors of *Academic Medicine* decided to undertake this project in conjunction with the Group on Educational Affairs and the "Review Criteria for Research Manuscripts" task force in an effort to promote scholarship. We depend on peer reviewers to inform publication decisions. Peer review also provides us and the authors of manuscripts with advice about how to improve submissions so that the published report represents the highest-quality scholarship possible. The "Review Criteria for Research Manuscripts" is intended to familiarize our reviewers, both new and experienced, with the purposes of review, approaches to or best practices for reviewing, and criteria for superlative research. We hope this information will help reviewers organize their recommendations and communicate them effectively to *Academic Medicine* and similar journals. We also hope that the "Review Criteria for Research Manuscripts" will help researchers by explaining the criteria we use to evaluate their submissions.

This guide would not have been possible without the concerted efforts of our project task force, led by associate editors Steven Durning, MD, PhD, and Jan Carline, PhD, and staff editor Elizabeth S. Karlin, MA. I also want to acknowledge editorial assistant John Remski, who helped manage the logistics of the update, and the other members of the task force. Their commitment to this project has produced a high-quality document that goes beyond the original purpose—updating previous information—to providing a resource of general value for faculty development. I am also grateful to the individual authors of chapters who donated their time and expertise to succinctly capture the essence of complex topics.

Peer review is a time-consuming and imperfect process, but I have often been pleasantly surprised at how the process can result in radically improved research reports that ultimately have substantial impact on our thinking about difficult problems. I am hopeful that our community will find that this updated "Review Criteria for Research Manuscripts" guide contributes to our efforts to improve peer review and, in so doing, improves the quality of published research.

*Funding/Support:* None reported.
*Other disclosures:* None reported.

# Foreword
# A Measure of Success

**Elizabeth A. Nelson, MD, Karen Szauter, MD, and Brian Mavis, PhD**

**E. A. Nelson** is associate professor, Department of Medicine, Dell Medical School, Austin, Texas.

**K. Szauter** is professor, Department of Internal Medicine, University of Texas Medical Branch Galveston, Galveston, Texas.

**B. Mavis** is associate professor, Office of Medical Education Research and Development, College of Human Medicine, Michigan State University, East Lansing, Michigan.

As three past and present chairs of the Association of American Medical Colleges' Group on Educational Affairs (GEA), we are encouraged by the expansion of medical education scholarship over the past decade, and we are excited to introduce the second edition of "Review Criteria for Research Manuscripts," first published in 2001. The updated, expanded, and new chapters of this guide will support our growing community of medical education scholars.

In his seminal work, Ernest L. Boyer[1] proposed that scholarship be broadened beyond an emphasis on discovery. Glassick[2] then helped define standards of excellence in the scholarship of teaching, and medical educators called for academic recognition of legitimate activities vital to the fulfillment of medical schools' educational mission.[3] The GEA echoed this call at its 2006 consensus conference.[4] The following year, an article in *Medical Education*[5] expanded the community's understanding of scholarship, stretching research beyond traditional forms and definitions to include teaching, curriculum design, educational leadership, mentoring, and learner assessment. With this broader understanding of what research entails, medical education scholarship expanded in scope, format, and venue. Despite this growth in scholarship, few—if any—programs have been developed to help foster skill development in peer review.

The original "Review Criteria for Research Manuscripts" guide was created to be an aid for reviewers to improve the quality of their peer reviews.[6] High-quality reviews are vital both for ensuring the excellence of published scholarship and as a way to provide authors with important feedback to use to improve their scholarship. The importance of the reviewer's guide to faculty development cannot be overestimated.

The original guide has helped demystify what many novice scholars feel is the "black box" of the peer-review process. Many faculty members have shared the guide with colleagues and trainees as a writing resource and to provide insights into what to expect during peer review. The guide has also been used as a resource in many faculty development programs, including the GEA-sponsored Medical Education Research Certificate (MERC) program.

This updated version of the guide includes modified chapters from the original work and several new sections. As dissemination of scholarly work in medical education continues to evolve, review guidelines for different types of work are needed. The new collection includes chapters on reporting qualitative findings, reviewing literature reviews, and reviewing reports on innovations. The Checklist of Review Criteria, a very useful document to keep on hand when reviewing a report, has also been updated to reflect the guide's content. As with the first edition, this collection serves as a wonderful resource for faculty development, mentoring, and personal use.

A goal of the GEA is to promote excellence in the education of physicians throughout their professional lives and, thereby, to contribute to improving the health of the public. Part of the GEA's mission is "the advancement of research in medical education and the dissemination of the results of that research."[7] To that end, the "Review Criteria for Research Manuscripts" task force undertook this project in collaboration with the GEA's Medical Education Scholarship Research and Evaluation (MESRE) section and *Academic Medicine*. Steven Durning, MD, PhD (Uniformed Services University of the Health Sciences), Jan Carline, PhD (University of Washington School of Medicine), and Elizabeth Karlin, MA (*Academic Medicine*), led a distinguished group of faculty writing teams in the revision of this guide. We thank the members of the task force and the authors who contributed to the chapters for their time and expertise.

*Acknowledgments:* The authors would like to thank Katherine McOwen for all her support of the Group on Educational Affairs.
*Funding/Support:* None reported.
*Other disclosures:* None reported.

## References

1. Boyer EL. Scholarship Reconsidered: Priorities of the Professoriate. Princeton, NJ: Carnegie Foundation for the Advancement of Teaching; 1990.
2. Glassick CE. Boyer's expanded definition of scholarship, the standards for assessing scholarship, and the elusiveness of the scholarship of teaching. Acad Med. 2000;75:877–880.
3. Fincher RM, Simpson DE, Mennin SP, et al. Scholarship in teaching: An imperative for the 21st century. Acad Med. 2000;75:887–894.
4. Simpson D, Fincher RM, Hafler JP, et al. Advancing educators and education: Defining the components and evidence of educational scholarship. In Proceedings from the Association of American Medical Colleges Group on Educational Affairs Consensus Conference on Educational Scholarship, February 9–10, 2006, Charlotte, N.C. Washington, DC: Association of American Medical Colleges; 2007.
5. Simpson D, Fincher RM, Hafler JP. Advancing educators and education by defining the components and evidence associated with educational scholarship. Med Educ. 2007;41:1002–1009. Epub September 5, 2007.
6. Joint Task Force of Academic Medicine and the GEA-RIME Committee. Review criteria for research manuscripts. Acad Med. 2001;76:897–978.
7. Association of American Medical Colleges. About the GEA. Mission Statement. 2015. https://www.aamc.org/members/gea/about_the_gea/. Accessed March 16, 2015.

# Preface
# Enhancing a Resource for Reviewers: An Update to the Guide

**Steven J. Durning, MD, PhD, and Jan D. Carline, PhD**

**S. J. Durning** is professor, Medicine and Pathology, Department of Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland.

**J. D. Carline is** professor, Biomedical Informatics and Medical Education, Departments of Family Medicine and Pharmacy, University of Washington, Seattle, Washington.

The guide that you are reading represents an update to a seminal work in the field of medical education research, "Review Criteria for Research Manuscripts," published in *Academic Medicine* in 2001. As with the first edition of the guide, our intention is to provide a resource for reviewers to use as they perform reviews of research in health professions education.[1] While we believe that this guide will be most helpful to more junior reviewers, we hope that it will assist reviewers across the continuum of experience by illuminating the tasks and the focus of this professional role. A secondary intent is to help authors with their writing through providing insight into what reviewers are looking for in research reports. Finally, as with the previous guide, we have sought to orient individuals to how journals work and to the review process itself.

Like the first guide, this work represents a collaboration between *Academic Medicine* and the Group on Educational Affairs (GEA). We are very grateful for the work of the 2014–2015 "Review Criteria for Research Manuscripts" task force, most of whose members were also members of the initial task force. This core group of educators first updated the review criteria through phone calls and face-to-face meetings. We then contacted the authors of the initial guide, almost all of whom agreed to refresh their chapters. We encouraged the original authors to invite lead authors who are up and coming in our field. Along with these more junior colleagues, they diligently and thoughtfully updated their content to reflect what we believe to be the science of reviewing as it stands today. We are also grateful for our peer reviewers. The product before you would not have been possible without everyone's input.

We would like to say a few words about why and how we updated the reviewer guide. First, the medical education research community has commented on the usefulness of the resource but also expressed an interest in expanding it to cover more recent research methods and newer formats. Although all the original sections of the guide are present, changes have been made and new sections have been added. Examples of what is new include recommendations for reviewing qualitative research reports, literature reviews, and descriptions of innovations. We have intentionally not included reviewer guidelines on articles, commentaries, or perspectives in this resource; rather, we have sought to build on the science of performing peer review of research reports and related papers. We have also updated the chapter on research ethics to reflect current concerns, such as quality-improvement projects and the protection of human participants. Finally, the authors have refreshed references to reflect recent scholarship in publishing.

The guide is not a replacement either for the technical knowledge and skills needed to critique statistical or methodological content or for the experience with scholarship required to suggest improvements to rationale or structure. While individual reviewers must judge their own skills in assessing these aspects of a manuscript, the guide suggests general approaches to the task of critical review and underscores the scope and importance of this work to the field of medical education. We hope that the guide continues to provide support and assistance to those engaged in the vital work of reviewing.

Some readers may wonder how to use this guide. We have updated it with the same expectations for its use as for the initial edition.[2] We sought to provide purposefully brief and explicit criteria for reviewers to follow. Additionally, because our intent is to enable reviewers and authors alike to envision what success looks like for each section of a research manuscript, each criterion is stated in positive terms. Also, through our vetting process, we attempted to simplify the list of criteria and to ensure that the each criterion is as general as possible so that the list might apply to the growing body of manuscript types submitted to health professions education journals. Still, each criterion should not be applied to every manuscript—certain criteria are applicable only to specific submission types. Similarly, we do not suggest that each applicable criterion be weighed equally. We also do not advocate using the list of criteria as a rigid formula for determining the quality of a submission. Rather, the review guide is meant to serve as a general resource.

The criteria presented here represent a consensus among a number of communities of practice—reviewers, authors, and editors—whose collective expertise spans a growing number of fields of research and who all seek the highest-possible quality standards for research published in the field of medical education. Readers may note a natural redundancy to some of the criteria, which we grappled with as a

group. In the end, we felt the redundancy reflected the iterative process of reviewing (and writing) a research report such that the "story" builds with each subsequent section. Finally, as in the 2001 edition of the guide, the format of the chapters is similar; each one focuses on one topic and begins with the list of relevant criteria. The list is followed by an in-depth discussion of the criteria and additional considerations about the topic at hand.

So, given this, where does a reviewer begin? The order of the chapters suggests a reasonable approach, starting with the initial establishment of a research question to answer, then moving through the selection of methods, then analysis, and, finally, the discussion and conclusions. Reviewers may also choose to use this resource on an as-needed basis, referring to it only when they have a question about a particular section of a manuscript and want to consider applying relevant criteria. In other words, these guidelines are just that: guidelines. This list and discussion of criteria are not meant to establish a singular approach. The guide does not represent complete or detailed criteria by which to judge all manuscripts. Some research methods may be subject to criteria beyond those included here, and reviewers may need to reference other sources for appropriate guidance about methods or even presentation in manuscripts. The criteria and discussions presented here can help to remind established reviewers of important considerations for reviewing manuscripts in medical education, and they can help in the training or mentoring of reviewers, educators, and scholars new to the review process.

*Funding/Support:* None reported.
*Other disclosures:* None reported.
*Disclaimer:* The views expressed herein are the authors' own and do not necessarily represent those of the Uniformed Services University of the Health Sciences, the U.S. Department of Defense, or the federal government of the United States.

## References

1. Bordage G, Caelleigh AS. Introduction: A tool for reviewers: "Review Criteria for Research Manuscripts." Acad Med. 2001;76:904–908.
2. Bordage G, Caelleigh AS. Introduction: How to read "Review Criteria for Research Manuscripts." Acad Med. 2001;76:909–910.

## Review Criteria for Research Manuscripts Task Force Members

**Christopher S. Candler, MD, EdD**
Professor of Medicine
Senior Associate Dean Academic Affairs
University of Oklahoma College of Medicine
University of Oklahoma
Oklahoma City, Oklahoma

**Jan D. Carline, PhD**
Professor of Biomedical Informatics and Medical Education
Department of Family Medicine and School of Pharmacy
University of Washington, Seattle, Washington
Associate Editor
*Academic Medicine*
Association of American Medical Colleges
Washington, DC

**Sonia Crandall, PhD, MS**
Professor
Department of Physician Assistant Studies
Wake Forest School of Medicine
Winston-Salem, North Carolina

**Mary Beth DeVilbiss**
Managing Editor
*Academic Medicine*
Association of American Medical Colleges
Washington, DC

**Steven J. Durning, MD, PhD**
Professor of Medicine and Pathology
Department of Medicine
Uniformed Services University of the Health Sciences
Bethesda, Maryland
Associate Editor
*Academic Medicine*
Association of American Medical Colleges
Washington, DC

**Anne L. Farmakidis, MPS**
Senior Director of Publishing
Association of American Medical Colleges
Washington, DC

**Larry Gruppen, PhD**
Professor of Medical Education
Department of Learning Health Sciences
University of Michigan
Ann Arbor, Michigan

**Elizabeth S. Karlin, MA**
Staff Editor
*Academic Medicine*
Association of American Medical Colleges
Washington, DC

**Karen V. Mann, MSc, PhD**
Professor Emeritus
Division of Medical Education
Dalhousie University
Halifax, Nova Scotia, Canada

**Elizabeth Nelson, MD**
Associate Professor and Associate Dean of Undergraduate Medical Education
Department of Medical Education
Dell Medical School
Austin, Texas

**Patricia S. O'Sullivan, EdD**
Professor of Medicine and Surgery
Director, Research and Development in Medical Education, School of Medicine
University of California San Francisco
San Francisco, California

**Judy A. Shea, PhD**
Professor of Medicine
Associate Dean
Medical Education Research and Assessment
Perelman School of Medicine
University of Pennsylvania
Philadelphia, Pennsylvania

**David Sklar, MD**
Associate Dean Emeritus, Distinguished Professor Emeritus
Graduate Medical Education
University of New Mexico Health Science Center
Department of Emergency Medicine
Albuquerque, New Mexico
Editor-in-Chief
*Academic Medicine*
Association of American Medical Colleges
Washington, DC

**Karen Szauter, MD**
Professor
Internal Medicine
University of Texas Medical Branch Galveston
Galveston, Texas

**Colin P. West, MD, PhD**
Professor of Medicine, Medical Education, and Biostatistics
Department of Medicine and Department of Health Sciences Research
Mayo Clinic
Rochester, Minnesota

## The Checklist of Review Criteria

| | |
|---|---|
| **Group 1: Problem Statement, Conceptual Framework, and Research Question** | |
| | 1. The introduction builds a logical case and provides context for the problem statement. |
| | 2. The problem statement is clear and well articulated. |
| | 3. The conceptual framework is explicit and justified. |
| | 4. The research purpose and/or question (as well as the research hypothesis, where applicable) is clearly stated. |
| | 5. The constructs being investigated are clearly identified and presented. |
| **Group 2: Reference to the Literature and Documentation** | |
| | 1. The literature review is comprehensive, relevant, and up-to-date. |
| | 2. The literature is analyzed and critically appraised; gaps in the literature are identified as a basis for the study. |
| **Group 3: Relevance** | |
| | 1. The study is relevant to the mission of the journal or its audience. |
| | 2. The study addresses important problems or issues; the study is worth doing. |
| | 3. For quantitative studies: the study has generalizability because of the selection of participants, setting, and educational intervention or materials. |
| | 4. For qualitative studies: the study offers concepts or theories that are generalizable or transferable to other contexts, people, etc. |
| **Group 4: Research Design** | |
| | 1. The research paradigm or approach is identified. |
| | 2. The design is appropriate for the research purpose or question. If a mixed-methods approach is used, the rationale is provided for the relationship between and sequencing of quantitative and qualitative aspects of the study. |
| | 3. For quantitative studies: the design has internal validity, and potential confounding variables or biases are addressed. |
| | 4. For quantitative studies: the design has external validity, including participants, settings, and conditions. |
| | 5. For qualitative studies: the study design incorporates techniques to ensure trustworthiness. |
| | 6. For studies with interventions: the intervention is described in sufficient detail (objectives, activities, time allocation, training) to be able to assess the likelihood of the intervention having the desired effect and/or to permit the study to be replicated. |
| | 7. The research methods are defined and clearly described, and they are sufficiently detailed to provide transparency or permit the study to be replicated. |
| **Group 5: Instrumentation, Data Collection, and Quality Control** | |
| | 1. The development and content of the instrument(s)—as well as the preparation of observers, interviewers, and raters, as appropriate—are sufficiently described or referenced and are sufficiently detailed to permit transparency and/or replication. |
| | 2. For qualitative studies: the characteristics of the researchers that may influence the research are described and accounted for during data collection. |
| | 3. The measurement instrument is appropriate given the study's variables; the scoring method is clearly defined. |
| | 4. The psychometric properties and procedures are clearly presented and appropriate. |
| | 5. The data set is sufficiently described or referenced. |
| | 6. Data quality control is described and is adequate. |
| **Group 6: Population and Sample** | |
| | 1. For quantitative studies: the population is defined in sufficient detail to permit the study to be replicated. |
| | 2. The sampling procedures are described in sufficient detail to permit transparency, replication, or theory generation. |
| | 3. Samples are appropriate to the research purpose or question. |
| | 4. Selection bias is addressed. |

| | |
|---|---|
| **Group 7: Data Analysis and Statistics** | |
| | 1. Data-analysis procedures are described in sufficient detail. |
| | 2. Data-analysis procedures conform to the research design; hypotheses, models, or theory drives the data analyses. |
| | 3. Statistical tests are appropriate. |
| | 4. Topics such as effect size or functional significance, multiple tests or comparisons, and adjustment of significance level for chance outcomes were considered. |
| | 5. Power issues are considered in studies that make statistical inferences. |
| | 6. For qualitative analysis: how members of the research team contributed to coding, identifying themes, and/or drawing inferences is described; methods used to ensure trustworthiness of the analysis are also described. |
| **Group 8: Presentation of Results** | |
| | 1. All results are presented. The results align with the methods and study questions. |
| | 2. The amount of data presented is sufficient, balanced, accurate, and supportive of inferences or themes. |
| | 3. Tables, graphs, or figures are used judiciously and agree with the text. |
| | 4. The statistics are reported correctly and appropriately. |
| **Group 9: Discussion and Conclusion—Interpretation** | |
| | 1. The conclusions are clearly stated; key points stand out. |
| | 2. The conclusions follow from the design, methods, and results. |
| | 3. The study limitations are discussed. |
| | 4. Findings are placed in the context of relevant literature, and alternative interpretations are considered as needed. |
| | 5. Practical significance or theoretical implications are discussed; guidance for future studies is offered. |
| **Group 10: Title, Authors, and Abstract** | |
| | 1. The title is clear, informative, and representative of the content. |
| | 2. The abstract contains essential details. |
| | 3. The conclusions in the abstract are justified by the information in the abstract and the text. |
| | 4. There are no inconsistencies in detail among the abstract, text, tables, and figures. |
| | 5. All the information in the abstract is present in the text. |
| **Group 11: Presentation and Documentation** | |
| | 1. The text is well written and easy to follow. |
| | 2. The manuscript is well organized. |
| **Group 12: Scientific Conduct** | |
| | 1. Ideas and materials of other authors are correctly attributed. (There are no instances of plagiarism). |
| | 2. Prior publication by the author(s) of substantial portions of the data or study is appropriately acknowledged. |
| | 3. Any apparent conflict of interest is appropriately disclosed. |
| | 4. There is an explicit statement of ethical review and approval (e.g., by an institutional review board [IRB]) for studies directly involving human subjects or data about them. |

# Part 1

# Chapter 1
# Review Process

**C. Jessica Dine, MD, MSHPR, Addeane S. Caelleigh, MA, and Judy A. Shea, PhD**

**C. J. Dine** is assistant professor of medicine, Division of Allergy, Pulmonary, and Critical Care, and associate program director, Internal Medicine Residency Program, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

**A. S. Caelleigh** was the Liaison Committee of Medical Education coordinator, the Office of Medical Education, and she is now a visiting faculty member, University of Virginia School of Medicine, Charlottesville, Virginia.

**J. A. Shea** is professor of medicine and associate dean, Medical Education Research and Assessment, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

The review process starts as soon as authors submit a manuscript. The first decision focuses on whether the report should be sent for external review. Practices vary across journals, but typically, one or more editors read the manuscript and decide whether to ask experts (the authors' peers) to review the manuscript or to reject it without review. General factors that the editor considers when making this first decision include, but are not limited to, the following:

- fit in terms of both the mission of the journal and the match between the focus of the manuscript and the interests of the journal subscribers,
- the quality of the presentation,
- the soundness of the research methodology, and
- whether the manuscript adds to the current literature and how important the article seems to be.[1]

The primary reason the editor decides to reject at this stage is to minimize reviewer fatigue by decreasing the number of reviews he or she needs to solicit. In almost all cases, reviewers do not receive compensation for the significant amounts of time, ranging from a few hours to a full work day, that they invest in reading and critiquing each manuscript.[2] (Yankauer and colleagues report that the time to review typically ranged from 45 minutes to 8 hours; the median time was 2.7 hours.[2]) Some journals, including *Academic Medicine*, now offer continuing medical education credits for high-quality reviews.

If the initial decision is to send the manuscript out for review, reviewers are selected and invited to review (see also Chapter 2). Journals have different types of review systems. Two common systems are "board review" and "pool review." In the former, all the reviewers are members of a review board or an editorial board that maintains responsibility for collectively reviewing the journal's manuscripts; in the latter, a large number of specialists of many types join the journal's pool of potential reviewers.[3,4] Most journals use a mixed or hybrid form, in which the editor relies heavily on pool review for initial reviews of manuscripts but also uses board review, especially to add another perspective to, or resolve discrepancies among, the initial reviews.

The exact mechanism by which journals select reviewers also varies. Some journals may ask authors to nominate potential reviewers for their manuscripts, and in some cases, reviewers are invited to nominate other potential reviewers. Reviewers may be chosen based on their content or their methodologic expertise, and the mix of reviewers for a particular manuscript depends on the unique needs of that manuscript as well as the established review standards of the journal.

Once potential reviewers have been identified, each receives an invitation to review the manuscript. The invitation to review typically includes a requested time frame. Potential reviewers should carefully consider the deadline when deciding whether to accept the invitation. Journals differ in the amount of time they give to reviewers, the timing and number of reminders they send before and/or after deadlines, and the processes they use for deciding if and when to seek additional reviewers when a promised review has not been received.

Journals also differ in the number of reviews they seek for each manuscript. Most journals request reviews from two or three external reviewers, while others use more. One decision journal editors must make is whether to request more reviews than necessary, in the hope that at least the minimum number will be completed in a reasonable time. Alternatively, the policy may be to request a small number and then follow up with invitations to additional reviewers when either the initial reviews are contradictory or one or more is unexpectedly delayed.

A key aspect of the review process is whether manuscript authors are blinded to the identity of the reviewers and vice versa. Policies about masking reviewers and authors vary from journal to journal. The traditional practice is to conceal each reviewer's identity from the authors, but some believe that identifying reviewers makes their comments more fair and balanced and that this openness helps define the review process as a collegial "dialogue" from which both authors and reviewers benefit.[5] Studies suggest that asking reviewers to sign their reviews does not affect their recommendations to publish or reject manuscripts.[6,7] Further, another study has shown that the quality of reviews did not change even when the reviewers knew that their signed reviews might be made available online.[8] However, more requests to review are declined when the identity of the reviewer is going to be made public than when reviewers remain anonymous.[7,8] Informally, it is fair to say that the balance of reviewers' sentiment is for anonymous reviews.[9] Power imbalances could make it

difficult for some reviewers (junior faculty members, for example) to be appropriately critical if they have to sign their reviews. Few educational research journals require reviewers to sign their reviews.

Concealing the authors' identities from the reviewer is a relatively uncommon practice, and findings from studies of the practice are mixed. Some research has shown that masking authors' identities leads to less-biased reviews,[10–12] but other work suggests that masking authors' identities does *not* improve the quality of reviews.[5,13,14] Although, importantly, most evidence indicates that masking authors' identities has no effect on the publication decision,[10] a few studies suggest that the knowledge of certain characteristics of the authors may influence recommendations to accept or reject, after all.[15] One study, for example, has demonstrated that authors who were English speaking, from the United States, or from a prestigious institution were more likely to have their submitted abstract for a national meeting accepted than authors who did not meet any of these criteria.[16]

Completely masking the authors' identities from the reviewer is difficult, especially if the reviewer is an expert in the field—particularly a niche field—who may know others working in the area. Also, because many authors cite their previous or ongoing work in the manuscript,[17] many scholars, especially reviewers themselves, believe that blinding is not uniformly effective since reviewers can often identify the authors through those citations.[10,14,18,19] Interestingly, though neither the quality nor the quantity of the evidence is strong, the research to date seems to show that reviewers are not as accurate as they think they are in guessing authors' identities.[5,20] Regardless, most journals do not currently blind the reviewer to authors' identities, but they do ask the reviewer to identify any areas of possible conflict of interest (see also Chapter 23).

Another form of masking is very rarely used but worth noting: blinding the *editor* to the identity of the authors, the reviewers, or both until he or she makes the final publication decision. This practice is based on the same reasoning that underlies blinding reviewers to the identity of the authors, which is to limit any positive or negative influence on the decision process.

Regardless of the type of blinding used, after reviewers accept the invitation to review a manuscript, they receive access to files that include the manuscript and any exhibits (e.g., tables, figures), as well as the electronic review forms, instructions, and sometimes associated materials (e.g., instructions to authors, sample reviews). The goal of review is to garner a high-quality assessment of the manuscript that can inform not only the editor making the publication decision but also the authors, who may use any constructive feedback to strengthen their manuscript.

Reviewers are also usually asked to suggest a recommendation about publication. A recommendation is a judgment based on a reviewer's overall assessment of the worth of the manuscript. Naturally, this recommendation will vary depending on the number of categories the journal allows (e.g., accept/reject, accept/provisional accept/revise/reject, accept/revise/reject). The reviewer should base this recommendation on both a quantitative assessment (e.g., ratings, Likert-type scales) and a qualitative assessment (i.e., narrative comments) of the manuscript. The narrative comments may be especially helpful; looking at the number of major issues allows the reviewer (and editor) to assess whether it is possible for the authors to address these in a manner that would ultimately lead to a high-quality publication. If a reviewer identifies a methodological or other problem that cannot be remedied, noting this fatal flaw will help the reviewer make a recommendation, and the editor make a decision, about whether or not to accept the manuscript for publication.

Sometimes, reviewers have the option to write confidential comments to the editor. Reviewers may want to use this opportunity to suggest that a manuscript be referred for a statistical consultation or to suggest that someone with specific expertise in certain methodologies (e.g., qualitative research) or additional training (e.g., statistics) review the manuscript as well.

Once the reviews, including any confidential comments and the decision recommendation, are available, the editor must make a publication decision. The complexity of making that decision is the topic of Chapter 4, "Publication Decision."

In addition to the process of deciding how and whether to solicit reviews for a manuscript, other procedures can support or round out the general review process (see Figure 1). For example, the editor and editorial staff must decide whether the reviewers' comments will go to the authors and, if they do, whether the editorial staff will edit them. Two advantages to sending them unedited are that doing so reduces the summary and distillation process for the staff and gives the authors a sense that they are getting as much feedback as possible. On the other hand, the comments from multiple reviewers may conflict or reviewers may suggest a direction the editor does not want the authors to pursue. And sometimes, reviews contain unhelpful or even inappropriate comments. There is no right or wrong policy—deciding how to handle review comments is simply another matter the editor and editorial staff must consider.

Other issues about which editors (and their editorial boards) make decisions include the following:

- whether to inform reviewers of the publication decision, either by copying them on the communication to the authors or by sending some other kind of summary;
- whether to involve the same reviewers in assessing a second version of a manuscript; and
- whether to solicit reviewers' opinions about all or parts of the review process.

Obviously, the editor can address these issues in a multitude of ways to create a review process that is unique for each journal.

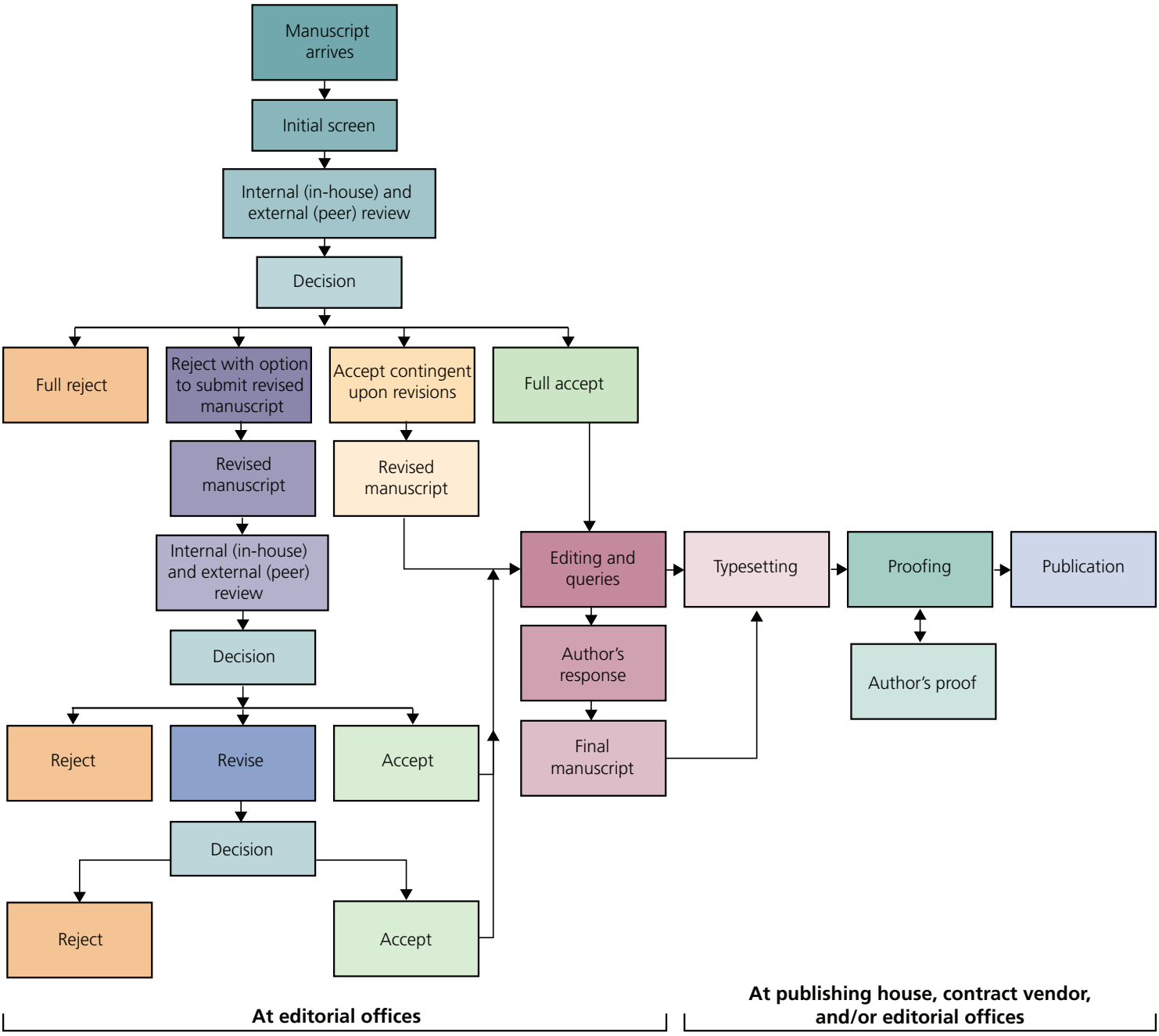


**Figure 1. The review process.**

*Acknowledgments:* The authors would like to acknowledge Louis Pangaro and Ann Steinecke as contributors to previous versions of this work.
*Funding/Support:* None reported.
*Other disclosures:* None reported.

## References

1. Bordage G. Reasons reviewers reject and accept manuscripts: The strengths and weaknesses in medical education reports. Acad Med. 2001;76:889–896.
2. Yankauer A. Who are the peer reviewers and how much do they review? JAMA. 1990;263:1338–1340.
3. Weller AC. Editorial peer review in US medical journals. JAMA. 1990;263:1344–1347.
4. Hargens LL. Variation in journal peer review systems: Possible causes and consequences. JAMA. 1990;263:1348–1352.
5. Jefferson T, Alderson P, Wager E, Davidoff F. Effects of editorial peer review: A systematic review. JAMA. 2002;387:2784–2786.
6. Godlee F, Gale CR, Martyn CN. Effect on the quality of peer review of blinding reviewers and asking them to sign their reports: A randomized controlled trial. JAMA. 1998;280:237–240.
7. van Rooyen S, Godlee F, Evans S, Black N, Smith R. Effect of open peer review on quality of reviews and on reviewers' recommendations: A randomised trial. BMJ. 1999;318:23–27.
8. van Rooyen S, Delamothe T, Evans SJ. Effect on peer review of telling reviewers that their signed reviews might be posted on the web: Randomised controlled trial. BMJ. 2010;341:c5729.
9. Regehr G, Bordage G. To blind or not to blind? What authors and reviewers prefer. Med Educ. 2006;40:832–839.
10. McNutt RA, Evans AT, Fletcher RH, Fletcher SW. The effects of blinding on the quality of peer review. JAMA. 1990;263:1371–1376.
11. Fisher M, Friedman SB, Strauss B. The effects of blinding on acceptance of research papers by peer review. JAMA. 1994;272:143–146.
12. Jadad AR, Moore A, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? Control Clin Trials. 1996;17:1–12.
13. Justice AC, Cho MK, Winker MA, Berlin JA, Rennie D. Does masking author identity improve peer review quality? A randomized controlled trial. JAMA. 1998;280:240–242.
14. van Rooyen S, Godlee F, Evans S, Smith R, Black N. Effect of blinding and unmasking on the quality of peer review: A randomized trial. *JAMA*. 1998;280:234–237.
15. Davidoff F. Masking, blinding, and peer review: The blind leading the blinded. Ann Intern Med. 1998;128:66–68.
16. Ross JS, Gross CP, Desai MM, et al. Effect of blinded peer review on abstract acceptance. JAMA. 2006;295:1675–1680.
17. Katz DS, Proto AV, Olmsted WW. Incidence and nature of unblinding by authors: Our experience at two radiology journals with double-blinded peer review policies. Am J Roentgenol. 2002;179:1415–1417.
18. Lock S. A Difficult Balance: Editorial Peer Review in Medicine. Philadelphia, PA: ISI Press; 1985: 122–123.
19. Cho MK, Justice AC, Winker MA, et al. Masking author identity in peer review: What factors influence masking success? JAMA. 1998;280:243–245.
20. Rosenblatt A, Kirk SA. Recognition of authors in blind review of manuscripts. J Soc Service Res. 1980;3:383–394.

# Chapter 2
# Selection and Qualities of Reviewers

**C. Jessica Dine, MD, MSHPR, Addeane S. Caelleigh, MA, and Judy A. Shea, PhD**

**C. J. Dine** is assistant professor of medicine, Division of Allergy, Pulmonary, and Critical Care, and associate program director, Internal Medicine Residency Program, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

**A. S. Caelleigh** was the Liaison Committee of Medical Education coordinator, Office of Medical Education, and she is now a visiting faculty member, University of Virginia School of Medicine, Charlottesville, Virginia.

**J. A. Shea** is professor of medicine and associate dean, Medical Education Research and Assessment, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

The quality and usefulness of a journal rest on the quality of the research submitted, the reviewers' critiques, and the editor's judgment. The reviewer plays a key role in the publication of excellent scholarship, and reviewing provides an excellent opportunity for the reviewer to contribute to his or her field. One of the editor's primary responsibilities is to create and maintain a pool of high-quality, productive reviewers, so editors are concerned with each reviewer's knowledge, judgment, constructiveness, ability to write clearly, and willingness to work within the journal's guidelines.

Journal editors retain responsibility for recruiting and organizing their reviewers. The editor and journal staff identify possible reviewers through contacts at professional meetings, personal acquaintances, editorial boards, literature searches, society membership lists, and manuscript bibliographies.[1] Occasionally, potential reviewers seek out journals, sending unsolicited offers to help. Some journals encourage authors to suggest reviewers for their manuscripts. Studies suggest that the *quality* of the review does not differ between author- or editor-recommended reviewers; however, reviewers suggested by authors tend to provide more favorable recommendations for publication.[2,3] Editorial offices typically track how often each reviewer submits reviews, and they take various approaches to monitoring their reviewers' performances, often rating each review and noting the reviewer's timeliness.

The editor or editorial staff selects the reviewers for a particular manuscript, taking into account the reviewers' expertise and availability. The selection often begins with matching the manuscript topic to a reviewer's area of expertise, and it may extend beyond this simple alignment to ensure that, for example, technical expertise is augmented by particular professional experiences (e.g., having a particular role or position, such as dean or chief financial officer, within an organization). Given the increasing breadth of work submitted in the field of health professions education, matching a manuscript with a potential reviewer is becoming increasingly difficult. Typically, the invitation to review, sent from the editorial office, comes with an abstract of the submitted manuscript to aid the reviewer in assessing whether his or her expertise matches the topic of the work. For some manuscripts, both content and methodological reviewers are needed.

To be most useful to a journal, a reviewer's expertise must be complemented by sound judgment; therefore, the journal editor is concerned about whether the reviewer can make balanced decisions, keep a sense of proportion when assessing the strengths and weaknesses of a research project, apply appropriate standards (such as those outlined in this reviewer guide), and give definite, well-supported opinions. To give these opinions, the reviewer must be able to express complex ideas clearly—and, as much as possible, concisely—for both the editor and the authors, whose needs are not the same. Indeed, many journals put considerable emphasis on reviewers' ability and willingness to give useful feedback to authors. Further, journals expect reviewers to give their opinions (even, when necessary, stern judgments) in a collegial spirit accompanied by concrete suggestions for improvement. Reviewers should aim to provide detailed and constructive criticism to help improve the manuscript and its potential for publication.

The journal editor and staff look for different attributes and qualities in reviewers depending on the nature of the journal, the role of the reviewer, and the editor's or sponsoring society's policies.[4] Unfortunately, little empirical evidence is available to determine the profile of the ideal or even the good reviewer. Studies conducted by clinical journals to assess the characteristics of reviewers who produce high-quality reviews report conflicting findings.[5–9] Although the authors of these studies define "quality" slightly differently, the core categories they used in assessing reviews and reviewers are very similar.

In the earliest study (in 1985), Stossel found that reviewers with lower academic or professional status produced better critiques of manuscripts than the higher-status reviewers did—and were less likely to decline to review.[5] Several years later, Evans and colleagues reported that the reviewers who produced the best reviews were less than 40 years old, known to the editors, and from highly respected (highly rated) academic institutions.[6] A later study confirmed the positive correlations between, on the one hand, higher-quality reviews and, on the other, younger age and affiliation with an academic institution.[7] This study also found no correlation between the quality of the review and the reviewer's gender, academic rank, or subspecialty.[7] Black and colleagues found that ratings of reviewer quality increased with the time spent on a review—up to three hours but not beyond that.[8] According to their study, reviewers' younger age was an independent predictor of the editors' (but not the authors') rating of a review as high quality, while reviews by members of an editorial board were rated of poorer quality by authors (but not by editors).[8]

Finally, the only significant factor associated with higher ratings of reviews by both editors and authors was whether the reviewer had training in epidemiology or statistics.[8] A later study suggested, however, that although there was significant variation in the quality of reviews provided even among experienced reviewers, no type of formal training (including formal training in critical appraisal or statistics) or previous experience (such as being a principal investigator of a grant) could predict a higher-quality review.[9] The variety of characteristics, or lack thereof, that can predict high-quality reviewers led Fletcher and Fletcher[10] to conclude "that editors should not have fixed views of what kinds of reviewers might return good reviews. Because the characteristics of good reviews might vary from one setting to another, it seems editors should continue the common practice of grading their own reviewers but recognize that this is an imperfect predictor of their future performance."

Fletcher and Fletcher's caution aside, previous experience with a reviewer may be the most useful guide. As mentioned, many journal editors evaluate the reviews they receive and refer to the quality ratings of past reviews when considering a reviewer for a manuscript. They also use the ratings when deciding whether to retain a reviewer. Many editors use simple rating systems that serve primarily as rough triage systems, but some are developing more sophisticated, even standardized, instruments to assess reviewers' work.[11–13] Editors may study reviewers' effectiveness in specific aspects of review, as well as in their overall effectiveness throughout the review process.

Baxt and colleagues, for example, examined how well reviewers at a clinical journal could identify major and minor flaws in a manuscript.[14] Callaham and colleagues studied the reliability of the editors' subjective ratings of review quality and found them to be moderately reliable and correlated with the reviewers' abilities to report manuscript flaws.[15] Generally, editors and journals create and retain reviewer assessments for internal purposes only, but sometimes reviewers request an evaluation of their review as well. More and more journals are sending a copy of the final decision and all reviewers' comments to the reviewers; reviewers can then compare their comments with those of the other reviewers.

Journals are aware of the intrinsic, unconscious biases in the review process and may blind the reviewer to the identity of the authors in an attempt to overcome some of them (see Chapter 1).[11,16,17] These biases may relate to intellectual positions or ideology, preferences for positive (or hypothesis-confirming) research outcomes, and personal social or political convictions. They may also be the oft-cited prejudices having to do with ethnicity, nationality, gender, and status. Journals ask reviewers to disclose any potential conflicts of interest that may interfere with the quality of the review.

In summary, the reviewer is crucial in selecting high-quality manuscripts for publication and in providing constructive feedback to help improve submitted manuscripts. Although there is no consensus on the qualities of an ideal reviewer, editors do assess reviews and reviewers in an effort to produce high-quality scholarship.

*Acknowledgments:* The current authors would like to acknowledge Gary Penn as a contributor to a previous version of this work.
*Funding/Support:* None reported.
*Other disclosures:* None reported.

## References

1. Weller AC. Editorial peer review in US medical journals. JAMA. 1990;263:1344–1347.
2. Schroter S, Tite L, Hutchings A, Black N. Differences in review quality and recommendations for publication between peer reviewers suggested by authors or by editors. JAMA. 2006;295:314–317.
3. Wager E, Parkin EC, Tamber PS. Are reviewers suggested by authors as good as those chosen by editors? Results of a rater-blinded, retrospective study. BMC Med. 2006;4:13.
4. Polak JF. The role of the manuscript reviewer in the peer review process. Am J Roentgenol. 1995;165:685–688.
5. Stossel TP. Reviewer status and review quality: Experience of the Journal of Clinical Investigation. N Engl J Med. 1985;312:658–659.
6. Evans AT, McNutt RA, Fletcher SW, Fletcher RH. The characteristics of peer reviewers who produce good-quality reviews. J Gen Intern Med. 1993;8:422–428.
7. Kliewer MA, Freed KS, DeLong DM, Pickhardt PJ, Provenzale JM. Reviewing the reviewers. Comparison of review quality and reviewer characteristics at *the American Journal of Roentgenology*. Am J Roentgenol. 2005;184:1731–1735.
8. Black N, van Rooyen S, Godlee F, Smith R, Evans S. What makes a good reviewer and a good review for a general medical journal? JAMA. 1998;280:231–233.
9. Callaham ML, Tercier J. The relationship of previous training and experience of journal peer reviewers to subsequent review quality. PLoS Med. 2007;4:e40.
10. Fletcher RH, Fletcher SW. The effectiveness of editorial peer review. In: Godlee F, Jefferson T, eds. Peer Review in Health Sciences. London: BMJ Books; 1999:45–56.
11. McNutt RA, Evans AT, Fletcher RH, Fletcher SW. The effects of blinding on the quality of peer review: A randomized trial. JAMA. 1990;263:1371–1376.
12. Feurer ID, Becker GJ, Picus D, Ramirez E, Darcy MD, Hicks ME. Evaluating peer reviews: Pilot testing of a grading instrument. JAMA. 1994;272:98–100.
13. van Rooyen S, Black N, Goodlee F. Development of the review quality instrument (RQI) for assessing peer reviews of manuscripts. J Clin Epidemiol. 1999;52:625–629.
14. Baxt WG, Waeckerle JF, Berlin JA, Callaham ML. Who reviews the reviewers? Feasibility of using a fictitious manuscript to evaluate peer reviewer performance. Ann Emerg Med. 1998;32(3 Pt 1):310–317.
15. Callaham ML, Baxt WG, Waeckerle JD Wears RL. Reliability of editors' subjective quality ratings of peer reviews of manuscripts. JAMA. 1998;280:229–231.
16. Gilbert JR, Williams ES, Lundberg GD. Is there gender bias in JAMA's peer review process? JAMA. 1994;272:139–142.
17. Ernst E, Resch KL. Reviewer bias: A blinded experimental study. J Lab Clin Med. 1994;124:178–182.

# Chapter 3
# Review Forms and Reviewer Comments

**C. Jessica Dine, MD, MSHPR, Addeane S. Caelleigh, MA, and Judy A. Shea, PhD**

**C. J. Dine** is assistant professor of medicine, Division of Allergy, Pulmonary, and Critical Care, and associate program director, Internal Medicine Residency Program, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

**A. S. Caelleigh** was the Liaison Committee of Medical Education coordinator, Office of Medical Education, and she is now a visiting faculty member, University of Virginia School of Medicine, Charlottesville, Virginia.

**J. A. Shea** is professor of medicine and associate dean, Medical Education Research and Assessment, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

In this chapter, we discuss the different review forms used by many journals to help focus the reviewer and improve the efficiency of peer review. Our examples are from research journals in the health professions, but they could have been taken from any research field in the biosciences and social sciences.

A brief review of the process discussed in Chapter 1, "Review Process," may be helpful here. Typically, the editorial office invites a potential reviewer to review a manuscript, and the invitation often includes the manuscript's abstract. A reviewer should judge whether the topic falls within his or her area of expertise and whether any potential conflicts of interest exist. *Academic Medicine*, for example, advises that a "conflict of interest exists when professional judgment concerning a primary interest may be influenced by secondary interests."[1] A conflict of interest does not necessarily disqualify someone from reviewing, but the potential reviewer should, a priori, discuss any questions or concerns with the editor or editorial staff who will determine whether a conflict of interest precludes the reviewer from accepting the invitation. Finally, the potential reviewer should look at the deadline to ensure he or she is able to provide a timely assessment of the manuscript in question.

Once the potential reviewer accepts an invitation, the editorial staff will provide access to a review form as well as the manuscript to review. Almost all journals now ask reviewers to submit their recommendations online. Reviewers may find that looking over the review form before evaluating the manuscript is helpful; doing so will likely help them anticipate what specific information the journal requires or wants.

No standard review form is used across journals, not even among the journals in any scientific field; further, we are not aware of any studies that demonstrate the best format for review forms.

The features of review forms vary widely depending on the type of information each editor seeks from reviewers—some request narrative comments alone,[2] while others combine checklists, global ratings, and narrative comments.[3] Many journals ask the reviewer to provide both quantitative and qualitative assessments of the reviewed manuscript (discussed further below).

Journals use their review forms—no matter the format—to solicit the same general categories of information: the importance of the topic discussed, the quality of the research as presented, the justifiability of the conclusions, and the appropriateness of the manuscript for publication in the journal. The review form often includes sections for the reviewer to provide specific and constructive feedback to the authors, comments to the editor not visible to the authors if needed, and a recommendation on whether or not to accept the manuscript for publication with or without required revisions. The specific questions and their order may vary among journals' review forms, but the overall content is similar, as is the goal: to help the reviewer provide a focused, constructive, and useful review.

Different response formats (e.g., global ratings, checklists, narrative comments) offer the reviewer, the editor, and authors different advantages, so they are often used in combination. The review form typically starts with a list of specific questions. These questions direct the reviewer to evaluate specific aspects of a manuscript, as in this example from *Medical Education*, which guides reviewers to consider a research report's Results section[4]:

- Are the results clearly presented?
- Are they consistent with both the methods used and the problem the authors are trying to address?
- Do they yield a clear answer to the research question?

Some journals present questions for the reviewer to answer using a dichotomous (yes/no) scale, and many allow the reviewer to leave comments that expound on the answers. Other journals use rating scales (e.g., unacceptable, poor, satisfactory, excellent). One major benefit of a checklist is that it identifies the key elements the reviewer should judge for soundness, particularly those having to do with potential research methods. Further, the reviewer's evaluation of these key elements often provides a source of specific feedback for the authors.

The risk that reviewers will not identify methodological flaws in their narrative comments alone is high,[5] and the risk that they will not identify such flaws in their checklist responses is probably just as great.[6] However, in combination with an opportunity to discuss a manuscript's strengths and weaknesses in narrative form, checklists can be particularly valuable for identifying significant problems within a manuscript.

Most journals ask the reviewer to provide some form of narrative comments after the checklist. These are typically elicited through open-ended questions or a section for free response. Narrative

comments allow the reviewer to respond with the greatest level of detail. *Academic Medicine*'s review form, for example, asks the reviewer to respond to a series of brief, open-ended items about the manuscript:

> Please provide substantive comments that will help the author(s) strengthen this manuscript.
>
> Please include feedback on the contribution to the literature, generalizability, and whether the conclusions are justified. For papers with quantitative or qualitative findings, please also provide feedback on the method, statistical or other analysis, reporting of the results, and limitations.

These requests free the reviewer from preformulated responses and allow detailed critique. Specific, detailed, and constructive comments are the most helpful for the editor, who must decide whether to accept a manuscript, and for the authors, who may seek to improve their work. A high-quality review typically includes an overall summary of the reviewer's impression of the manuscript and how it adds to the current literature. Often, such a summary is followed by a description of the major and minor issues the reviewer has identified in the manuscript.[7,8]

As mentioned, most review forms typically include a space for the reviewer to provide narrative comments to the editor that will not be visible to the authors. The reviewer can use this space to

- recommend additional review by someone with specific expertise,
- make specific comments on the quality of the manuscript,
- provide opinions about the relevance or significance of the work, or
- raise potential ethical concerns.

The reviewer can also use this space to give the editor more nuanced and detailed information and to explain the severity of any problems detected in the manuscript, along with the likelihood that the authors can address the problems through a revision. The reviewer should ensure that his or her confidential comments to the editor do not contradict the comments directed toward the authors and that they provide information that is relevant only to the editor.

Review forms typically begin (or end) with a global rating score that asks the reviewer to make a recommendation for publication, such as *Academic Medicine*'s current four-point scale[9]:

1. Accept;
2. Reconsider after Minor Revisions;
3. Reconsider after Major Revisions; and
4. Reject.

These scales are efficient, and they enable the editor to quickly assess the potential of a manuscript; however, global rating scores alone are not sufficient to allow the editor to make a judgment. Additional information is needed to provide the editor with a nuanced discussion of the manuscript's strengths and weaknesses and the authors with constructive feedback. For these reasons, global ratings are rarely used in isolation from other question formats.

As an alternative to structured review forms, some journals simply give reviewers an opportunity to make any comments they wish about the manuscript. The *British Medical Journal*, for example, does not provide a formal review form. Instead, the reviewer receives only a copy of the instructions for review; he or she is expected to return their comments on stationery.[10] The instructions include a list of questions to consider, such as, "Is the article important?" and "Will the article add enough to existing knowledge?"

While free-text comments (whether alone or to complement quantitative critique questions) allow the reviewer to provide a great deal of detail, they also present drawbacks for the reviewer. They require more time and effort to complete than do global ratings and checklists, and they do not provide a guide for the reviewer.

Reviewers should be aware of two ways that review forms may inadvertently affect their reviews:

1. When completing a review using a form that combines response formats, reviewers may inadvertently contradict what they have indicated on the checklist with what they have written in their comments. For example, a reviewer's global ratings for a manuscript may be quite positive, whereas his or her narrative comments may illuminate a methodologic problem that cannot be overcome.
2. The content and directions of the questions included on the review form may bias the reviewer. For example, the absence of a question about methodology may imply that the Methods section is not important to the editorial staff.

Because of the potential influence of review forms on the reviewer, they are usually carefully designed. There is ongoing debate about whether they should be tested for inter-rater reliability. Many studies have demonstrated a lack of agreement between reviewers' recommendations; correlations average between 0.2 and 0.3.[7,8] Editors of clinical specialty journals, especially those who rely most heavily on reviewers' recommendations, value high inter-rater reliability,[11,12] and tested review instruments—with high reliability and validity—have been described.[13] It is not known whether these tested review instruments, designed specifically for clinical manuscripts, offer the same reliability for health professions education manuscripts.

Although review forms for clinical and health journals share many characteristics and aims of forms for health professions education journals, particularly for assessing study design and methods, editors in the latter fields may prefer forms that allow competing viewpoints. Some research suggests that competing or contradictory reviews may actually strengthen the quality of the decision-making process.[14] Specifically, Eckberg suggests that differences in reviewers' comments may reflect different biases, so higher reviewer reliability may mean that the biases of the reviewers are similar. In addition, he suggests that when the expertise of reviewers differs, the reviewers may detect different flaws within the methods or interpretation of the submitted manuscript.[14]

In summary, the editor uses the information from the reviewer to help decide whether to publish or reject the article, whether to ask the authors for revisions, and whether to send the manuscript for further review. The authors use the same reviewer-provided information to better understand the publication decision, to gain insight into the strengths and weaknesses of their manuscript, and as a guide to revise the manuscript. The review form is thus a tool for the reviewer to use to communicate information for different uses by the editor and authors.

*Acknowledgments:* The current authors would like to acknowledge Ann Steinecke as a contributor to previous versions of this work.
*Funding/Support:* None reported.
*Other disclosures:* None reported.

## References

1. Academic Medicine. Conflict of Interest: Authorial. Updated September 2014. http://journals.lww.com/academicmedicine/Pages/InstructionsforAuthors.aspx#conflictofinterest. Accessed April 13, 2015.
2. Bligh J. What happens to manuscripts submitted to the journal? Med Educ. 1998;32:567–570.
3. Nyleena M, Riis P, Karlsson Y. Multiple blinded reviews of the same two manuscripts: Effects of referee characteristics and publication language. JAMA. 1994;272:149–151.
4. Medical Education. Guidelines for reviewers: Results. 2015. http://onlinelibrary.wiley.com/journal/10.1111/%28ISSN%291365-2923/homepage/guidelines_for_reviewers.htm. Accessed March 19, 2015.
5. Bordage G. Reasons reviewers reject and accept manuscripts: The strengths and weaknesses in medical education reports. Acad Med. 2001;76:889–896.
6. Jefferson T, Alderson P, Wager E, Davidoff F. Effects of editorial peer review: A systematic review. JAMA. 2002;287:2784–2786.
7. Lovejoy TI, Revenson TA, France CR. Reviewing manuscripts for peer-review journals: A primer for novice and seasoned reviewers. Ann Behav Med. 2011;42:1–13.
8. Cullen DJ, Macaulay A. Consistency between peer reviewers for a clinical specialty journal. Acad Med. 1992;67:856–859.
9. Academic Medicine. For Reviewers: *Academic Medicine* Guide to Reviewer Recommendations. 2015. http://journals.lww.com/academicmedicine/Documents/Academic%20Medicine%20Guide%20to%20Reviewer%20Recommendations.pdf. Accessed April 13, 2015.
10. British Medical Journal. Guidance for peer reviewers. 2015. http://www.bmj.com/about-bmj/resources-reviewers/guidance-peer-reviewers. Accessed April 13, 2015.
11. Kravtiz RL, Franks P, Feldman MD, et al. Editorial peer reviewers' recommendations at a general medical journal: Are they reliable and do editors care? PLoS One. 2010;5:e10072.
12. Cicchetti DV. The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. Behav Brain Sci. 1991;14:119–135.
13. Cho MK, Bero LA. Instruments for assessing the quality of drug studies published in the medical literature. JAMA. 1994;272:101–104.
14. Eckberg DL. When nonreliability of reviews indicates solid science. Behav Brain Sci. 1991;14:145–146.

# Chapter 4
# Publication Decision

**C. Jessica Dine, MD, MSHPR, Addeane S. Caelleigh, MA, and Judy A. Shea, PhD**

**C. J. Dine** is assistant professor of medicine, Division of Allergy, Pulmonary, and Critical Care, and associate program director, Internal Medicine Residency Program, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

**A. S. Caelleigh** was the Liaison Committee of Medical Education coordinator, Office of Medical Education, and she is now a visiting faculty member, University of Virginia School of Medicine, Charlottesville, Virginia.

**J. A. Shea** is professor of medicine and associate dean, Medical Education Research and Assessment, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

The publication decision is the last step of the review process. Journals vary in the process they use for making publication decisions, but typically, they publish their process on their website where authors can also find instructions on how to prepare and submit manuscripts. Many journal websites also include information on appealing the final decision.

The principal actor in the publication decision process is the editor. Generally, editors consider reviewers' comments *along with other factors* when they decide how to proceed with a manuscript. Most do not defer to reviewers' comments for the publication decision,[1] although one survey of North American medical journal editors showed that almost half of the respondents did base their decisions solely on reviewers' comments.[2,3] Typically, journal editors rely on reviewers' comments only for advice and reserve the final decision for themselves.[3] Using reviewers' comments strictly as helpful information is a major area of consensus among health professions journals and among the leading journals of most fields and disciplines. Reviewers make *recommendations* to editors.

Here is a good place to clarify the use of the term *editor* as a singular noun. While this term applies generally to the organizational and decision-making structure at some journals, the exact organization of a journal may vary, and several people may have decision-making responsibilities. For example, each section of a journal may have a separate editor who functions and makes decisions independently. Other journals are structured such that associate editors are assigned their "own" manuscripts. They make publication recommendations that are then discussed by the larger group of either many associate editors or multiple associate editors and an editor-in-chief. Often, in such a structure, the group members make the final publication decision together.

The decision about whether or not to publish a manuscript that has been sent out for review is rarely obvious. Occasionally, reviewers agree that a manuscript should not be considered further—that it clearly should not be published. Often, such an agreement will precede an editor's decision to reject a manuscript, and if the adjoining reviews are well written, they will help elucidate this decision for the authors. Very rarely, reviewers unanimously agree that a manuscript should be published directly as written. Most often, reviewers recommend some middle strategy (e.g., revise), or they offer conflicting recommendations for and against publication.

Regardless of the agreement (or lack thereof) among reviewers, the editor must make the final publication decision. Reviewers can help by submitting detailed and specific concerns that include the severity of the identified flaws (e.g., major or minor concerns). A reviewer's comments can help the editor understand which revisions to recommend prior to publication or whether the concerns can even be mediated through revisions at all.

The editor often knows the general field well but is not (and cannot be expected to be) a subject expert for every manuscript. How, then, does he or she make a decision after reading the reviews? The editor must consider each reviewer's opinions, suggestions, and comments—balanced with the collective input from several reviewers, his or her own thoughts, and additional factors (see below). The editor is not likely to give equal weight to all reviewers' comments; for example, he or she may weigh those from known experts more heavily. Also, reviews that offer constructive feedback, are more than a couple of sentences long, identify original issues or flaws, and/or use details from the manuscript to support their comments may receive special attention. In short, the editor is looking for a review that is thoughtful, helpful, and specific.

The editor must balance the reviewer's suggestions with many other factors. A manuscript's originality (or, at the opposite end of the spectrum, its redundancy) is one very important consideration. Depending on the manuscript's topic, either pole may be desirable. The reviewer can help the editor assess the balance between new and superfluous by offering his or her opinion on whether the manuscript adds to the existing literature and whether publishing the findings would aid the larger community (i.e., the journal's readers). Other issues that may factor in the decision include length (sometimes short is good; sometimes long is better), the relevance of the topic, and what the journal has published recently. An editor may also decide to reject a manuscript based on the concern that the amount of editing necessary would require substantial staff time and effort.

These factors are often secondary to the external reviews, which focus primarily on quality. These additional concerns may be especially important either to justify a publication decision or to direct future revisions. For example, if an excellent manuscript is rejected because it is too similar to one that was just published, the journal staff or editor may take steps to ensure that the authors

hear both points: the manuscript is excellent but redundant. An editor can also work with authors to shorten or expand a manuscript.

The decision-making process is highly complex, multifactorial, and unique for each paper. It is subjective, but it is neither capricious nor uninformed. In many ways, a manuscript is in competition with all the other good manuscripts being processed at the same time. Often, the editor may select only a few from a large field. This point is very important, yet authors may forget or overlook it: a manuscript that is generally acceptable may be rejected primarily because it is not quite as "strong" (e.g., interesting, methodologically sound, different) as current competitors. Beyond stating how many reviewers most journals typically request and making public the general review criteria journals use, there is little about the process that can be succinctly generalized—except to say that editors take reviewers' suggestions seriously.

Reviewers' comments are almost always the most important component in an editor's decision-making process. If the content and direction (e.g., mostly in favor of publication) of the comments agree, the editor is likely to follow the reviewers' recommendations. Unfortunately, agreement is rare, and editors must rely on their own experience and judgment in weighing the reviewers' comments.

The final step in the decision-making process is conveying (usually via email) the decision to the authors. Several aspects of this step are relevant to the reviewer, and, again, journals differ in how they handle them:

- Does the editorial office forward the reviewer's comments to the authors? If so, are they sent as received, or does the editor pick and choose among the comments and compile them in the decision letter?
- Does the reviewer see a copy of the letter sent to the authors?
- Do reviewers see one another's comments?

In some instances, the review process is open or unblinded such that the authors and reviewers are aware of one another's identities; other times, everyone is anonymous; and still other times, only the identities of the reviewers are hidden (see Chapter 1). The editor is usually willing to communicate, either in writing or verbally, the important elements in any particular decision to the authors.

In summary, while journals have well-defined policies and processes designed to support efficient and fair publication decisions, the final decision about each manuscript is made within a larger context.

*Funding/Support:* None reported.
*Other disclosures:* None reported.

## References

1. Olson CM. Peer review of the biomedical literature. Am J Emerg Med. 1990;8:356–358.
2. Wilkes MD, Kravitz RL. Policies, practices, and attitudes of North American medical journal editors. J Gen Intern Med. 1995;10:443–450.
3. Kravtiz RL, Franks P, Feldman MD, Gerrity M, Byrne C, Tierney WM. Editorial peer reviewers' recommendations at a general medical journal: Are they reliable and do editors care? PLoS One. 2010;5:e10072.

# Chapter 5
# Manuscript Revision and Final Editing

**C. Jessica Dine, MD, MSHPR, Addeane S. Caelleigh, MA, and Judy A. Shea, PhD**

**C. J. Dine** is assistant professor of medicine, Division of Allergy, Pulmonary, and Critical Care, and associate program director, Internal Medicine Residency Program, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

**A. S. Caelleigh** was the Liaison Committee of Medical Education coordinator, Office of Medical Education, and she is now a visiting faculty member, University of Virginia School of Medicine, Charlottesville, Virginia.

**J. A. Shea** is professor of medicine and associate dean, Medical Education Research and Assessment, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

Manuscript revisions typically occur at two points. Some journals ask authors to make revisions as a condition of having the manuscript reconsidered or accepted, and some require changes as part of the final editing process before the manuscript is published. Most require revisions at both points. Some journals have policies—and the resources—to help authors make the revisions, but many do not. And a few, including *Academic Medicine*, provide substantive editing, while others do only the most basic copyediting. These processes are, in part, the link between the peer reviewer's comments and the editor's decision. They also represent, in part, the journal's contribution to improving manuscripts.

In deciding whether to publish a manuscript, the editor may accept it as submitted, reject it, or offer authors the option of revising their manuscript before a final decision is made. The editor has a range of reasons for offering revisions and for choosing which revisions to request or require. In making these decisions, the editor must maintain a balance among the appropriate use of the reviewer as adviser, the originality of the submitted manuscript, and the rights and responsibilities of the editor. In other words, the editor

- must not expect, or ask, reviewers to provide all the revisions required to ready a manuscript for publication;
- may not request revisions that change the authors' original intention or work; and
- must work to protect the standards of the journal and of publication ethics.

The editor's concerns include factors intrinsic to the study described in the manuscript (e.g., methodologic standards) and others that are outside the authors' sphere, such as the focus of the journal, the nature of its readership, reports recently published in the editor's and other journals, and any special interests that the editor or journal may be emphasizing at a particular time.

The reviewer can help the editor determine if and when to ask for revisions and what revisions to request by

- providing detailed but constructive criticism of the manuscript,
- prioritizing suggestions for revisions (e.g., distinguishing between major and minor recommendations),
- explicitly discussing the severity of any identified flaws, and
- judging the likelihood that the authors can address these comments through revisions.[1]

For many journals, the reviewer also has the opportunity to make confidential comments to the editor that will not be forwarded to the authors. Reviewers may want to use the "confidential comments" space to suggest that a statistician or an expert in a particular topic review the manuscript before it is accepted. Other confidential feedback may address any overlap with prior work, the appropriateness of the work for the journal's audience, ethical concerns, and the length of the manuscript balanced with the importance of its contribution.

The editor uses part or all of the reviewers' comments in giving instructions to the authors about revisions.[2] If an editor wants the authors to address a concern one of the reviewers raised in confidential comments, the comment will likely appear with other instructions from the editor or editorial office. These instructions may specify which revisions are required, which are encouraged but not required, and which are entirely optional. The editor should resolve any conflict between different reviewers' comments so that the authors will know which comments they should follow in making particular revisions.[3] The editor should also outline clearly what will happen after the journal receives the revised manuscript. For example, will it go out for further review, or will the next decision be made by the editor alone? Finally, the letter should set out a timeline or a deadline for the authors' reply.

After a manuscript is officially accepted, final editing is almost always required before publication.[2] As mentioned, the depth and extent of this editing vary widely from journal to journal. Requested edits may be as simple as copyediting (e.g., correcting spelling or punctuation, or formatting the manuscript to align with journal style), or they may be more involved (e.g., justifying information in the abstract with information in the report or checking the accuracy of data and references). Some journal offices engage in substantive editing, including incorporating suggested revisions proposed by reviewers, clarifying content for readers, or restructuring content for consistency. Editorial offices will typically outline requested edits in a letter to the authors (either at the time of acceptance or later, when the manuscript is scheduled and ready for final editing).

In summary, the information the reviewer relays to the editor is vital; the most useful reviews may affect the decision to publish the manuscript, shape the revisions requested, and, in turn, shape the final report.

*Funding/Support:* None reported.
*Other disclosures:* None reported.

## References

1. Benos DJ, Kirk KL, Hall JE. How to review a paper. Adv Physiol Educ. 2003;27:47–52.
2. Purcell GP, Donovan SL, Davidoff F. Changes to manuscripts during the editing process: Characterizing the evolution of a clinical paper. JAMA. 1998;280:227–228.
3. Shattell MM, Chinn P, Thomas SP, Cowling WR 3rd. Authors' and editors' perspectives on peer review quality in three scholarly nursing journals. J Nurs Scholarsh. 2010;42:58–65.

# Part 2

# Chapter 6
# Problem Statement, Conceptual Framework, and Research Question

**C. Jessica Dine, MD, MSHPR, William C. McGaghie, PhD, Georges Bordage, MD, PhD, and Judy A. Shea, PhD**

**C. J. Dine** is assistant professor of medicine, Division of Allergy, Pulmonary, and Critical Care, and associate program director, Internal Medicine Residency Program, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

**W. C. McGaghie** was director, Ralph P. Leischner, Jr. MD Institute for Medical Education, Loyola University Chicago Stritch School of Medicine, Maywood, Illinois, at the time this work was done. He is now professor of medical education, Northwestern University Feinberg School of Medicine, Chicago, Illinois.

**G. Bordage** is professor of medical education, Department of Medical Education, University of Illinois at Chicago, Chicago, Illinois.

**J. A. Shea** is professor of medicine and associate dean, Medical Education Research and Assessment, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

## Review Criteria

1. ***The introduction builds a logical case and provides context for the problem statement.***
2. ***The problem statement is clear and well articulated.***
3. ***The conceptual framework is explicit and justified.***
4. ***The research purpose and/or question (as well as the research hypothesis, where applicable) is clearly stated.***
5. ***The constructs being investigated are clearly identified and presented.***

## Issues and examples related to criteria

A scholarly manuscript starts with an Introduction that tells a story, even if the Introduction is not explicitly labeled as such. The Introduction orients the reader to the topic of the report, moving from broad concepts to more specific ideas.[1,2] The Introduction should convince the reader—and all the more, the reviewer—that the authors have thought the topic through and identified a "researchable" problem. The Introduction should move logically from the known to the unknown.

The aspects of each Introduction (including length, complexity, and organization) vary depending on the type of study being reported, the traditions of the research community or discipline in which it is based, and the style and tradition of the journal receiving the manuscript. It is helpful for the reviewer to evaluate the Introduction by thinking about its overall purpose and its individual components: problem statement, conceptual framework, and research question. Two related chapters, "Reference to the Literature and Documentation" and "Relevance," follow this one.

### Problem statement

The Introduction to a research manuscript articulates a problem statement. This essential element conveys the issues and context that gave rise to the study by summarizing what the current problem is and how the research may address it.[3] To illustrate, the following problem statement provides both background information (about the effectiveness of teaching palliative care skills and concepts across the continuum of medical education) and the issue at hand (the lack of consensus on how best to teach these competencies):

> Good evidence now demonstrates that palliative care competencies can be successfully taught at the undergraduate, graduate, and practicing physician level; however, the content of existing curricula and methods of instruction are inconsistent across academic centers, and national standards for medical school and residency palliative care education are needed.[4].

A well-articulated problem statement also helps readers anticipate the goals of each study (in this example, to assess which methods and curricula may be most effective in teaching palliative care). In laying out the issues and context, the authors should avoid broad generalizations or sweeping claims that are not backed up in their literature review (see Chapter 7).

### Conceptual framework

Most research reports cast the problem statement within the context of a conceptual or theoretical framework.[5] A description of this framework contributes to the research report in at least two ways: (1) it identifies research variables, and (2) it clarifies relationships among the variables.[6] Along with the problem statement, the conceptual framework sets the stage for the presentation of the specific research question that drives the investigation being reported. For example, the conceptual framework and research question for a study based on a formative evaluation framework would be different from the framework and question for a study based on a summative evaluation framework, even though their variables might be similar.

Cook and colleagues have shown that only 55 percent of published experimental studies in medical education included an explicitly articulated conceptual framework[7]; however, scholars argue that a conceptual or theoretical framework always underlies a research study, even if the authors do not explicitly articulate the framework.[8] This assertion may seem to be overreaching or inaccurate since many research problems and investigations originate from practical educational or clinical activities. Questions often arise from researchers' or educators' work; for example, a

clinical instructor might wonder why residents' test-interpretation skills did not improve after they received feedback. Further, some researchers undertake a study simply to report or describe an event, such as pass rates for women versus men on high-stakes examinations such as the United States Medical Licensing Examination (USMLE) Step 1. Nevertheless, the reviewer is usually able to construct at least a brief theoretical rationale for the study or place it in a range or combination of conceptual frameworks. In the USMLE example, the rationale may be to study gender equity and bias.

Each conceptual or theoretical framework provides a vocabulary for explaining the design and describing the results of the study. Using the vocabulary of a particular framework helps articulate the problem and ensure that the authors have accounted for and measured all the potential variables generally associated with the framework.[6] It also allows the reader and reviewer to better evaluate and interpret the methods and results.[7] The description of a manuscript's framework is usually more elaborate and detailed if the topic of the research has a long scholarly history (e.g., cognition, psychometrics) with well-established theories in which the authors have embedded their own empirical work.

In summary, the framework should provide the reader and reviewer with a clear sense of how the authors chose to frame or think about their research and where the manuscript fits into the existing literature.

## Research question

A more precise and detailed expression of the problem statement cast as a specific research question is usually stated at the end of the Introduction. To illustrate, a recent research report on teaching effectiveness ratings (TERs)[9] states,

> Our first objective was to identify the principal components of teaching effectiveness. Our second objective was to explore whether the relationship between these components and TER is modified by physical attractiveness. We predicted that if physical attractiveness impacts TER, there should be an independent association between the type of picture displayed and TER, or an interaction between ratings of the principal components of teaching effectiveness and TER.

A well-articulated research question, such as the example above, shows that, in experimental research, the logic revealed in the Introduction (through a well-considered problem statement and a thorough description of the conceptual framework) might result in explicitly stated hypotheses that would specify the dependent and independent variables.[10] For example, a paper on teaching effectiveness (dependent variable) might state that an objective was to understand how characteristics of the teacher and learners (independent variables) affected the teaching effectiveness.

By contrast, much of the research in medical education is not experimental. In such cases, it is more typical to state general research questions. One example reads, "To explore what third-year medical students learn from residents and which teaching strategies are used by excellent resident teachers in their interactions with students in the clinical workplace environment."[11] The authors may not specify a hypothesis, but their research questions should still specify or imply the dependent (e.g., learning) and independent (e.g., teaching strategies) variables.

For a few journals, the tradition is to define the main study variables (e.g., medical student learning) in the Introduction, but the tradition for most journals (including *Academic Medicine*) is for the variables to be defined in the Methods section. Whether the authors provide specific hypotheses and objectives or more general exploratory goals in their research questions, the reader—and the reviewer—should be able to anticipate what will be revealed in the Methods section from reading the research question.

## Summary

The purpose of the Introduction is to construct a logical "story" that will prepare the reader with background information and justification for the study that follows. The order of the components may vary—sometimes the problem statement appears after the conceptual framework, sometimes it appears in the first paragraph to prepare the reader for what to expect—however, in all cases, the Introduction should engage readers, encouraging them to finish the report.

*Funding/Support:* None reported.
*Other disclosures:* None reported.

## References

1. Zeiger M. Essentials of Writing Biomedical Research Papers, 2nd ed. London: McGraw-Hill; 1999.
2. Browner WS. Publishing and Presenting Clinical Research. Baltimore, MD: Williams & Wilkins; 1999:153–163.
3. Beckman TJ, Cook DA. Developing scholarly projects in education: A primer for medical teachers. Medical Teacher. 2007;29:210–218.
4. Schaefer KG, Chittenden EH, Sullivan AM, et al. Raising the bar for the care of seriously ill patients: Results of a national survey to define essential palliative care competencies for medical students and residents. Acad Med. 2014;89:1024–1031.
5. Chalmers AF. What Is This Thing Called Science? St. Lucia, Queensland, Australia: University of Queensland Press; 1982.
6. Bordage G. Conceptual frameworks to illuminate and magnify. Med Educ. 2009;43:312–319.
7. Cook DA, Beckman TJ, Bordage G. Quality of reporting experimental studies in medical education: A systematic review. Med Educ. 2007;41:735–745.
8. Hammond KR. Introduction to Brunswikian theory and methods. In: Hammond KR, Wascoe NE, eds. Realizations of Brunswik's Representative Design, New Directions for Methodology of Social and Behavioral Sciences, No. 3. San Francisco, CA: Jossey–Bass, 1980:1–11.

9. Rannelli L, Coderre S, Paget M, Woloschuk W, Wright B, McLaughlin K. How do medical students form impressions of the effectiveness of classroom teachers? Med Educ. 2014;48:831–837.
10. Fraenkel JR, Wallen NE. How to Design and Evaluate Research in Education, 4th ed. Boston: McGraw-Hill, 2000.
11. Karani R, Fromme HB, Cavea D, Muller D, Schwartz A, Harris IB. How medical students learn from residents in the workplace: A qualitative study. Acad Med. 2014;89:490–496.

# Chapter 7
# Reference to the Literature and Documentation

**Sonia J. Crandall, PhD, MS, and Addeane S. Caelleigh, MA**

**S. J. Crandall** is professor, Department of Physician Assistant Studies, Wake Forest School of Medicine, Winston-Salem, North Carolina.

**A. S. Caelleigh** was the Liaison Committee of Medical Education coordinator, Office of Medical Education, and she is now a visiting faculty member, University of Virginia School of Medicine, Charlottesville, Virginia.

## Review Criteria

1. ***The literature review is comprehensive, relevant, and up-to-date.***
2. ***The literature is analyzed and critically appraised; gaps in the literature are identified as a basis for the study.***

## Issues and examples related to the criteria

One of the reviewer's most important contributions is his or her assessment of the authors' use of the relevant literature and its presentation to the reader. A reviewer undertaking an assessment of any manuscript's literature review and documentation must take into account the type of research being reported. Research questions come from observing phenomena or reading the literature. Regardless of what inspired the research, authors must adequately review the existing literature to understand the scope of the issues relevant to their questions.

Before we delve deeper, we acknowledge that some authors conduct formal systematic reviews of the literature. These are addressed extensively in Chapter 18, "Reviewing a Review Manuscript." Here, we discuss the evidence base or literature review that authors should conduct before undertaking and presenting original research.

Authors of research manuscripts should achieve three key research aims through a careful analysis of the literature: (1) refinement of their research questions; (2) defense of their research design; and, ultimately, (3) support for their interpretations of their findings and for their conclusions. Thus, in the report, the reviewer should find a clear demonstration of the manuscript's contribution to the research questions and its context—that is, the authors should use the literature to make a compelling case for the need for and the publication of their own study.[1,2]

Before discussing the specifics of each of the three aims, it is important to offer some distinctions regarding the quantitative-qualitative research continuum. Where research fits along the continuum shapes how authors use literature within a study, although, notably, even at the ends of the spectrum, there are no rigid rules about how to use the literature.

Typically, at the *quantitative* end of the spectrum, researchers review the bulk of the literature at the beginning of the study to establish the theoretical or conceptual framework for the research question or problem. This review often includes a critique of alternative theories.[2] Researchers use the literature, adding citations to the Methods or other sections of the manuscript as necessary, to validate the application of specific methods, instruments, and statistical analyses. Researchers using more quantitative methods often cite additional references to pertinent literature in the Discussion section to highlight how their results support, refute, or expand previous research outcomes.

At the *qualitative* end of the spectrum, researchers weave relevant literature into all phases of the study, using it to guide the evolution of their thinking as they gather, transcribe, excerpt, analyze, and present data.[3] Researchers use the literature, as needed, to reframe the problem as the study evolves. Because the literature is more integrated throughout a report of qualitative research, the format may differ from the common IMRaD (Introduction, Methods, Results, and Discussion) structure used in many scientific research reports. To organize the reader (and the reviewer), there may be more subheadings in various sections of the manuscript in qualitative than in quantitative research reports. Unlike in many typical quantitative research reports, the authors may reference the literature in the Results section to articulate the relevancy of the themes that emerged from the data analyses.

Although the distinction is neither crystal clear nor unchanging, the reviewer might view the difference between the ends of the quantitative-qualitative continuum as the difference between *testing* theory-driven hypotheses (i.e., quantitative) and *generating* theory-building hypotheses (i.e., qualitative). The reviewer should ensure that authors all along that continuum use literature to inform the early development of their research interests, problems, and questions and, later, to conduct their research and interpret their findings.

### Refining the research question

The first key aim of the literature review is to refine the research question. The authors' review of the relevant literature should set the stage for the study. The review should provide a logically organized worldview either of the authors' questions or of their observations of the clinical or educational milieu. The literature should illuminate what knowledge relevant to the research question already exists, how the question or problem has been previously studied (e.g., study designs, methodological concerns), and which concepts and variables have been shown to be associated with the problem or question.[4] In other words, in defining the research question, the authors should use their literature review to evaluate previous work "in terms of its relevance to the research question of interest"[5] and to synthesize current knowledge, noting relationships that have been well studied and identifying

areas for elaboration, questions that remain unanswered, or gaps in understanding.[1,5]

To define their research aim, the authors should document the history and present the status of their study's question or problem. The literature they review should not only be current but also reflective of the contributions of salient, older research—whether published (e.g., journal articles, articles indexed in databases such as ERIC) or unpublished (e.g., dissertations). These dated materials may reflect significant evolutions in the topic. Regardless of perspective—qualitative, quantitative, or mixed method—the reviewer must be satisfied that the authors have framed the problem or research questions as precisely as possible from a chronological, developmental, and disciplinary perspective, given the confines of the literature.[3] For example, when presenting the tenets of adult learning as the basis for a program evaluation, authors would be remiss, if a historical perspective is germane to their project, to omit the foundational writings of Houle,[6] Knowles,[7] and, perhaps, Lindeman.[8]

## Defending the research design

The second aim of a literature review is to develop an evidence-based study, which means that authors must identify current knowledge and use it both to defend and support their study and to inform its design and methods.[1,5] To accomplish this aim, authors should

- interpret and weigh the available evidence,
- explicitly draw connections between the literature and their study design,
- present logical reasons for their use of specific methods, and
- describe in detail the variables or concepts that they scrutinize.

The reviewer's task, therefore, is to determine whether the authors have provided an adequate map for guiding the reader to the conclusions that the current study is important and necessary, the design is appropriate to answer the questions, and the study expands what is already known.[5]

In some cases, authors design a study and choose its methods based on experience and observation, rather than on a critical review of the literature, and then later, choose literature to support their study design. Such an approach often produces less reliable and sometimes flawed studies, but it is difficult for a reviewer to detect. One sign of such an approach may be that the supporting literature seems to fit the chosen design too perfectly—that is, the authors cite only literature that wholly supports their design. Another sign may be that the supporting literature seems less nuanced or less complex than normal, given the particular research area.

## Supporting interpretations of the outcomes

The third reason for including other literature in a report is to support the interpretation of the findings. Authors offer explanations, challenge assumptions, and make recommendations taking into consideration the literature they used initially to frame the research problem. They may cite some of the most salient literature at the end of the manuscript to support their conclusions (fully or partially), to refute current knowledge, to revise a hypothesis, or to reframe the problem.[9] The reviewer should be convinced that the authors have used literature to bring the reader back to the theory being tested (i.e., quantitative) or generated (i.e., qualitative).

## Further considerations

The reviewer must not only consider the pertinence of the literature to defining the research question, defending the methodology, and interpreting the results, but also evaluate the types of resources the authors cite and the balance of those resources' perspectives.

The reviewer should assess whether the references are general sources (e.g., textbooks),[5] primary sources (e.g., research or theory-based articles written by those who conducted the research or developed the theory),[5] or secondary sources (e.g., articles wherein authors describe the work of others).[5] References should predominantly be primary sources, whether published or unpublished. Secondary sources are acceptable, and desirable, if they provide a review (e.g., a meta-analysis) of what is known about the research problem or if primary sources are not available. Authors may use general resources as a basis for describing, for example, a theoretical or methodological principle or a statistical procedure.

Authors are obligated to comply with the manuscript-length requirements of journals, so they must be discriminating about what references they cite. Even though online-only journals have more flexibility about space limitations than do those that produce a hard-copy edition, authors should still select only the most foundational, relevant, and recent sources. They should not present an exhaustive literature review. They should, however, be credibly comprehensive, and their review should sufficiently identify the critical knowledge as well as the gaps in the research. The authors should convince the reviewer that they have critically examined (even if they have not cited) the full literature.

The reviewer should examine not only the breadth of resources but also their source. Although the Internet has made an abundant amount of material readily available, the scholarly quality of some of those materials is not assured. Online-only journals are prevalent. Some of these are indexed in databases such as SCOPUS and Web of Science, and the articles listed in these are typically of high quality; however, articles not vetted for quality are also easily available online through search engines such as Google Scholar.

The reviewer should assess whether references cover the entire body of existing literature. The pertinent literature may not all be published in commonly known journals or indexed in frequently used citation databases. Relevant articles may not be indexed in PubMed, but may be available through ERIC or other online citation databases. The reviewer may be able to determine

whether the authors searched multiple databases by the breadth of the disciplines represented by their citations or by the breadth of resources they reference (e.g., newspaper articles or primary historical documents). Have the authors restricted themselves too much or taken a too narrow view of the research question?

If the authors have relied solely on reports published in recognized online databases such as PubMed, then the viewpoint they present may be prejudiced toward only statistically significant outcomes,[10] or the authors may have neglected perspectives (e.g., law, the arts, humanities) that are not always covered in such databases. Further, when considering the perspectives presented by the authors, the reviewer should pay attention to whether the Discussion presents all the viewpoints that exist in the literature. Do the authors offer both confirming and conflicting views, both consensus and controversial opinions?[9,11]

The reviewer should be wary of authors who have not conducted a sufficient literature review. They may report a paucity of research in their area when, in fact, ample research exits. In these cases, the reviewer may push the authors to do more.

Finally, it is important that the authors explain how they found their resources (e.g., naming the databases they searched, providing the dates of searches) to give the reader (and reviewer) a sense of how comprehensive the literature search was. At the very minimum, the reviewer should comment on whether the authors have described, to the reviewer's satisfaction, how they found study-related literature and the criteria they used to select specific articles.

*Funding/Support:* None reported.
*Other disclosures:* None reported.

## References

1. Fink A. Conducting Research Literature Reviews: From the Internet to Paper, 4th ed. Thousand Oaks, CA: Sage Publiations, Inc.; 2014.
2. Lovejoy TI, Revenson TA, France CR. Reviewing manuscripts for peer-review journals: A primer for novice and seasoned reviewers. Ann Behav Med. 2011;42:1–13.
3. Haller EJ, Kleine PF. Using Educational Research: A School Administrator's Guide. New York, NY: Longman; 2001.
4. Boote DN, Beile P. Scholars before researchers: On the centrality of the dissertation literature review in research preparation. Educ Res. 2005;34:3–15.
5. Fraenkel JP, Wallen NE III, Hyun HH. How to Design and Evaluate Research in Education, 8th ed. New York, NY: McGraw-Hill; 2012.
6. Houle CO. The Inquiring Mind. Madison, WI: University of Wisconsin Press; 1961.
7. Knowles MS. The Modern Practice of Adult Education: From Pedagogy to Andragogy. Chicago, IL: Follett; 1980.
8. Lindeman EC. The Meaning of Adult Education. Montreal, Québec: Harvest House; 1961.
9. Martin PA. Writing a useful literature review for a quantitative research project. Appl Nurs Res. 1997;10:159–162.
10. Dwan K, Gamble C, Williamson PR, Kirkham JJ. Systematic review of the empirical evidence of study publication bias and outcome reporting bias—An updated review. PLoS One. 2013;8:e66844.
11. Bruette V, Fitzig C. The literature review. J N Y State Nurses Assoc. 1993;24:14–15.

# Chapter 8
# Relevance

**Louis Pangaro, MD, and William McGaghie, PhD**

**L. Pangaro** is professor and chair, Department of Medicine, F. Edward Hébert School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland.

**W. C. McGaghie** was director, Ralph P. Leischner, Jr. MD Institute for Medical Education, Loyola University Chicago Stritch School of Medicine, Maywood, Illinois, at the time this work was done. He is now professor of medical education, Northwestern University Feinberg School of Medicine, Chicago, Illinois.

## Review Criteria

1. ***The study is relevant to the mission of the journal or its audience.***
2. ***The study addresses important problems or issues; the study is worth doing.***
3. ***For quantitative studies: the study has generalizability because of the selection of participants, setting, and educational intervention or materials.***
4. ***For qualitative studies: the study offers concepts or theories that are generalizable or transferable to other contexts, people, etc.***

## Definitions and examples related to criteria

A key consideration for the editor in deciding whether to publish a manuscript is its relevance to the community (or, usually, communities) that the journal serves. The denotation of the term *relevance* is pertinence and relationship to something else. Indeed, one thing is often spoken of as being "relevant *to*" something else, and that "something" is the necessary context that establishes relevance. In this case, the context is the mission of the journal and the work of its readership. Relevance for an academic journal has several connotations, and all are judged with reference to a specific group of professionals and to the tasks of that group.

## Conditions of relevance

**Importance of content, the question at hand.** First, the editor and the reviewer must gauge the suitability of the manuscript to the journal's focus and to the interests of its readership. *Relevance* is a judgment about the propriety of the manuscript for the journal; does the topic "touch," or overlap, an area of interest for the journal? *Importance* is a judgment about the priority or magnitude of that overlap. A manuscript about cancer screening with no reference to educational issues would not be suitable for a journal focusing on health professions education, and it is likely to be dismissed out of hand. Each manuscript must be judged with respect to its propriety to the mission of the specific journal. The reviewer helps the editor judge this basic issue of relevance, irrespective of the intrinsic merit or quality of the report.

Within the journal's field of interest, relevance may be established in several ways. The reviewer and the editor can judge a manuscript's importance by examining several criteria related its main topic or subject:

- Does the manuscript address a serious problem? Does the problem present a barrier to effective or efficient work, and would its solution be of great help?
- Is the problem common? Is it prevalent enough to affect a large proportion of the community in question?
- Will the study contribute to solving a mystery of the mechanism of how problems arise? Does it address root causes?
- Does the question at hand reflect broader societal concerns that are current or persistent?

The reviewer is often able to determine whether a study meets the initial standard of relevance by assessing the article's Introduction.

**Importance of execution, rigor of methods.** Another condition of relevance, which the reviewer and the editor must consider once they have established that the topic of the manuscript is important to the readership, is whether the study's methods (including how participants were selected) are sufficient to enable readers to accept its results with confidence and to apply them in another setting. This is not so much a judgment about the adequacy of the research methodology in its own setting, as it is of its applicability in other settings—in other words, its relevance to the journal's readership. The reviewer may ask,

- Will a rigorous answer to this study's articulated question affect what readers do in their daily work?
- Will the study's results be generalizable to other institutions or settings?
- Does the study affect what other researchers will do in their next study or even, potentially, what policymakers may decide?

## Feasibility and applicability

Notably, the relevance of a topic is related to, but is not the same as, the feasibility of answering a research question. *Feasibility* is related to study design and deals with whether an answer is possible and, if so, how the answer can be derived. *Relevance* more directly addresses whether the question is significant enough to be worth asking.[1] The relevance of a manuscript is more complex than that of the topic per se; the relevance includes the importance of the topic *plus* whether the execution of the study or the discussion is powerful enough to affect what others in the field think or do. For example, some specific education methods (e.g., using actors

to simulate patients) may have seemed infeasible initially, yet those approaches are relevant and important.[2]

Additionally, a manuscript might be relevant even if its findings are not immediately applicable or practical. If the authors have provided important insights for understanding a theory or if they have suggested innovations that could advance the field, their report may be relevant to the community. In these cases, a journal leads its readership and does not simply reflect its interests.

## The relevance of different kinds of studies

The relevance of a manuscript is, as noted, often most immediately apparent in its first paragraphs, especially in how the authors pose or frame the research question. As discussed in Chapter 6, "Problem Statement, Conceptual Framework, and Research Question," a cogent report explicitly states the issue to be addressed in the form of either a question to be answered or a controversy to be addressed. A conceptual or theoretical framework underlies a research question, and a manuscript is stronger when the authors have made this framework explicit. An explicit presentation of the conceptual framework helps readers (including the reviewer) focus and helps clarify the study's importance or relevance.[3] It may be equally helpful to readers when the research question pertains to one of the several current fields of academic discourse listed by Hodges (e.g., "performance," emphasizing assessments in which skills are demonstrated, or "reflection," emphasizing self-regulation).[4]

The reviewer may gauge the relevance of a research manuscript by assessing its purpose or the intention of the study. Here, a vocabulary drawn from clinical research is applicable. Feinstein classifies research according to its "architecture"—that is, the effort to create and evaluate research structures that have both "the reproducible documentation of science and the elegant design of art."[5 (p 4)]

**Studies that quantify effects and justify action.** Studies that compare new methods with old or that compare two newly available methods of educational instruction or assessment have the potential to influence what the journal's readers might do or how they might use available resources. This potential exists even if a study's findings are "negative"—that is, they do not confirm the hypothesis at hand. Such studies are direct demonstrations of the benefits (or liabilities) of an instructional or assessment method. They have been described as "fruitful" by Bacon[6] and, therefore, justify action.[7] For studies without hypotheses (for instance, a systematic review of prior research or a meta-analysis), the same question applies: Does this review achieve a synthesis that will directly affect what readers do? A study may thus be relevant even though its immediate, practical application has not been worked out.

In cause-effect research, authors make specific comparisons (for instance, to the participants' baseline status or to a separate control group) to reach conclusions about the efficacy or impact of an intervention (e.g., a new public health campaign or an innovative curriculum). The relevance of such research architecture derives from its power to establish the causality, or at least the strong effects, of innovations. The relevance of research that deals with processes or, as defined by Feinstein,[5 (pp 15–16)] the products of a new protocol or the performance of a particular procedure (e.g., a new tool for the assessment of clinical competence) emanates from the quality or value of the process or procedure itself. In such cases, relevance does not derive from a cause-and-effect relationship but from a new measurement tool that readers can apply to a wide variety of educational settings.[1]

**Studies that clarify and explore mechanisms.** A manuscript, especially one involving qualitative research, may be pertinent to the community by virtue of its contribution to theory building, hypothesis generation, or methodology development.[8] In this sense, the manuscript clarifies or critiques issues of mechanism or cause[6,7] that, for example, underlie the teaching and practice of medicine, such as cognitive psychology, ethics, and epistemology. Some studies may be relevant if they address the mechanism of how or why a proven instructional method works—even if they do not add to the evidence that it does work. For example, Bennett and colleagues recently *applied* a social theory (Activity Systems Analysis) to explore how the tensions and factors for success of peer-assisted learning differ between the clinical workplace and small groups of students only.[9]

**Studies that describe new approaches.** Finally, "descriptive" studies simply explain a new method without providing any details to prove that the method is, in fact, effective.[7] The editor has to judge the potential interest and usefulness of these studies—without data that confirm benefit—to readers. Descriptive research provides collections of data that characterize a problem or provide information. The study design entails no comparisons, and the observations may be used for policy decisions or to prepare future, more rigorous studies. Many reports in social science journals, including those in health professions education, derive their relevance from such an approach.

## Judgments about a manuscript's relevance

Judging relevance is, at times, a dichotomous (yes/no) decision; the relevance is clear and obvious, or it is missing. However, relevance is often a matter of degree, as illustrated by the criteria listed at the beginning of this chapter. In this more common circumstance, determining relevance involves making a summary conclusion rather than a simple observation. It entails making a judgment supported by the applicability of a manuscript's principles, methods, instruments, and findings. Importantly, one reviewer's or reader's opinion of what is relevant may not align with another's, and the judgment of relevance is not infallible. In fact, one study of clinical research reports showed that readers did not always agree with reviewers on the relevance of studies to their own practice.[10]

The editor must choose among competing manuscripts because space is limited in each journal issue. Many journals ask reviewers, as part of their recommendations to the editor (see Chapter 22) to describe or, in some cases, rate using a scale how important a

manuscript is. The reviewer's comment on relevance provides the editor with a summary of the importance of a manuscript's subject, thesis, and conclusions to the journal's readership. The reviewer may want to consider the following specific questions in assessing a manuscript's relevance:

- Would a large part of the journal's community—or parts of several of its overlapping communities—consider the manuscript worth reading? Editors and reviewers of manuscripts for health professions education journals must be careful to consider the perspectives of educational practitioners in their judgments of relevance to ensure that the journal's content reflects the concerns of these readers.

- Is it important that this report be published even though the journal can publish only a small percentage of the manuscripts it receives each year?

## Summary

Relevance is a necessary but not a sufficient criterion for publishing an article. The rigorous study of a trivial problem or one already well-studied would not earn pages in a journal that deals with competing submissions. However, even in an electronic journal without page limitations, reviewers and editors must decide whether the question asked is worth answering at all. This is often referred to, colloquially, as the "so what?" issue: Will the solutions or answers to the question posed in the manuscript contribute, immediately or in the longer term, to the work of health professions education? Will the manuscript will be *helpful* to the journal's readership?

*Funding/Support:* None reported.
*Other disclosures:* None reported.
*Disclaimer:* The views expressed herein are the authors' own and do not necessarily represent those of the Uniformed Services University of the Health Sciences, the U.S. Department of Defense, or the federal government of the United States.

## References

1. Fraenkel JR, Wallen NE. How to Design and Evaluate Research in Education, 4th ed. Boston: McGraw-Hill; 2000:30–37.
2. Barrows HS. Simulated patients in medical teaching. Can Med Assoc J. 1968;98:674–676.
3. Bordage G. Conceptual frameworks to illuminate and magnify. Med Educ. 2009;43:312–319.
4. Hodges BD. The shifting discourses of competence. In: Hodges BD, Lorelei L, eds. The Question of Competence: Reconsidering Medical Education in the Twenty-First Century. Ithaca, NY: Cornell University Press; 2013:14–41.
5. Feinstein AR. Clinical Epidemiology: The Architecture of Clinical Research. Philadelphia, PA: W. B. Saunders; 1985.
6. Bacon F. The new organon. In: Jardine L, Sliverthorne M, eds. Francis Bacon: The New Organon. Cambridge Texts in the History of Philosophy. Cambridge, UK: Cambridge University Press; 2000.
7. Cook DA, Bordage G, Schmidt HG. Description, justification and clarification: A framework for classifying the purposes of research in medical education. Med Educ. 2008;42:128–133.
8. Mays N, Pope C. Qualitative research in health care. Assessing quality in qualitative research. BMJ. 2000;320:50–52.
9. Bennett D, O'Flynn S, Kelly M. Peer assisted learning in the clinical setting: An activity systems analysis. Adv Health Sci Educ Theory Pract. 2014; Oct. 1. [Epub ahead of print]
10. Justice AC, Berlin JA, Fletcher SW, Fletcher RH, Goodman SN. Do readers and peer reviewers agree on manuscript quality? JAMA. 1994;272:117–119.

# Chapter 9
# Research Design

**William C. McGaghie, PhD, Georges Bordage, MD, PhD, Sonia Crandall, PhD, MS, and Louis N. Pangaro, MD**

**W. C. McGaghie** was director, Ralph P. Leischner, Jr. MD Institute for Medical Education, Loyola University Chicago Stritch School of Medicine, Maywood, Illinois, at the time this work was done. He is now professor of medical education, Northwestern University Feinberg School of Medicine, Chicago, Illinois.

**G. Bordage** is professor, Department of Medical Education, the College of Medicine, University of Illinois at Chicago, Chicago, Illinois.

**S. Crandall** is professor, Department of Physician Assistant Studies, Wake Forest University School of Medicine, Winston-Salem, North Carolina.

**L. Pangaro** is professor and chair, Department of Medicine, F. Edward Hébert School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland.

## Review Criteria

1. ***The research paradigm or approach is identified.***
2. ***The design is appropriate for the research purpose or question. If a mixed-methods approach is used, the rationale is provided for the relationship between and sequencing of quantitative and qualitative aspects of the study.***
3. ***For quantitative studies: the design has internal validity, and potential confounding variables or biases are addressed.***
4. ***For quantitative studies: the design has external validity, including participants, settings, and conditions.***
5. ***For qualitative studies: the study design incorporates techniques to ensure trustworthiness.***
6. ***For studies with interventions: the intervention is described in sufficient detail (objectives, activities, time allocation, training) to be able to assess the likelihood of the intervention having the desired effect and/or to permit the study to be replicated.***
7. ***The research methods are defined and clearly described, and they are sufficiently detailed to provide transparency or permit the study to be replicated.***

## Issues related to the criteria

We agree with Dannels that "[t]he definition of research design is deceptively simple: it is a plan that provides the underlying structure to integrate all elements of a quantitative study so that the results are credible, free from bias, and maximally generalizable."[1] Research design has two key functions: (1) to provide answers to research questions and (2) to provide a road map for conducting a study using a planned and deliberate approach that controls or explains quantitative variation or organizes qualitative observations.[2] The design helps the investigator plan an orderly approach to addressing a research question through the collection, analysis, and interpretation of data. Research manuscripts should state clearly the quantitative (e.g., randomized trial, case-control study) or qualitative (e.g., ethnographic participant-observation) design for a reported study, and the reviewer should comment if the authors have not specified the design or, worse, if the design is not clear.

Research designs can fit anywhere along a continuum ranging from controlled laboratory investigations to observational studies. This continuum of designs is seamless, not sharply segmented, and goes from structured and formal to evolving and flexible. The reviewer should consider two basic questions related to research design:

1. Have the authors selected the best research design options to address their specific research question?
2. How well do the authors account for the design's strengths and limitations in the specific research context?

Selecting either a strictly quantitative or a strictly qualitative approach may not work because research excellence in many areas of inquiry often requires the best of both—that is, a mixed-methods approach.[3] If a mixed-methods approach is used, the reviewer should ensure that the authors have provided a rationale for the relationship between and the sequencing of quantitative and qualitative features of the study.

The reviewer should take into account key research design features when evaluating research manuscripts. The key features vary across different specialties and disciplines, and expert reviewers know about the critical research design features in their specific field. Health professions education research usually relies on research designs and methods from the social and behavioral sciences that involve research on human behavior. The key features for such studies relate to five issues or categories: design propriety, internal validity, external validity, unexpected outcomes, and plausibility.

### Design propriety

Is the research design appropriate, or as good as possible, for the research question? The matter of congruence, or "fit," is at issue because much research in health professions education is descriptive, comparative, or correlational, and much of it addresses new developments. For example, health professions education research studies have addressed creating new measurement scales,[4] adopting new standard-setting methods,[5] and revising achievement standards.[6] Other examples of health professions research include empirical demonstrations of novel practices such as mastery learning[7,8] and social network analysis to study clinical teams.[9]

Designs for these and other studies will differ depending on the focus of each research question.

Scholars have presented many different ways of classifying or categorizing research designs. For example, Fraenkel, Wallen, and Hyun have recently identified seven general research methods in education: experimental, correlational, causal-comparative, survey, content analysis, qualitative, and historical.[10] Their classification illustrates the overlap, and sometimes the confusion, that can exist among designs, data-collection strategies, and data analyses.[10] For example, investigators could structure a study as a prospective, comparative experimental design and then collect data via an open-response survey and analyze the written answers using content analysis.[11]

Each method or design category can be subdivided further. To illustrate, Fraenkel, Wallen, and Hyun break down experimental research into four subcategories: weak experimental designs, true experimental designs, quasi-experimental designs, and factorial designs.[10] Rigorous attention to the details of the specific design or subdesign encourages an investigator to focus the research method on the research question, which brings precision and clarity to a study.

As mentioned, health professions education research reports should clearly articulate the link between their research question and the research design; the reviewer's task is to make sure that link is transparent and the authors have embedded, within their description of the design, citations to the methodological literature to demonstrate awareness of fine points. In their review of 105 experimental studies in medical education, Cook, Beckman, and Bordage found that only 16 percent of the articles had an explicit statement of the research design.[12]

## Internal validity

The reviewer should assess whether the research study, as designed by the authors, has the internal validity (i.e., integrity or credibility) to address the question rigorously.[13] This calls for paying attention to a potentially long list of sources of bias or confounding variables, including but not limited to selection bias, attrition of research participants, the strength and integrity of any interventions, measurement bias, reactive effects, and study management.[14,15]

For studies using interventions, the reviewer should assess whether the authors have described the intervention in enough detail (covering, for example, objectives, activities, time allocation, and training) for readers to determine the likelihood that the intervention has the intended impact and to permit the study to be replicated. Effective educational interventions have strength and integrity and are sustained and measured over time.[16] Weak, one-shot educational interventions are unlikely to yield durable effects.

For qualitative studies, the study design should contain features or techniques (e.g., member checking, audit trail, triangulation, achieving saturation) that ensure the trustworthiness of the data. The research design, execution, and report must, together, convince the reviewer that the investigation represents an authoritative and dependable portrait of the persons, events, or conditions being investigated.[17–19]

## External validity

The reviewer must also determine whether the research study, as the authors have designed and described it, has external validity.[13] Are the results generalizable to participants, settings, and conditions beyond the research situation? Ensuring generalizability is frequently, but not exclusively, a matter of purposefully sampling participants, settings, and conditions as deliberate variables or features of the research design.

Health professions education research has been cast as translational science embedded within the familiar National Institutes of Health (NIH) T1 (translation of basic science research results to humans in clinical studies), T2 (translation of clinical discoveries into practice in controlled environments), and T3 (translation of practice guidelines and recommendations to actual or real-world settings) clinical science framework.[20,21] Results from educational research studies—such as trainees' acquisition of clinical skills (e.g., central venous catheter, or CVC, insertion)—in an educational setting (e.g., simulation education laboratory) are termed T1 outcomes. Research that demonstrates the transfer of acquired T1 skills to better patient care practices (e.g., fewer CVC needle sticks, lower complications, less bleeding) in clinics and wards reveals T2 outcomes. Research that shows that powerful educational interventions yield downstream results measured as improved patient or public health outcomes (e.g., reduced infection rates) is termed T3 translational science.[21]

Translational science in health professions education research aims to establish the external validity of educational interventions and outcome measures. Translational outcomes can rarely be achieved in the context of a single research study. Instead, generalizable T2 and T3 outcomes are more likely achieved from rigorous educational and health services research *programs* that are thematic, sustained, and cumulative.[22,23]

## Unexpected outcomes

Another question the reviewer should ask when evaluating a manuscript is whether the research design permits unexpected outcomes or events to occur and be recognized. To illustrate, the design of a study by Barsuk and colleagues enabled the discovery of unexpected, yet welcome, "collateral effects."[24] The authors, who were investigating a simulation-based mastery learning curriculum on CVC insertion, observed steadily increasing pretest passing scores among successive cohorts of new resident physicians in the same clinical and educational setting.[24] The unexpected collateral effects made it necessary to "raise the bar"—that is, increase the minimum passing standard (MPS) from 79 to 88 percent correct, a nearly flawless performance level.[6] Any research design that is too rigid to accommodate the unexpected may not properly reflect real-world conditions or may stifle the expression of the phenomenon being studied.

## Plausibility

A reviewer must consider whether the research design will lead to plausible outcomes, given the research question, the intellectual context of the study, and the practical circumstances where the study is conducted. Common flaws in research design that may decrease plausibility include

- failure to randomize correctly in a controlled trial,
- a small sample size resulting in low statistical power,
- brief or weak experimental interventions, and
- missing or inappropriate comparison (control) groups.

Qualitative studies are subject to similar design flaws, such as inappropriate or inadequate sampling or lack of thorough triangulation strategies (e.g., observations, focus groups, document reviews, and individual interviews) to satisfactorily answer the research question.[18] Signs of research implausibility include the authors' failure to describe the research design in detail, failure to acknowledge context effects on research procedures and outcomes, and the presence of study features that appear to be unbelievable, such as perfect response rates or flawless data. Often, there are tradeoffs in research between theory and pragmatics, precision and richness, elegance and application. Is the research design attentive to such compromises? In summary, the research design must be plausible and reported in sufficient detail to provide transparency or to permit the study to be replicated.

Finally, study design is intimately linked to the conceptual framework or theory undergirding a research project or investigation (see also Chapter 6). Hammond explains the bridge connecting research design, conceptual framework, and theory:

> Every method … implies a methodology, expressed or not; every methodology implies a theory, expressed or not. If one chooses not to examine the methodological base of [one's] work, then one chooses not to examine the theoretical context of that work, and thus becomes an unwitting technician at the mercy of implicit theories.[2]

In other words, authors should consider the design of any investigation as part of a larger theory, and the reviewer should consider how the design elements contribute to that theory.

*Funding/Support:* None reported.
*Other disclosures:* None reported.
*Disclaimer:* The views expressed herein are the authors' own and do not necessarily represent those of the Uniformed Services University of the Health Sciences, the U.S. Department of Defense, or the federal government of the United States.

## References

1. Dannels SA. Research design. In: Hancock GR, Mueller RO, eds. The Reviewer's Guide to Quantitative Methods in the Social Sciences. New York: Routledge; 2010:343–355.
2. Hammond KR. Introduction to Brunswikian theory and methods. In: Hammond KR, Wascoe NE, eds. New Directions for Methodology of Social and Behavioral Sciences, Vol. 3: Realizations of Brunswik's Representative Design. San Francisco: Jossey-Bass; 1980:1–11.
3. Cresswell JW, Klassen AC, Plano Clark VL, Smith KC for the Office of Behavioral and Social Sciences Research. Best Practices for Mixed Methods Research in the Health Sciences. Bethesda, MD: National Institutes of Health; 2011. http://obssr.od.nih.gov/mixed_methods_research/pdf/Best_Practices_for_Mixed_Methods_Research.pdf. Accessed March 19, 2015.
4. Loughry ML, Ohland MW, Moore DD. Development of a theory-based assessment of team member effectiveness. Educ Psychol Meas. 2007;67:505–524.
5. Yudkowsky R, Tumuluru S, Casey P, Herlich N, Ledonne C. A patient safety approach to setting pass/fail standards for basic procedural skills checklists. Simul Healthc. 2014;9:277–282.
6. Cohen ER, Barsuk JH, McGaghie WC, Wayne DB. Raising the bar: Reassessing standards for procedural competence. Teach Learn Med. 2013;25:6–9.
7. Wayne DB, Butter J, Siddall VJ, et al. Mastery learning of advanced cardiac life support skills by internal medicine residents using simulation technology and deliberate practice. J Gen Intern Med. 2006;21:251–256.
8. Cohen ER, Barsuk JH, Moazed F, et al. Making July safer: Simulation-based mastery learning during intern boot camp. Acad Med. 2013;88:233–239.
9. Shoham DA, Mundt MP, Gamelli RL, McGaghie WC. The social network of a burn unit team. J Burn Care Res. 2014. [Epub ahead of print]
10. Fraenkel JR, Wallen NE III, Hyun HH. How to Design and Evaluate Research in Education, 8th ed. New York: McGraw-Hill; 2012.
11. Bordage G. Moving the field forward: Going beyond quantitative-qualitative. Acad Med. 2007;82:S126–S128.
12. Cook DA, Beckman TJ, Bordage G. Quality of reporting experimental studies in medical education: A systematic review. Med Educ. 2007;41:735–745.
13. Shadish WR, Cook TD, Campbell DT. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston: Houghton Mifflin; 2002.
14. Kazdin AE, ed. Methodological Issues & Strategies in Clinical Research, 3rd ed. Washington, DC: American Psychological Association; 2003.
15. Maxwell SE, Delaney HD. Designing Experiments and Analyzing Data: A Model Comparison Perspective, 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates; 2004.
16. Cordray DS, Pion GM. Treatment strength and integrity: Models and methods. In: Bootzin RR, McKnight PE, eds. Strengthening Research Methodology: Psychological Measurement and Evaluation. Washington, DC: American Psychological Association; 2006:103–124.

17. Patton MQ. Qualitative Evaluation and Research Methods, 2nd ed. Newbury Park, CA: Sage Publications; 1990.
18. Shenton AK. Strategies for ensuring trustworthiness in qualitative research projects. Educ Inform. 2004;22:63–75.
19. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: A synthesis of recommendations. Acad Med. 2014;89:1245–1251.
20. Dougherty D, Conway PH. The "3T's" road map to transform US health care: The "how" of high-quality care. JAMA. 2008;299:2319–2321.
21. McGaghie WC. Medical education research as translational science. Sci Transl Med. 2010;2:19cm8.
22. McGaghie WC, Issenberg SB, Cohen ER, et al. Translational educational research: A necessity for effective health-care improvement. CHEST. 2012;142:1097–1103.
23. McGaghie WC, Issenberg SB, Barsuk JH, Wayne DB. A critical review of simulation-based mastery learning with translational outcomes. Med Educ. 2014;48:375–385.
24. Barsuk JH, Cohen ER, Feinglass J, McGAghie WC, Wayne DB. Unexpected collateral effects of simulation-based medical education. Acad Med. 2011;86:1513–1517.

# Chapter 10
# Instrumentation, Data Collection, and Quality Control

**C. Jessica Dine, MD, MSHPR, William C. McGaghie, PhD, Louis Pangaro, MD, and Judy A. Shea, PhD**

**C. J. Dine** is assistant professor of medicine, Division of Allergy, Pulmonary, and Critical Care, and associate program director, Internal Medicine Residency Program, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

**W. C. McGaghie** was director, Ralph P. Leischner, Jr. MD Institute for Medical Education, Loyola University Chicago Stritch School of Medicine, Maywood, Illinois, at the time this work was done. He is now professor of medical education, Northwestern University Feinberg School of Medicine, Chicago, Illinois.

**L. Pangaro** is professor and chair, Department of Medicine, F. Edward Hébert School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland.

**J. A. Shea** is professor of medicine and associate dean, Medical Education Research and Assessment, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

## Review Criteria

1. ***The development and content of the instrument(s)—as well as the preparation of observers, interviewers, and raters, as appropriate—are sufficiently described or referenced and are sufficiently detailed to permit transparency and/or replication.***
2. ***For qualitative studies: the characteristics of the researchers that may influence the research are described and accounted for during data collection.***
3. ***The measurement instrument is appropriate given the study's variables; the scoring method is clearly defined.***
4. ***The psychometric properties and procedures are clearly presented and appropriate.***
5. ***The data set is sufficiently described or referenced.***
6. ***Data quality control is described and is adequate.***

In research, *instrumentation* refers to the selection or development, and the later use, of tools that will be used to make observations about variables in a research study. These observations are the primary data that authors collect, record, analyze, and interpret. A manuscript should also clearly describe how and what types of data were collected so that the reviewer can assess whether the authors executed their methods as planned. This transparency in the methodology and data collection is necessary to judge the validity of the authors' conclusions.

## Instrumentation

In social and behavioral science research—including health outcomes, medical education, and patient education research—the tools or instruments are often questionnaires, either "paper-and-pencil" tools (participants respond in writing) or online, electronic questionnaires. Increasingly, observations and data come through qualitative methods in which the instruments are the researchers or observers themselves and the resulting transcripts or record of their observations (see also below). Research in the biological and physical sciences, in contrast, usually relies on tools such as microscopes and CAT scans that produce descriptions and/or other laboratory technologies such as clinical chemistry and immunoassays that produce quantified outputs. Regardless of the actual tools used, the goals and processes of developing and using instruments are similar across the physical and social sciences. The focus and examples in this chapter are from the social sciences—in particular, from health professions education research.

Instrumentation builds on a study's design, problem statement, and underlying conceptual framework (see Chapter 6). If the authors have appropriately and carefully considered and specified the design, research question, and theoretical framework for their investigation, they are likely to select the methodology, including the instruments, that will allow them to obtain the most unambiguous conclusion possible. (Some ambiguity is likely; the goal is to minimize it.) Specifically, the reviewer should focus on the rigor with which the authors selected or developed the instruments, as described below.

**Selecting the right instrument.** Describing the instrumentation for a particular study starts with specifying in what ways the authors have captured and/or measured the variables. The reviewer needs to know what observations the authors used as surrogates for the concept under investigation and how the data were collected (i.e., what were their means of obtaining observations?). Authors can choose from many different types of instruments for collecting data. Some of these instruments include multiple-choice tests and written examinations, measures of attitude, checklists, surveys, abstraction forms, interview schedules and guides, notes from field work, and rating forms.

Additionally, authors may use haptic measures such as sensors to detect, for example, depth of chest compression in advanced cardiovascular life support (ACLS) skill-acquisition studies or resident activity in a study of duty hours. Indeed, scholars recommend that investigators use multiple methods to address the same research construct, a process called *triangulation*.[1] The authors should specify whether the purpose of any additional instruments is to confirm the accuracy of the first instruments or to deliberately look at a different aspect that might provide divergent results.

"Selection" implies choice, and the reviewer should ask whether the choice of methods is justified by reference to a framework for analysis. For instance, a study examining whether exposure to conflict-of-interest policies during residency is associated with changes in prescribing patterns might be grounded in a typical analytic framework of decision making or habit formation.

The selection of the proper tool or instrument may be relatively straightforward because existing and well-known tools are available for authors who wish to capture a particular variable of interest (e.g., the Medical College Admission Test [MCAT exam] for medical school "readiness" or "aptitude"; National Board of Medical Examiners [NBME] subject examinations for "acquisition of medical knowledge"; the Association of American Medical Colleges [AAMC] Graduation Questionnaire for "curricular experiences"). Sometimes, however, the process of determining which tool to use is less straightforward. For example, authors may not have ready tools for describing and measuring the clinical competence of medical students after a required core clerkship. One approach for evaluating that particular variable of interest might be to use direct observations of students performing a clinical task, perhaps with standardized patients. Another approach might be to use a written test to learn what students would do in alternative, hypothetical situations, based on the premise of adaptive expertise. Yet another option would be to collect ratings and/or observation notes from clerkship directors at the end of the clerkship that attest to students' clinical skills. Other alternatives include using peer- and self-ratings of competence or collecting patient satisfaction data. Choosing among several possible measures of a variable is a key decision authors make when planning a research study.

If the authors use an existing instrument, the reviewer needs to learn from the manuscript the rationale for the choice. The authors should also demonstrate that they have validated the use of the instrument in their population. Importantly, the instruments they select should have a record of prior use that includes details about their measurement properties. In other words, authors should cite studies that have used the instruments and/or reported their psychometrics, or they may reference nontraditional resources such as MedEdPORTAL[2] and HaPI.[3]

**Creating a new instrument.** Often, a suitable measurement instrument is not available, and authors must develop instruments de novo. As a rule, when new instruments are used for research, more detail about their development (including calibration and piloting) is expected than when existing methods are used. Authors who develop new instruments for research take on the additional burden of proving the validity of the scores resulting from the instrument. In general, the data set used to validate an instrument should not be used to test the study's hypothesis. The reviewer does not have to be an expert in instrument development, but does need to be able to assess that the authors did the right things to ensure the validity and reliability of their instrument for the question at hand. Numerous publications describe the methods that should be followed in developing academic achievement tests,[4,5] rating and attitude scales,[6,7] checklists,[8] and surveys.[9–11] There is no single best approach to instrument development, but the process should be described rigorously and in detail, and the reviewer should look for citations in the manuscript so readers will be able to access information about the development of any instruments.

Instrument development starts with

- specifying the content domain (e.g., patient satisfaction),
- conducting a thorough literature review to see what similar or relevant instruments exist, and then, only if necessary,
- beginning to create a new instrument.

When authors develop new tools, they may draw the content from many sources, including potential subjects or participants, other instruments, the literature, and/or experts. What the reviewer needs to see is that the authors followed a rigorous process. The authors need to show that their instrument was not simply the result of a single investigator (or two) putting thoughts on paper. The reviewer should make sure that individual items on a written tool were critically reviewed for clarity and meaning and that the instrument was pilot tested and revised as necessary.

For some instruments, such as a data-abstraction form, pilot testing might mean as little as trying out the form on a sample of hospital charts. More stringent testing is needed for instruments that are administered to individuals. There are no rules for pilot testing in terms of how many people must test or sample each version, how many versions the developers must create, or what methods should be used for testing each version. The onus is on the authors to communicate the credibility of the process for developing the instrument, of the methods used, and of the resulting scores.

**Collecting data in qualitative traditions.** If the authors are using qualitative methods, they probably developed the instrument they used (e.g., script for interviews or focus groups) de novo. The thoughtful process of developing the guide questions and follow-up probes parallels that for a written tool, described above. The primary questions should be linked to the study aims and purpose. The authors need to consider language, using a vocabulary and syntax that are accessible to the study participants. Because qualitative methods often rely heavily on a conversation, it is imperative that the questions be open-ended and neutral[12] and that they follow a sensible and logical order. One way of developing such questions is mimicking a figure 8—that is, starting with broad questions, getting progressively more detailed and sometimes personal, and then broadening out with general, closing questions. Pilot testing for clarity, length, and ability to elicit a conversation is also necessary.[13]

**Considering confounders.** It is important to acknowledge that the characteristics of the researchers, observers, and/or raters may influence the research, including the design or selection of the instrument and the methods used to gather, analyze, and interpret

the findings. No research, especially qualitative research, can be completely free of bias; however, some aspects of a manuscript allow the reviewer to draw some conclusions about whether the authors have attempted to reduce bias, whether the authors have been transparent about their research processes or methods, and whether these processes seem replicable. The manuscript should provide an explanation of how the authors used the conceptual framework relevant to their work to guide the design of their study and the interpretation of their results. The authors should describe the training of raters, so that readers, including the reviewer, can determine whether the training was adequate.

The reviewer should also be able to assess whether the selection and use of the instrument, as well as the presentation of the results, minimized the researchers' potential bias.[14] For example,

- Who conducted the focus groups and interviews?
- Were these interviewers trained and if so, how?
- Were the interviewers aware of the study aims and primary research questions?
- Were they open to hearing opposing and unexpected viewpoints?
- If the authors conducted focus groups or interviews, did they describe their decision to end data collection?
- Did they conduct enough research and gather enough data to ensure that they did not miss new information—that is, did they reach saturation?

## Scoring or operationalizing data

**Quantitative data.** The reviewer needs to discern how scores or classifications are derived from an instrument. For example, how did the authors sum and dichotomize questionnaire responses such that respondents were grouped into those who "agreed" and "disagreed" or those who were judged to be either "competent" or "not competent"? An understanding of scores and classifications is also necessary for results that are quantified; if the scores were numerical, were they grouped into small-medium-large, and have the authors adequately explained the groupings? Notably, some instruments are not useful for yielding aggregated scores. In any case, the reviewer needs to be clear about how investigators operationalized research variables and judged the technical properties (i.e., reliability and validity) of their research data.

The authors also need to convey their understanding or use of terms and their decisions about cut-scores to readers. For example, in a study of the perceived frequency of feedback from preceptors and residents to students, the definition of "feedback" needs to be reported and justified. For example, is it a report of *any* feedback through any means, or is it formal feedback provided through a structured assessment tool? Investigators make many decisions in the course of conducting a study, and while they do not need to report all of them, they should present enough to allow readers to understand how they operationalized their variables of interest.

**Qualitative data.** In qualitative analysis, scoring is analogous to coding data. Multiple methods of coding or processing qualitative data are available to authors, such as content analysis, immersion/crystallization, or grounded theory.[15] While there are clear steps specific to each method, in general, coding processes allow for "discovering" themes within the data. The analysis of qualitative data often begins with developing a codebook for use across coders. The codebook is often quite detailed, listing subthemes within broader themes. It is developed through coding or analyzing multiple transcripts in an iterative manner. Once the codebook has stabilized, the codes are applied systematically across the remaining scripts; authors should check for inter-rater and intra-rater agreement, as applicable (see below). Sometimes, qualitative methods result in listings or tables of themes with some descriptive data, such as the prevalence of each theme among all interviewees.[15,16] Finally, it is important for the authors to constantly evaluate whether the emerging results address the study questions. Excellent references for evaluating qualitative studies are available.[17]

**Psychometric data.** If a manuscript is *about* an instrument, as opposed to the more typical case in which authors *use* an instrument to assess some question, then the authors might present methods for formal scale development and evaluation, often focusing on subscale definition, reliability estimation, reproducibility, and homogeneity.[18] Authors reporting large development projects for instruments designed to measure individual differences on a variable of interest will also need to pay attention to validity issues, sensitivity, and stability of scores.[19] The reviewer should ensure that a manuscript focusing on instrument development is generalizable—that is, that the authors have investigated and reported whether this instrument can be used in other settings.

**Administrating the tool.** The reviewer also needs to know the steps the authors took to ensure that the instrument is used properly. If the authors administered any tests, questionnaires, or forms themselves, the reviewer should look for important information concerning incentives for participants and processes used to gather complete data (e.g., contact of nonresponders, location of missing charts). For instruments that may be more reactive to the person using the forms (e.g., rating forms, interviews), the authors must have summarized coherently the actions they took to minimize differences related to the instrument user. These actions typically involve discussions of rater or interviewer training and the computation of inter-rater (between two or more raters) or intra-rater (within the same rater) reliability coefficients.[20] In other words, the reviewer should be able to identify how observers, interviewers, and raters were trained to ensure reliability and that the authors tested for potential variation between them.

**Defining secondary data.** In health professions education, researchers often use secondary data (i.e., data they themselves did not collect), so they have no input into data quality. Fortunately, for widely used datasets (e.g., NBME examination scores, AAMC questionnaires), quality is not an issue. However, authors do need to sufficiently describe the data set and its origins:

- Are the data the results of a survey?
- If so, to whom were instruments administered and when?
- What were the response rates?

Further, an explanation of how the authors used or categorized scores applies equally to existing data sets, such as the AAMC Faculty Roster or the American Medical Association (AMA) MasterFile, and to new data sets derived from the authors' work. In fact, data from existing databases may actually create more problems for the authors in terms of explaining and justifying their analytic decisions. A focus of these manuscripts should be on how the authors selected, cleaned, and manipulated data. For example, if the AMA Master File is being used for a study on primary care providers, how exactly have the authors defined their sample—by training, board certification, or self-report? Does "primary care" include both family medicine and internal medicine? The reviewer must look for evidence that the authors made sound decisions about defining their sample and treating missing data (e.g., how did the authors handle physicians in the dataset for whom no specialty information was available?).

## Quality control

In addition to reviewing the details about the actual instruments used in the study and the data derived from those instruments, the reviewer needs to gain a sense that the authors executed their methods as planned.[21] In most cases, it is impossible and unnecessary to report internal methods that were put in place for monitoring data collection and quality. This level of detail might be expected for a proposal or application, but it does not fit in most manuscripts. Still, depending on the methods of the study under review, the reviewer must assess a variety of issues, such as

- the unbiased recruitment and retention of subjects,
- the effects of conflicts of interest due to any funding sources (whether reported or not),
- the appropriate training of data collectors,
- the use of sensible definitions of analytic variables, and
- for studies conducted over many years, whether and how the integrity of the data set was preserved over time.

These are generic concerns relevant to any study. It would be too unwieldy to consider here all possible elements affecting the quality of research, but the reviewer needs to be convinced that the methods are sound. If the reviewer notes any carelessness, incompleteness, or inconsistency in reporting (or worse), he or she should be concerned and should mention any potential lapses in quality control to the editor.

In the end, the reviewer must be convinced that appropriate rigor was used in selecting, developing, and using measurement tools for the study. Without being an expert in measurement, the reviewer can look for relevant details about instrument selection and subsequent score development. If reviewers do not feel they can provide an adequate review of a new instrument, they may certainly suggest that an expert in instrument design review the manuscript. In recommending a paper for publication, the reviewer should be confident and clear about the procedures that the authors followed in selecting or developing and implementing data-collection tools.

*Funding/Support:* None reported.
*Other disclosures:* None reported.
*Disclaimer:* The views expressed herein are the authors' own and do not necessarily represent those of the Uniformed Services University of the Health Sciences, the U.S. Department of Defense, or the federal government of the United States.

## References

1. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait–multimethod matrix. Psychol Bull. 1959;56:81–105.
2. The Association of American Medical Colleges. MedEdPORTAL. https://www.mededportal.org/. Accessed March 18, 2015.
3. Behavioral Measurement Database Services (BMDS). HaPi: Health and Psychosocial Instruments. http://bmdshapi.com/. Accessed March 13, 2015.
4. Linn RL, Gronlund NE. Measurement and Assessment in Teaching, 8th ed. Englewood Cliffs, NJ: Prentice-Hall; 2000.
5. Millman J, Green J. The specification and development of tests of achievement and ability. In: Linn RL, ed. Educational Measurement. 3rd ed. Phoenix: American Council on Education and Oryx Press; 1993:335–366.
6. Scientific Advisory Committee of the Medical Outcomes Trust. Instrument review criteria. Med Outcomes Trust Bull. 1995;3:I–IV.
7. DeVellis RF. Scale Development: Theory and Applications. Applied Social Research Methods Series, Vol. 26. Newbury Park, CA: Sage; 1991.
8. Hales B, Terblanche M, Fowler R, Sibbald W. Development of medical checklists for improved quality of patient care. Int J Qual Health Care 2008;20:22–30.
9. Woodward CA. Questionnaire construction and question writing for research in medical education. Med Educ. 1998;22:347–363.
10. Rickards G, Magee C, Artino AR, Jr. You can't fix by analysis what you've spoiled by design: Developing survey instruments and collecting validity evidence. JGME. 2012;4:407–410.
11. Magee C, Rickards G, A Byars L, Artino AR Jr. Tracing the steps of survey design: A graduate medical education research example. JGME. 2013;5:1–5.

12. Flick U. An Introduction to Qualitative Research, 4th ed. Los Angeles: Sage; 2009.
13. Krueger RA. Focus Group Kit 6: Analyzing and Reporting Focus Group Results. Thousand Oaks, CA: Sage; 1998.
14. Nays N, Pope C. Rigour and qualitative research. BMJ. 1995;311:109–112.
15. Bernard HR, Ryan GW. Analyzing Qualitative Data: Systematic Approaches. Los Angeles: Sage; 2010.
16. Strauss AL. Qualitative Analysis for Social Scientists. Cambridge: Cambridge University Press; 1987.
17. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: A synthesis of recommendations. Acad Med. 2014;89:1245–1251.
18. Kerlinger FN. Foundations of Behavioral Research, 3rd ed. New York: Holt, Rinehart and Winston; 1986.
19. Nunnally JC. Psychometric Theory. New York: McGraw-Hill; 1978.
20. Streiner DL, Norman GR. Health Measurement Scales: A Practical Guide to Their Development and Use, 2nd ed. Oxford: Oxford University Press, 1995.
21. McGaghie WC. Conducting a research study. In: McGaghie WC, Frey JJ, eds. Handbook for the Academic Physician. New York: Springer-Verlag; 1986:217–233.

# Chapter 11
# Population and Sample

**William C. McGaghie, PhD, and Sonia Crandall, PhD, MS**

**W. C. McGaghie** was director, Ralph P. Leischner, Jr. MD Institute for Medical Education, Loyola University Chicago Stritch School of Medicine, Maywood, Illinois, at the time this work was done. He is now professor of medical education, Northwestern University Feinberg School of Medicine, Chicago, Illinois.

**S. Crandall** is professor, Department of Physician Assistant Studies, Wake Forest University School of Medicine, Winston-Salem, North Carolina.

## Review Criteria

1. ***For quantitative studies: the population is defined in sufficient detail to permit the study to be replicated.***
2. ***The sampling procedures are described in sufficient detail to permit transparency, replication, or theory generation.***
3. ***Samples are appropriate to the research purpose or question.***
4. ***Selection bias is addressed.***

## Issues and examples related to the criteria

Investigators conducting research in health professions education (and in other, related domains such as health outcomes, public health, and clinical practice) are expected to describe the research populations, sampling procedures, and research samples of the empirical studies they undertake. These descriptions must be clear and complete enough to allow reviewers and research consumers to decide whether the research results are valid internally and can be generalized externally to other research samples, settings, and conditions. Given necessary and sufficient information, reviewers and consumers can judge whether an investigator's population, sampling methods, and research sample are appropriate to the research question.

Sampling from populations has become a key issue in 20th- and 21st-century applied research. It allows for research efficiency without sacrificing accuracy. To illustrate, the Gallup Organization achieves highly accurate (±3 percentage points) estimates about opinions of the U.S. population (319 million) using samples of approximately 1,200 individuals.[1]

Samples in health professions education research come from at least two dimensions: (1) subjects or participants (e.g., U.S. medical students) and (2) stimuli or conditions (e.g., curriculum requirements, clinical problems or cases). Some investigators use a third approach—matrix sampling—to address research participants and stimuli simultaneously.[2] In all cases, however, the reviewer should find that the authors have defined their participant and stimulus populations, as well as their sampling procedures, clearly and thoroughly.

### Quantitative study sampling

The reviewer must assess how the authors, given their population of interest (e.g., U.S. medical students), have defined a population subset (or sample) for the practical matter of conducting their research study. Quantitative (and qualitative) investigators have provided detailed, scholarly descriptions of purist sampling procedures (e.g., multistage, random, systematic), which are difficult to achieve.[3,4] Other scholars have offered more practical guides. For example, Fraenkel and colleagues have identified five quantitative sampling methods that a researcher may use to draw a representative subset from a population of interest so that research results can be generalized broadly: random, simple, systematic, stratified random, and cluster.[5] The reviewer may note which of these methods the authors have used, whether the method is the most appropriate one, and whether it permits generalizing to the broader population.

### Qualitative study sampling

Qualitative approaches to participant and stimulus sampling, in contrast to quantitative, focus on the uniqueness of persons and conditions rather than on their general characteristics. Qualitative sampling need not be representative on statistical grounds (i.e., it does not require reports of numerical point estimates and confidence intervals). Instead, samples in qualitative research are chosen for their singularity, their special qualities, which, when revealed by research, permit an in-depth understanding through hermeneutics, or the science of interpretation. Qualitative research rigor has less to do with systematic sampling procedures than with the richness of the stories the research reveals.

For example, Gergen, Josselson, and Freeman state that "by far, the most practical form of qualitative inquiry is narrative research. Here it is assumed that one of the major ways in which we understand our lives is through stories."[6] The classic study of medical students in training at the University of Kansas reported by Howard Becker and colleagues in *Boys in White: Student Culture in Medical School* illustrates the power of qualitative, narrative research to reveal and explain everyday regularities in health professions education.[7] If the manuscript the reviewer is evaluating reports qualitative research, the task is to determine whether the sample provides fresh insight into or a greater appreciation of the population of interest.

### Convenience samples

Most investigators who are either conducting health professions education research with students or investigating educational interventions use convenience samples—that is, groups of students or conditions that are readily or easily used—even though this is not a recognized or formal method of sampling for representativeness. The reviewer must be aware that generalizing the results of studies using convenience samples may be misleading, unless there is a close match between research

participants and the target population to which the results are applied. Medical students, for example, are very homogeneous academically due to stringent selection criteria.[8] However, medical students display demographic variation (e.g., ethnicity, age, gender) that should be controlled or explained in research projects.

Sometimes, research is deliberately done on "significant"[9] or specifically selected samples, such as Nobel laureates[10] or astronauts and cosmonauts,[11] where descriptions of particular subjects rather than generalization to a population is the scholarly goal.

In intervention research, once a research sample is identified and drawn, its members may be assigned to certain study conditions (e.g., treatment and control groups). By contrast, conditions are uniform for all members of the sample in single-group observational studies, such as those examining statistical correlations among variables or those measuring possible improvement resulting from an educational intervention.[12] Qualitative observational studies of intact groups, such as the pulmonary and critical care residents described in *Life and Death in Intensive Care*[13] and the radiology residents described in *CT Suite: The Work of Diagnosis in the Age of Noninvasive Cutting*,[14] follow a similar approach, but the investigators use words, not numbers, to describe their research samples.

Systematic sampling of participants from a population of interest allows an investigator to generalize research results beyond the information obtained from the sample values. The degree to which the results are generalizable is an index of external validity. For example, research results obtained from a rigorous, stratified, random sample of U.S. medical students at a specific time has a high probability of being generalizable to the broader population of U.S. medical students at that same time. Stratified, random sampling is rarely accomplished for logistical and practical reasons. Instead, investigators rely on convenience samples—despite the pitfalls, selection biases, and attrition biases that reduce confident generalization of the research findings. Replications of research studies using new convenience samples that yield similar results boost confidence that research outcomes are stable and generalizable.

Similar logic—the need for convenience—applies to ensuring the internal validity of the convenience samples of the stimuli or independent variables involved in a research enterprise (e.g., clinical cases and their features in problem-solving research). Careful attention to stimulus sampling is the cornerstone of representative research.[15–17] To illustrate, medical learners and practitioners are expected to solve clinical problems of varying degrees of complexity as one indicator of their clinical competence. However, to date, neither the population of eligible problems nor clear-cut rules for sampling clinical problems from the parent population have been made plain. Thus, researchers select the problems, often expressed as cases, for evaluating medical personnel without a systematic sampling plan. This haphazardness likely contributes to the frequently cited, yet controversial,[18] finding of case specificity (i.e., nongeneralizability) of medical problem solving. A rival hypothesis is that case specificity has more to do with how the cases are selected or designed than with the problem-solving skill of physicians in training or practice.

Recent work on the construction of examinations of academic achievement in general[19,20] and medical licensure examinations in particular[21] is giving direct attention to stimulus sampling and representative design. Conceptual work in the field of facet theory and design[22] also holds promise as an organizing framework for research that takes stimulus sampling seriously.

As mentioned, research protocols that make provisions for systematic, simultaneous sampling of participants and stimuli alike use matrix sampling.[2] This approach is especially useful when an investigator aims to judge the effects of an overall program on a broad spectrum of participants and stimuli. Isolating and ruling out sources of bias is a persistent problem when identifying research samples for matrix sampling because of the nonrepresentativeness of participants and stimuli or a small number of cases.

## Selection bias

Selection bias is more likely to occur when

- investigators fail to specify and use explicit inclusion and exclusion criteria,
- there is differential attrition (drop out) of participants from study conditions, or
- samples are too small to give a valid estimate of population parameters and have low statistical power.

The reviewer must be attentive to these potential flaws. Research reports should also describe the use of incentives, compensation for participation, recruitment, informed consent, and whether the research participants are volunteers.

*Funding/Support:* None reported.
*Other disclosures:* None reported.

## References

1. Gallup. Election Polls—Accuracy Record in Presidential Elections. http://www.gallup.com/poll/9442/election-polls-accuracy-record-presidential-elections.aspx. March 19, 2015.
2. Thomas N, Raghunathan TE, Schenker N, Katzoff MJ, Johnson CL. An evaluation of matrix sampling methods using data from the National Health and Nutrition Examination Survey. Survey Method. 2006;32:217–231.
3. Lohr SL. Coverage and sampling. In: deLeeuw ED, Hox JJ, Dillman DA, eds. International Handbook of Survey Methodology. New York: Lawrence Erlbaum Associates; 2008:97–112.
4. Denzen NK, Lincoln YS, eds. The Sage Handbook of Qualitative Research, 4th ed. Thousand Oaks, CA: Sage Publications; 2011.

5. Fraenkel JR, Wallen NE, Hyun HH. How to Design and Evaluate Research in Education, 8th ed. New York: McGraw-Hill; 2012.
6. Gergen KJ, Josselson R, Freeman M. The promises of qualitative inquiry. Am Psychol. 2015;70:1–9.
7. Becker HS, Geer B, Hughes EW, Strauss AL. Boys in White: Student Culture in Medical School. Chicago: University of Chicago Press; 1961.
8. Prideaux D, Roberts C, Eva K, et al. Assessment for selection for the health care professions and specialty training. In: McGaghie WC, ed. International Best Practices for Evaluation in the Health Professions. London: Radcliffe Publishing, Ltd.; 2013:77–96.
9. Simonton DK. Significant samples: The psychological study of eminent individuals. Psychol Meth. 1999;4:425–451.
10. Zukerman H. Scientific Elite: Nobel Laureates in the United States. New Brunswick, NJ: Transaction Publishers; 1996.
11. Santy PA. Choosing the Right Stuff: The Psychological Selection of Astronauts and Cosmonauts. Westport, CT: Praeger; 1994.
12. Kratochwill TR, Levin JR. Single-Case Intervention Research: Methodological and Statistical Advances. Washington, DC: American Psychological Association; 2014.
13. Cassell J. Life and Death in Intensive Care. Philadelphia: Temple University Press; 2005.
14. Saunders BF. CT Suite: The Work of Diagnosis in the Age of Noninvasive Cutting. Durham, NC: Duke University Press; 2008.
15. Brunswik E. Systematic and Representative Design of Psychological Experiments. Berkeley, CA: University of California Press; 1947.
16. Hammond KR. Human Judgment and Social Policy. New York: Oxford University Press, 1996.
17. Maher BA. Stimulus sampling in clinical research: Representative design revisited. J Consult Clin Psychol. 1978;46:643–647.
18. Norman G, Bordage G, Page G, Keane D. How specific is case specificity? Med Educ. 2006;40:618–623.
19. Schmeiser CB, Welch CJ. Test development. In: Brennan RL, ed. Educational Measurement, 4th ed. Westport, CT: American Council on Education and Praeger Publishers; 2006:307–353.
20. Haladyna TM, Rodriguez MC. Developing and Validating Test Items. New York: Routledge; 2013.
21. Clauser BE, Margolis MJ, Case SM. Testing for licensure and certification in the professions. In: Brennan RL, ed. Educational Measurement, 4th ed. Westport, CT: American Council on Education and Praeger Publishers; 2006:701–732.
22. Shye S, Elizur D, Hoffman M. Introduction to Facet Theory: Content Design and Intrinsic Data Analysis in Behavioral Research. Applied Social Research Methods Series Vol. 35. Thousand Oaks, CA: Sage; 1994.

# Chapter 12
# Data Analysis and Statistics

**William C. McGaghie, PhD, and Sonia Crandall, PhD, MS**

**W. C. McGaghie** was director, Ralph P. Leischner, Jr. MD Institute for Medical Education, Loyola University Chicago Stritch School of Medicine, Maywood, Illinois, at the time this work was done. He is now professor of medical education, Northwestern University Feinberg School of Medicine, Chicago, Illinois.

**S. Crandall** is professor, Department of Physician Assistant Studies, Wake Forest University School of Medicine, Winston-Salem, North Carolina.

## Review Criteria

1. ***Data-analysis procedures are described in sufficient detail.***
2. ***Data-analysis procedures conform to the research design; hypotheses, models, or theory drives the data analyses.***
3. ***Statistical tests are appropriate.***
4. ***Topics such as effect size or functional significance, multiple tests or comparisons, and adjustment of significance level for chance outcomes were considered.***
5. ***Power issues are considered in studies that make statistical inferences.***
6. ***For qualitative analysis: how members of the research team contributed to coding, identifying themes, and/or drawing inferences is described; methods used to ensure trustworthiness of the analysis are also described.***

## Issues and examples related to the criteria

Researchers must perform and report their data analysis—no matter where their research falls along the seamless web of quantitative and qualitative research (see Chapter 9)—according to scholarly conventions. The conventions apply to statistical treatments of data expressed as numbers and to qualitative data expressed as descriptions of observations, field notes, interview transcriptions, abstracts from hospital charts, or other written archival records. When it comes to data analysis, investigators must "get it right" to ensure that the research progression of design, methods (including data analysis), results, and interpretation (Discussion and Conclusion) is orderly and integrated. The amplification of the six data-analysis and statistical review criteria in this section, meant to help the reviewer assess a manuscript's data-analysis section, underscores this assertion. (Additionally, the chapters on reporting results, Chapters 13–15, extend these ideas.)

### Quantitative

Statistical, or quantitative, analysis of research data is *not* the cornerstone of science. It does, however, appear in a large proportion of the research papers submitted to health professions education journals. The reviewer should expect a clear and complete description of research samples and data-analysis procedures in such manuscripts.

Statistical analyses of quantitative health professions education research data have three aims:

1. to establish data quality in terms of reliability estimates, which may be calculated in a number of ways depending on measurement scales, for the research variables,
2. to describe the central tendency (e.g., mean) and distribution of research data sets, and
3. to draw inferences about the meaning of research data (outcomes, findings) either from associations among variables or as differences resulting from educational interventions.

The quality of a research data set should be established before statistical analyses are performed. Investigators should report reliability coefficients, appropriate to variable measurement scales, for their research data, especially their dependent variables. Statistical analyses of low-quality (unreliable) data are unlikely to yield useful or interpretable results.[1] The reviewer may check for high-quality data sets by evaluating reliability coefficients for quantitative variables (or by studying evidence of data trustworthiness in qualitative research).

The authors should report the statistics that *describe* their research data set in a way that provides research consumers—and reviewers—with a clear portrait of the research variables and their properties. Continuous variables should be described in terms of central tendency (mean, median) and dispersion (standard deviation). Nominal data can be described as categorical frequencies or percentages. Descriptive statistics inform readers about the range and structure of a research data set in advance of the inferential data analyses.

Inferential data analyses should be performed only after the authors have established data quality and clearly described the data set. Researchers may perform a number of different statistical tests or analyses. The guiding principle the reviewer should follow is that data-analysis procedures must conform to the theory, hypotheses, and research design that underlie the research study. Statistical analyses must "fit" or be appropriate to answer the research question. Investigators should choose the simplest possible data-analysis procedure to address the research question—that is, what Wilkinson has termed "a minimally sufficient analysis."[2] Wilkinson continues,

> The enormous variety of modern quantitative methods leaves researchers with the non-trivial task of matching analysis and design to the research question. Although complex designs and state-of-the-art methods are sometimes necessary to address research questions effectively, simpler classical approaches often can provide elegant and sufficient answers to important questions. Do not choose an analytic method to impress your readers or to deflect criticism. If the assumptions and strength of a simpler method are reasonable for your data and research problem, use it. Occam's razor applies to methods as well as to theories.[2]

One family of statistical analyses seeks to reveal associations among measured variables. These analyses range from simple correlation coefficients to complex regression models, factor analyses, and cluster analyses.[3] Still other statistical methods used to study associations among measured variables (and among other research entities) include multidimensional scaling[4] and social network analysis.[5–7]

A second family of statistical analyses aims to study group differences, usually as a result of an educational treatment or intervention. These analyses range from simple *t* tests and analyses of variance (ANOVA) to very complex multivariate models.[3]

Measures of effect such as odds ratios and relative risk, which are frequently reported in clinical epidemiology studies,[8] have also been used in medical education research. Some journal editors also insist that authors report an index of effect size, such as Cohen's *d* coefficient, to amplify and support statistical analysis results.[9] Meta-analysis—that is, quantitative integration of research data from separate studies of the same research problem—is yet another data-analysis procedure that some medical education researchers have used.[10,11]

The results of statistical analyses of data often rest on assumptions about the data, including the measurement properties of the data and the normality of data distributions. These assumptions must be satisfied (i.e., realized) to make the data analysis legitimate. Investigators should use nonparametric, or "distribution-free," statistical methods to evaluate correlations among variables or group differences when research measurements are in the form of categories (e.g., female-male, pass-fail) or ranks (e.g., postgraduate year 1, 2, and 3 residents).

The reviewer's task, then, is to determine whether the authors used the most appropriate tests to analyze the data, given the study's design, methodology, and resulting data set. Specifically, the reviewer needs to look for signs that the statistical analysis methods were based on sound assumptions about the characteristics of the data and research design. The reviewer must be satisfied that the statistical tests presented in a research manuscript have been used and reported properly. Signs of a flawed data analysis include inappropriate or suboptimal analyses (e.g., wrong methods or tests) and failure to specify post hoc analyses before collecting data.

Performing analysis of data sets without attention to an explicit research design or an a priori hypothesis can quickly become an exercise in "data dredging."[12] The availability of powerful computers, user-friendly statistical software, and large institutional data sets increases the likelihood of such mindless data analyses. The ability to perform hundreds of statistical tests in seconds is not a proxy for thoughtful attention to research design and focused data analysis. The reviewer should also be aware that, for example, in the context of only 20 statistical comparisons, one of the tests will be likely to achieve "significance" at a traditional *P* level ($P \leq 0.05$) solely by chance. Multiple statistical tests or comparisons may call for adjustment of significance levels (*P* values) using the Bonferroni or a similar procedure to ensure accurate data interpretation.[3]

Research studies that involve small numbers of participants often lack enough statistical power to demonstrate significant results.[9] This shortfall can occur even when a larger study would show a significant effect for an experimental intervention or for a correlation among measured variables. Whenever a reviewer encounters a "negative" study (i.e., one with disconfirming results), he or she should consider the question of power and determine whether lack of power was the reason for a nonsignificant result.

Shadish, Cook, and Campbell point out that thoughtful data analysis is needed to achieve what they term statistical conclusion validity (internal validity) and construct validity (external validity) when results are being generalized from samples to larger populations.[13] Internal and external validity are different from the valid interpretation of test scores and other forms of data, which have been the subject of recent discussion.[14]

In summary, the reviewer should note whether any quantitative data analyses in a manuscript are hypothesis driven, follow an orderly plan, are "minimally sufficient," and have sufficient statistical power. He or she should assess whether the authors have established the quality (reliability) of their data set, described data distributions, and demonstrated associations or differences among variables.

## Qualitative

Qualitative data analysis has a deep and longstanding research legacy in health professions education and medicine. Well-known and influential examples are *Boys in White*, the classic study of student culture in medical school, written by Howard Becker and colleagues[15]; psychiatrist Robert Coles' five-volume study, *Children of Crisis*, about children growing up under perilous conditions[16]; the classic participant-observation study of patient culture on psychiatric wards by David Rosenhan[17]; and Terry Mizrahi's observational study of the culture of residents on hospital wards, *Getting Rid of Patients*.[18] The reviewer should be familiar with the scholarly contribution of qualitative research in medical education.

Analysis of qualitative data, which involves the manipulation of words and symbols rather than of numbers, is also governed by rules and rigor. Investigators who have undertaken qualitative research should carefully plan their research and analyses; they

are expected to use established, conventional approaches to ensure both data quality and accurate analysis. Flaws in qualitative analysis include (but are not limited to) the following:

- inattention to data triangulation (cross-checking information sources);
- insufficient detail in depicting research observations (lack of "thick description");
- failure to use recursive (repetitive) data analysis and interpretation;
- failure to verify independent data through colleagues or groups similar to participants (peer debriefing);
- failure to verify independent data through other stakeholders (member checking);
- failure to enhance trustworthiness by seeking further data until no new categories or themes arise (e.g., saturation); and
- failure to express the investigators' personal orientations (e.g., homeopathy) early in the written report.

Prominent resources on qualitative research provide research insights and methodological details that would be useful for the review of a complex or unusual study.[19–24] Conventions for reporting qualitative medical education research have been published recently.[25]

In summary, analysis of qualitative data in health professions education research needs to be planned and rigorous, and it must conform to prevailing scholarly conventions and expectations.

## Conclusions

Analyzing data—irrespective of whether the work uses a quantitative or qualitative approach—is an essential part of conducting research. Data analysis of any variety needs to be performed with skill and care, and it should emanate directly from the investigator's research question and research design.

*Funding/Support:* None reported.
*Other disclosures:* None reported.

## References

1. Ghiselli EE, Campbell JP, Zedeck S. Measurement Theory for the Behavioral Sciences. San Francisco: W.H. Freeman; 1981.
2. Wilkinson L. Statistical methods in psychology journals: Guidelines and explanations. Am Psychol. 1999;54:594–604.
3. Maxwell SE, Delaney HD. Designing Experiments and Analyzing Data, 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates; 2004.
4. Young FW, Hamer RM. Multidimensional Scaling: History, Theory, and Applications. Hillsdale, NJ: Lawrence Erlbaum Associates; 1987.
5. Lurie SJ, Fogg TT, Dozier AM. Social network analysis as a method of assessing institutional culture: Three case studies. Acad Med. 2009;84:1029–1035.
6. Scott J. Social Network Analysis, 3rd ed. Thousand Oaks, CA: Sage Publications; 2013.
7. Shoham DA, Mundt MP, Gamelli RL, McGaghie WC. The social network of a burn unit team. J Burn Care Res. 2014. [Epub ahead of print]
8. Fletcher RH, Fletcher SW, Fletcher GS. Clinical Epidemiology: The Essentials, 5th ed. Philadelphia: Lippincott Williams & Wilkins; 2014.
9. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
10. McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Does simulation-based education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. Acad Med. 2011;86:706–711.
11. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Mastery learning for health professionals using technology-enhanced simulation: A systematic review and meta-analysis. Acad Med. 2013;88:1178–1186.
12. Norman G. Data dredging, salami-slicing, and other successful strategies to ensure rejection: Twelve tips on how to not get your paper published. Adv Health Sci Educ. 2014;19:1–5.
13. Shadish WR, Cook TD, Campbell DT. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston: Houghton-Mifflin; 2002.
14. Kane MT. Validating the interpretations and uses of test scores. J Educ Meas. 2013;50:1–73.
15. Becker HS, Geer B, Hughes EC, Strauss A. Boys in White: Student Culture in Medical School. Chicago: University of Chicago Press; 1961.
16. Coles R. Children of Crisis: A Study of Courage and Fear. Vols. 1–5. Boston: Little, Brown; 1967–1977.
17. Rosenhan DL. On being sane in insane places. Science. 1973;179:250–258.
18. Mizrahi T. Getting Rid of Patients: Contradictions in the Socialization of Physicians. New Brunswick, NJ: Rutgers University Press; 1986.
19. Glaser BG, Strauss AL. The Discovery of Grounded Theory: Strategies for Qualitative Research. Chicago: Aldine; 1967.
20. Denzen NK, Lincoln YS, eds. The Sage Handbook of Qualitative Research, 4th ed. Thousand Oaks, CA: Sage Publications; 2011.
21. Harris IB. Qualitative methods. In: Norman GR, van der Vleuten CPM, Newble DI, eds. International Handbook for Research in Medical Education. Dordrecht, NL: Kluwer Academic; 2002:45–95.
22. Gicomini MK, Cook DJ. User's guide to the medical literature. XXIII. Qualitative research in health care. A. Are the results of the study valid? JAMA. 2000;284:357–362.

23. Gicomini MK, Cook DJ. User's guide 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;284:478–482.
24. Gergen KJ, Josselson R, Freeman M. The promises of qualitative inquiry. Am Psychol. 2015;70:1–9.
25. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: A synthesis of recommendations. Acad Med. 2014;89:1245–1251.

# Chapter 13
# Results: Presentation

**Glenn Regehr, PhD, Lara Varpio, PhD, and Kulamakan Kulasegaram, PhD**

**G. Regehr** is professor, Department of Surgery, and associate director, Research, Centre for Health Education Scholarship, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, Canada.

**L. Varpio** is associate professor, Department of Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland.

**K. Kulasegaram** is assistant professor, Department of Family and Community Medicine, and scientist, The Wilson Centre, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada.

## Review Criteria

1. ***All results are presented. The results align with the methods and study questions.***
2. ***The amount of data presented is sufficient, balanced, accurate, and supportive of inferences or themes.***
3. ***Tables, graphs, or figures are used judiciously and agree with the text.***
4. ***The statistics are reported correctly and appropriately.***

## Issues and examples related to criteria

The Results section of a research report lays out the body of evidence collected within the context of the study to support the questions, discussions, and/or conclusions that the authors present. To be effective, the study results and their relation to the research questions and discussion points must be clear to the reader. Unless this relationship is clear, the reader cannot effectively judge the quality of the evidence or the extent to which it supports the claims made in the report.[1,2] Several devices can maximize the presentation of results, and the reviewer needs to be aware of them so that they can effectively review the results, express any concerns they may have to the editor, and provide useful feedback to the authors.

## Organization of the data and analyses

The organization of the data and analyses is critical to the coherence of the Results section. The authors should present the data and analyses in an orderly fashion, and the logic inherent in that order should be explicit.[3] There are several possible ways to organize data, and the choice of organization ought to be strategic, reflecting the needs of the journal's audience and the nature of the findings being presented. The reviewer should be alert to the organization and determine whether this particular organization is effective in conveying the results coherently.

One very helpful type of organization is to use a parallel structure across the entire research report—that is, to make the organization of the results consistent with the organization of the other sections of the report. Thus, the organization of the Results section would mirror the organization of the research questions that were established in the Introduction; it would follow the descriptions provided in the Methods section; and it would anticipate the organization of points to be elaborated upon in the Discussion.

If there are several research questions, hypotheses, or important findings, the Results section may be best presented as a series of subsections, each of which presents the results that are relevant to a given question, hypothesis, or set of findings. This type of organization clarifies the point of each set of results or analyses and, thus, makes it relatively easy to determine how the results or analyses speak to the research questions. In so doing, this organization also provides an easy way to determine whether the authors have addressed each of the research questions appropriately and completely, and it provides a structure for identifying post hoc or additional analyses and serendipitous findings that might not have been initially anticipated.

However, there are other ways to organize a Results section that also maintain clarity and coherence and may better represent the data and analyses. Many of these approaches are used in the context of qualitative research, but they may also be relevant to quantitative research designs. For example, the authors may group results according to the themes, recurrent patterns, or relationships that they identified during data collection and analysis. Notably, themes, patterns, or relationships often overlap in complex ways, so the authors must take care to focus the reader on the particular issue under consideration while simultaneously identifying and explaining its relationship to the others. Alternatively, the authors might organize the data in relation to a theory or a model they are developing, revising, or using as a framework for analysis. Yet another possibility is to organize the data according to the method of collection (e.g., interviews, observations, documents) or the critical phases in the data-analysis process (e.g., primary node coding and axial coding).

In short, authors can use several different organizational structures to present results. Regardless of the organization, however, if the Results section does not clearly establish the relevance of the data presented and the analyses performed to the questions and aims of the report, then the authors have not properly demonstrated the point they were trying to make with the findings, and the Results section has failed. If the results are not coherent, the reviewer must consider whether the problem lies in poor execution of the analyses or in poor organization of the Results section. If the former, the report is probably not acceptable. If the latter, the reviewer might merely want to suggest an organizational structure that would convey the results effectively.

## Selection of qualitative data for presentation

Qualitative research can produce great amounts of raw material. While the authors will order, vet, and explain this raw material through the analysis process, they may still possess an overwhelming set of possible data (e.g., participant quotes, field notes, text excerpts, or visual images) to cover in the Results section. Selecting which data to present in a Results section is, therefore, critical.

The logic that informs this selection process should be transparent and related explicitly to the research questions and objectives. If it is not clear to the reviewer, for example, why the authors chose a particular piece of data to explicate a point, or if an alternative interpretation of the data is reasonable, the reviewer should note this ambiguity or discrepancy and might reasonably ask for further clarification or justification for the choice. In addition, the authors should make clear any implicit relationships among the results, such as trends, contrasting cases, and voices from a variety of perspectives. The reviewer should be alert to any hints that that the process of selecting data might have distorted the overall gist of the entire data set. Often, discerning the relationship of the data selected to the full data set is difficult for a reviewer, but hints of distortions may arise through inconsistencies between pieces of data across the Results section, such as a quotation for one theme that seems to disconfirm a claim or interpretation made for another theme.

The variety of organizational structures available for the presentation of qualitative results requires an equally varied approach in data display. For example, data can be presented in tables or figures or as incorporated narrative descriptions or text blocks embedded in the description of results. Qualitative data may sometimes be "quantitized"[4] into numeric counts if that numeric representation supports a meaningful presentation of results (see Chapter 15). Regardless of display techniques, the decision of how much qualitative data to incorporate must be judicious. Authors should ensure that narrative excerpts are only as long as required to represent a theme or point of view but take care not to minimize them to the point of distorting their meaning or diluting their character. Determining whether the authors have excerpted just the right amount of text—not too little and not too much—can be a difficult, subtle task, but the balance is essential to the efficient yet accurate presentation of findings about complex social phenomena.

## The balance of descriptive and inferential statistics for quantitative data

In reports of quantitative, or hypothesis-testing, research, a rough parallel to the qualitative issue of selecting data for presentation is the balance of descriptive and inferential statistics. One common shortcoming in quantitative reports is that the Results section focuses very heavily on inferential statistics and not enough on descriptive statistics. Researchers—and, importantly, the reviewer—should remember that the inferential statistics are presented only to aid in the reasonable interpretation of the descriptive statistics. If the data (or patterns of data) to which the inferential statistics are being applied are not clear, then the point of the inferential statistics has not been properly established, and the Results section has failed. Again, however, balancing inferential and descriptive statistics can be a subtle task. Excessive presentation of descriptive statistics that do not speak to the research objectives may also make the Results section unwieldy and uninterpretable.

When authors present descriptive statistics, the reviewer must ensure that the numbers in any tables or other exhibits, the numbers in the text, and the prose description of the data—both in the abstract and the body of the report—are consistent. Literal mismatches are often obvious if the authors present the same data in more than one place, and the reviewer should note these. However, the reviewer should also be alert to inconsistencies in patterns across the data set that imply that the story the authors are constructing about the results might be too simplistic or overstated. As with reviews of the presentation of qualitative data, reviewers should be alert for potential patterns of data that might be negative or disconfirming, but are not properly represented in the textual description of the results.

When inferential tests are presented, the reviewer must ensure that they are appropriate for the research question, accurate, properly described, and interpretable by the typical reader of the journal. If the form of analysis is beyond the knowledge or experience of the reviewer, then the reviewer should acknowledge this in his or her review so that the editor can obtain a more technical review if needed. However, if the form of analysis is so complicated that it confuses a reviewer (especially one who has expertise in the methodologies used), the reviewer might legitimately raise concerns in the review that the statistics are not effectively described for an audience that is highly diverse in its statistical knowledge. Some level of lay explanation should be present so that, for the typical reader, the Results section is not merely a string of incomprehensible phrases, symbols, and numbers.

## The use of narration for quantitative data

In the context of quantitative studies, the Results section is generally not the place to elaborate on the implications of the data collected, how the data fit into the larger theory that is being proposed, or how they relate to other literature. Providing implications, context, and comparison is the role of the Discussion section. This being said, it is also true that the Results section of a quantitative, or hypothesis-testing, study should not be merely a string of numbers and Greek letters. Rather, the results should include a narrative description of the data, a brief justification of each analysis, and an explanation of what the resulting statistics signify in relation to the research question.

Striking an effective balance between a thorough description of the results and an extrapolation of the implications of the results requires skill and subtlety (like knowing how much and what kinds of data to present); however, the distinction between explaining findings and extrapolating implications is important. For example, sophisticated analysis techniques that use unfamiliar

or novel statistical tests may require explicit statements about the meaningfulness of the outcomes and their implications for the model or research question.

Thus, it is reasonable—in fact, expected—that a Results section include a statement such as, "Based on the pattern of data, the statistically significant two-way interaction in the analysis of variance (ANOVA) implies that the treatment group improved on our test of knowledge more than the control group." It is not appropriate for the Results section to include a statement such as, "The ANOVA demonstrates that the treatment is effective," or, even more extreme, "The ANOVA demonstrates that we should adopt this teaching methodology." The first statement is a narrative description of the data interpreted in the context of the statistical analysis. The second statement is an extrapolation of the results to the research question and belongs in the Discussion. The third is an extreme overinterpretation of the results—it presents a highly speculative value judgment about the importance of the outcome variables used in the study relative to the huge number of other variables and factors that must be weighed in any decision to adopt a new educational method (and it should not, at least in the form presented above, appear anywhere in the report). The reviewer is responsible for determining whether the authors have found the appropriate balance of description. If not, the reviewer should identify particular areas of concern (too little description or too much interpretation) in the feedback he or she provides to the authors and the editor.

## Contextualization of qualitative data

Notably, in qualitative research traditions, the distinction between the presentation and the discussion of results is less clear. This more nebulous relationship can add a layer of complexity for the reviewer in determining whether particular concepts belong in the Results or the Discussion section.[5] Nonetheless, issues parallel to the presentation of quantitative data are important for the narrative presentation of data in qualitative studies.

The presentation of qualitative results should not consist of merely a listing of narrative excerpts, visual images, and/or numeric summary counts. Narrative results, like numeric data, cannot stand on their own. They require descriptions of their origins in the data set and an explanation of the understandings they provide. In the process of selecting material from a set of qualitative data (e.g., when carving out relevant narrative excerpts from analyzed focus group transcripts), the data must not become "disconnected" and void of their original meanings. Authors should contextualize the presentation of qualitative data in relation to the main analytic findings, and they should frame the data with a descriptive summary.

The results may also be described in relation to a theory (either a preexisting theory or a theory that the authors are developing in the report). A good qualitative Results section provides a framework for the selected data to ensure that their original contexts are sufficiently apparent such that the reviewer can judge whether the authors' interpretation is faithful to and reflects those contexts.

## The use of tables and figures

Tables and figures present tradeoffs because they often are the best way to convey complex data, yet if they are overused, they may present data that are not relevant to the point of the research, disrupt the flow of the story being developed, or be inappropriately relied upon as a substitute for the effective narration of the results as described above. The reviewer must, therefore, evaluate whether the tables and figures are the most efficient or clearest way to present the data in a report[6] and whether the exhibits are used appropriately sparingly. Authors should make every effort to combine data into the fewest number of exhibits possible. In addition, if data are presented in tables or figures, they should not be repeated in their entirety in the text; rather, the text should be used to describe the table or figure, highlighting the key elements in the data as they pertain to the relevant research question, hypothesis, or analysis. When the results of statistical tests are in the form of tables or charts, the reviewer should be able to source or match these to a specific analysis outlined in the Methods and presented in the Results.

Finally, although somewhat mundane, an important responsibility of the reviewer is, as mentioned, to determine whether the data in the tables, the figures, and the text, including the abstract, are consistent. If the numbers or descriptions in the text do not match those in the tables or figures, the reviewer should have—and should note—serious concerns about the quality control used in the data analysis and interpretation.

## Summary

Whether the authors are reporting on a quantitative or qualitative study, the Results section represents the evidence from which they will make claims about their work. Thus, the Results section must be organized so as to make the connection between evidence and claims clear and convincing. The accuracy of the data and the analyses is, of course, critical; however transparency and interpretability are also vital. The reviewer's task with regard to the Results section, therefore, is not merely to vet the data and analyses, but also to assess the clarity of presentation. The next two chapters will extend this discussion by focusing on, respectively, the reporting of statistical analyses and the reporting of qualitative findings.

*Acknowledgments:* The current authors acknowledge Georges Bordage, Addeane S. Caelleigh, and Ann Steinecke for their work on the previous version of this chapter.
*Funding/Support:* None reported.
*Other disclosures:* None reported.
*Disclaimer:* The views expressed herein are the authors' own and do not necessarily represent those of the Uniformed Services University of the Health Sciences, the U.S. Department of Defense, or the federal government of the United States.

## References

1. Regehr G. The experimental tradition. In: Norman GR, van der Vleuten CPM, Newble D, eds. International Handbook of Research in Medical Education. Dordrecht, The Netherlands: Kluwer Academic Publishers; 2002:5–44.
2. Harris IB. Qualitative methods. In: Norman GR, van der Vleuten CPM, Newble D, eds. International Handbook of Research in Medical Education. Dordrecht, The Netherlands: Kluwer Academic Publishers; 2002:45–96.
3. American Psychological Association. Publication Manual of the American Psychological Association. 6th ed. Washington, DC: American Psychological Association; 2010.
4. Sandelowski M, Voils CI, Knafl G. On quantitizing. JMMR. 2009;3:208–222
5. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: A synthesis of recommendations. Acad Med. 2014;89:1245–1251.
6. Henry GT. Graphing Data: Techniques for Display and Analysis. Thousand Oaks, CA: Sage; 1995.

# Chapter 14
# Results: Reporting Statistical Analyses

**Kulamakan Kulasegaram, PhD, and Glenn Regehr, PhD**

**K. Kulasegaram** is assistant professor, Department of Family and Community Medicine, and scientist, The Wilson Centre, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada.

**G. Regehr** is professor, Department of Surgery, and associate director, Research, Centre for Health Education Scholarship, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, Canada.

## Issues and examples related to reporting statistical analyses

Authors are using ever more sophisticated and complex statistical techniques to analyze health professions education studies. Most statistical techniques use a significance-testing approach (e.g., *t* tests, analysis of variance [ANOVA], Pearson correlations)[1] or a model-fit approach (e.g., structural equation modeling),[2] both of which involve standard reporting guidelines. Psychometric analysis techniques in measurement or assessment (e.g., generalizability analyses, item-response modeling) also have their own set of standards for reporting.[3] Although specific technical issues in a research report's analysis may require specialized knowledge in these areas, some general guidelines apply for reporting all analysis methods.

## Appropriately performing and reporting analyses

Importantly, even if the planned statistical analyses as reported in the Methods section are plausible and appropriate, sometimes the implementation of those statistical analyses, as reported in the Results section, is not. Several issues may have arisen in performing the analyses that render them inappropriate. Perhaps the most obvious occurs when the data do not demonstrate many of the properties that the authors anticipated when they planned their data analysis.

For example, although the authors may have planned to calculate a correlation between two variables, the data from one or the other (or both) of the variables may demonstrate a restriction of range that invalidates the use of a correlation. When such a strong restriction of range exists, the correlation is bound to be low, not because the two variables are unrelated, but because the range of variation in the particular data set does not allow for the expression of the relationship in the correlation.[4]

Similarly, the authors may have planned to use a *t* test to compare the means of two groups, but upon reviewing the data, they note a bimodal distribution that raises doubts about the use of a mean and standard deviation to describe the data set. If the use of a mean to represent the central tendency is questionable, then the use of a *t* test to evaluate the differences between the two groups becomes questionable as well.

The reviewer should be alert to these potential problems and ensure, to the extent possible, that the data as collected and presented continue to be amenable to the statistical analyses that were originally intended. Assessing this relationship is often difficult for a reviewer because the data necessary to make this assessment are not presented. When the opportunity does present itself, however, the reviewer should evaluate the extent to which the data collected satisfy the assumptions of the statistical tests that are described in the Methods section and reported in the Results.

Notably, in some cases, the authors may have deviated from established standards or violated assumptions for the use of a statistical test, yet the application and results of the test or tests may still be appropriate. For example, in most situations, statistical tests are making inferences against a hypothesized *population*, so the assumptions of the test should be compared against the authors' hypothesis about the population, not necessarily the sample. What is important is that the authors adequately justify their decisions. The reviewer, therefore, must be attentive to the degree to which the analysis is meaningful and believable in the context of the aims of the report.

In studies based on modeling approaches, drawing inferences requires comparing multiple possible models to determine the best fit to the observed data. Authors should report the appropriate or commonly used fit indices that assess the adequacy of the statistical model in describing the data. Given that authors have a choice of several possible indices, the primary index or indices should be declared a priori along with the standard values for an adequate fit. The fit of statistical models can vary depending on the index chosen, so the authors should give an adequate justification for design choices in the analysis.

Another potential error that the reviewer should be alert to is the possibility that although authors have selected appropriate analyses, they have executed them poorly or inappropriately. Often, enough data are presented for the reviewer to determine that the results of the analysis are implausible given the descriptive statistics—that is, the numbers just don't add up. One example is the reporting of a significant *t* test when the means do not seem sufficiently different to warrant significance. Alternatively, the authors may have reported data and analyses insufficiently for the reviewer to determine their accuracy or legitimacy. These situations are problems that a reviewer should address explicitly in his or her review.

## Interpreting multiple comparisons

In addition to determining whether the individual analyses are complete, appropriate, and properly executed, the reviewer should assess and discuss the authors' interpretation of the results. First, for analyses that rely on significance testing (i.e., *P* values) to draw inferences, the reviewer should be mindful that as the number of statistical tests increases, so too does the likelihood that at least one of the analyses will be statistically significant by chance alone.[5] When analyses proliferate, the reviewer must determine whether the authors have appropriately taken this possibility into

account in their interpretation of their results. As one example, significance levels (*P* values) might be adjusted to reflect the need to be more conservative, using techniques such as "family-wise alpha" corrections. While such prescriptive approaches are not strictly necessary (in fact, scholars have debated the extent to which such prescriptive adjustments of alpha are even appropriate[6,7]), the reviewer should be alert to the authors' excessive use of inferential tests without some explicit acknowledgment of the potential for overinterpreting one or two of the results they found to be statistically significant.

As a related issue, the reviewer might identify, in the Results section, analyses that were not anticipated in the Methods section. In practice, the results of an analysis or a review of the data often lead to other obvious questions, which in turn lead to other obvious analyses that the authors may not have anticipated. This type of expansion of analyses is not necessarily inappropriate, thus the reviewer must determine whether the authors have undertaken further analyses with control and reflection. A study design to evaluate a set of specific outcomes may not have the statistical power or design rigor to address these unexpected findings. If the reviewer perceives an uncontrolled proliferation of analyses or if the new analyses appear without proper introduction or explanation, then he or she should raise this concern in the review. The reviewer may feel that the authors have fallen into a trap of chasing an incidental finding too far or that they have enacted an unreflective or unsystematic set of analyses to look for anything that is significant. While these analyses may be sources of new hypotheses or interpretations of the data, each may also indicate a misuse of statistical techniques.

## Statistical versus practical significance

Notably, statistical significance or adequate model fit does not necessarily imply practical significance. Tests of statistical significance inform an investigator about the probability that chance alone is responsible for study outcomes *assuming the null hypothesis*[1]; however, whether the resulting findings are statistically significant or not, inferential statistical tests do not reveal the strength of the association among research variables or effect size. A preliminary mechanism for presenting the strength or degree of an association to the reader is to provide descriptive statistics such as means and standard deviations, or (when applicable) raw counts, to help contextualize the inferential statistics. Additionally, when authors are presenting their findings as parameters (e.g., means, odds ratios, correlation coefficients), it is often helpful to provide confidence intervals[8] for those parameters. While confidence intervals tend to become smaller with increasing sample size, they generally provide a measure of certainty in determining the importance and extent of differences.

However, the strength of a finding can also be gauged explicitly using indices of the proportion of variance in the dependent variable that is "explained" or "accounted for" by the independent variables in an analysis. Common indexes of explained variation include eta$^2$ ($\eta^2$)[9] in ANOVA, Cohen's $d$[10] in $t$ tests, and $r^2$ (coefficient of determination)[11] in correlational analyses. Other indices such as odds ratios and prediction parameters may be appropriate depending on the type of analysis conducted. The reviewer must be alert to the reality that statistically significant research results tell only part of the story. If a result is statistically significant, but the independent variable accounts for only a very small proportion of the variance in the dependent variable, the result may not be sufficiently interesting to warrant extensive attention in the Discussion section. If none of the independent variables accounts for a reasonable proportion of the variance, then the study may not warrant publication.

On the other hand, sometimes, authors use the absence of a statistically significant difference in comparative statistical analyses to justify equivalency between groups or interventions. Strictly speaking, equivalency is not an appropriate interpretation. Lack of statistically significant differences may result from a lack of power (an inadequate sample size). Although there are research designs to test for equivalency, such equivalency or noninferiority trials have a different set of assumptions and hypothesis-testing standards from traditional superiority trials.[12] Thus, if the authors are attempting to claim equivalency, they should clearly articulate the nature of the study and conduct analyses as appropriate.

## Reporting psychometric and measurement statistics

Often, health professions education studies addressing assessment or measurement report the reliability, validation evidence, and other parameters of new tools or procedures. The statistics reported in these types of studies come with their own set of reporting and interpretation concerns, which the reviewer must evaluate and address.

First, it is important to realize that there are many potential sources of error in assessment (e.g., between raters, between items, between iterations, between test forms).[13] The authors must indicate both that they understand which of these sources of error are relevant to a particular circumstance and that they have addressed these sources of error in their reliability analysis. Addressing the theoretically and contextually relevant sources of error is critical to the legitimacy of the authors' claims. Thus, the reviewer should attend to the nature of the measurement error captured by the statistics, then determine whether each source of error represented in the analysis is relevant to the research question and whether all sources of error relevant to the research question are represented in the analysis. Informative methods of representing these sources of error include, for one, reporting the variance components from such analyses and clearly stating the assumptions of fixed and random facets of the analysis, so the reviewer might look for these statistics as a part of his or her review.

Second, researchers now generally recognize that reliability is not a property of the test or procedure, but a property of the use of that test or procedure for a given population.[13] If an inappropriate population is used, then the reliability statistics might be highly misleading. This use of an inappropriate population for the reliability analysis sometimes arises when authors intentionally include groups with different levels of ability in order to

demonstrate the test's capacity to discriminate between the groups as evidence of its validity. If the authors then perform a reliability analysis on the full range of participants, the results will likely overestimate the reliability for any one of the groups represented in the study.

Finally, analyses that support evidence for validation vary, and no single analysis can provide conclusive evidence for the validation of an instrument's use.[14] As with other statistical analyses, the reviewer should attend to whether the assumptions and purpose of the analysis are consistent with and appropriate for the model, tool, or procedure being validated; the goal is to ensure that the claims for the evidence of validation are supported and appropriately restricted.

## Alignment with overall question and results

The reviewer must remember that statistical analyses are but one part of the argument presented in a research report and should consider the relevance and appropriateness of the reported analyses in relation to the main purpose of the report. An analysis section in the Results can be considered complete if it addresses the central research question as outlined in the Introduction and Methods sections. The reviewer's task, therefore, includes not only a determination of whether the reporting meets accepted standards, but also a judicious evaluation of the level of detail required by the research question and the degree of interest the details will hold for potential readers. In this sense, the Results section should not simply present a string of findings resulting from various analyses; rather, it should explain the analyses in a manner that allows a typical reader to appreciate the purpose of the statistics and to interpret them in light of that purpose.

*Acknowledgments:* The current authors acknowledge Georges Bordage, Addeane S. Caelleigh, and Ann Steinecke for their work on the previous version of this chapter.
*Funding/support:* None reported.
*Other disclosures:* None reported.

## References

1. Fisher RA. Statistical methods for research workers. Edinburgh, UK: Oliver and Boyd; 1925.
2. Hoyle RH, ed. Structural Equation Modeling: Concepts, Issues, and Applications. Thousand Oaks, CA: Sage Publications; 1995.
3. Streiner DL, Norman GR. Health Measurement Scales: A Practical Guide to Their Development and Use, 4th ed. New York: Oxford University Press; 2008.
4. Pearson K. Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variable and correlation of organs. Philosophical Transactions of the Royal Society, London Series A. 1903;2001:1–66.
5. Braun HI, Tukey JW. Multiple comparison through orderly partitions: The maximum subrange procedure. In: Wainer H, Messick S, eds. Principals of Modern Psychological Measurement: A Festschrift for Frederic M. Lord. New York, NY: Routledge Talor & Francis Group; 1983:55–68.
6. O'Keefe DJ. Colloquy: Should familywise alpha be adjusted? Hum Commun Res. 2003;29:431–447.
7. Tutzauer F. On the sensible application of familywise alpha adjustment. Hum Commun Res. 2003;29:455–463.
8. Neyman J. Outline of a theory of statistical estimation based on the classical theory of probability. Philosophical Transactions of the Royal Society London Series A. 1937;236:333–380.
9. Cohen J. Eta-squared and partial eta-squared in fixed factor ANOVA designs. Educ Psychol Meas. 1973;33:107–112.
10. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Hillsdale, NJ: Erlbaum; 1988.
11. Pearson K. Notes on regression and inheritance in the case of two parents. Proc Roy Soc Lond. 1895;58:240–242.
12. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: The importance of rigorous methods. BMJ. 1996;313:36–39.
13. Shea JA, Fortna GS. Psychometric Methods. In: Norman GR, van der Vleuten C, Newble D, eds. International Handbook of Research in Medical Education. Boston, MA: Kluwer Academic; 2002:97–126.
14. Messick S. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. Am Psychol. 1995;50:741–749.

# Chapter 15
# Results: Reporting Qualitative Findings

**Lara Varpio, PhD, and Glenn Regehr, PhD**

**L. Varpio** is associate professor, Department of Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland.

**G. Regehr** is professor, Department of Surgery, and associate director, Research, Centre for Health Education Scholarship, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, Canada.

## Issues and examples related to reporting qualitative findings

Evaluating the reporting of qualitative results is difficult because qualitative findings elude rule-based appraisals. Qualitative research encompasses many traditions such as Grounded Theory[1] (including constructivist variations[2]), phenomenology[3] (in all its iterations, including hermeneutic[4] and transcendental[5]), and discourse analysis[6] (be that analysis informed by critical theories, as in Foucauldian discourse analysis,[7] or by disciplines, as in linguistic discourse analysis[8]). Every tradition can incorporate a wide range of data sources, including conversations, documents, and observations, as well as a wide range of analytic techniques.

## Reporting formats

Given the variety in qualitative traditions, the presentation of qualitative findings can take many forms, making rigid rules for communicating results inappropriate.[9] For example, qualitative results often take the form of narrative excerpts (usually from interviews, focus groups, documents, or observations); however, other forms of data displays are equally valid, including (but not limited to) tables, figures, diagrams, and visual images. In other words, while narrative excerpts are commonly used, it is wholly appropriate for qualitative results to be presented in other formats—including numeric representations.

To illustrate, if a scholar wants to describe patterns or symmetries found in the complexity of qualitative data, he or she may find that those data are most efficiently expressed as "quantitizations"[10] and so would present them numerically. Quantitization could be used, for example, to report the patterns of informal education of residents in clinical settings,[11] or the number of conversational turns made by doctors, patients, and students in bedside teaching encounters.[12] If data are presented numerically, the reviewer must examine whether the numbers represent an effective elaboration or extension of the analysis, or an inappropriate reduction of data that is unjustified or incompatible with the qualitative tradition being used.

Furthermore, while narrative comments are often interwoven in the text of the Results section, placing narratives outside the text of the report into other structures, such as tables or figures, is also entirely permissible. This technique can allow authors of qualitative research reports to adhere to word-count restrictions while simultaneously sharing with the reader an overview of complex coding structures and thematic relationships. For example, authors may choose to summarize key findings,[13] to illustrate conceptual frameworks,[14] or to present the relationship between themes and codes across a data set.[15] Importantly, when the authors use such approaches, the reviewer must explicitly examine the relationship between the textual component of the Results section and any associated tables or figures. While the text, tables, and figures should not overlap to the point of redundancy, they should be consistent and complementary.

In summary, evaluating the presentation of results requires an appreciation for the variability that is inherent to qualitative scholarship. The challenge for the reviewer is to decide whether the manner in which the authors present the results remains internally consistent (e.g., across text, tables, and figures), aligns with the research approach as articulated in the Methods section, and provides the reader with a clear, appropriate, and complete understanding of the findings.

## Using data effectively as evidence of claims

In assessing the quality of the analysis, the reviewer should ensure that both the substance and the format of the findings provide sufficient evidence to support each claim presented in the report. The Results section should not consist of a mere listing of themes with illustrative data excerpts. To think that the narrative data "speak for themselves" is a mistake. Instead, the reviewer should determine the extent to which the authors have established the "claim-grounds-warrant"[16] connection in conveying their results—that is, in reviewing the Results section, the reviewer should ensure that the authors have provided a clear set of arguments or positions (i.e., *claims*). Claims that are clear and logically connected should offer a persuasive and coherent argument. Further, each claim should be substantiated with data (i.e., *grounds*) that support the claim. In other words, each position should be supported with evidence—each claim corroborated with grounds. Finally, each claim-grounds dyad should conclude with a statement that interprets the data, showing the reader how the data effectively support the claim (i.e., *warrant*). Effective warrant statements should provide reasonable and logical descriptions of how the data specifically support the claims made.

When considered this way, the Results section is much more than a "data dump" of quotations or data excerpts. The results should provide a breadth of concrete examples, paraphrases, and/or summaries—all described in enough detail and all of sufficient evocative strength to support the full extent of each claim presented in the report. Thus, the reviewer must attend to the warrants (i.e., the relationship between the claim and the grounds provided) to ensure that the authors have reported qualifications to and conditions of their claims.

## Clear, informative, and evocative results

Another concern that the reviewer should be mindful of is the clarity of reporting. Qualitative research delves into the complex social relationships and contextual factors that inform individuals' and groups' experiences. Even though qualitative research investigates such complexities, the communication of these intricacies must be focused and direct. Clarity should be supported by writing simply, but without inappropriately simplifying the experiences reported.[17] The reviewer must determine, therefore, whether the authors have provided enough detail in their results to evoke a nuanced understanding of the phenomena, social relationships, and/or contexts being studied,[18] without overwhelming or confusing the reader. Describing the intricacies and richness of qualitative research with clear, focused, simple language is an important evaluative criterion of the Results section.

## Reflexivity

Another concern in the reporting of qualitative findings is transparency. Qualitative findings are generally representations of the authors' selections of data as evidence, as well as their interpretations and contextualizations of the data.[9,17] Consequently, the qualitative findings presented in Results section are never "objective" in the traditional sense. In qualitative results, the researcher communicates findings as a coherent whole[17] that represents the complex reality studied *as perceived by the researcher or research team that undertook the study.* This aspect of qualitative results has given rise to the notion of *reflexivity*—a hallmark of high-quality qualitative research. Reflexivity refers to an attitude of systematically and intentionally attending to how the researchers influence all aspects of the research process, and thus the knowledge that is constructed through that process.[9,19,20]

Typically, reflexivity is maintained by researchers' declarations (usually, in the Methods section) of their beliefs, their personal characteristics, and their relationships with participants. However, reflexivity is also supported by organizing and sufficiently describing results in a way that allows the reader to "see a clear correspondence between the empirical data and the interpreted findings."[21(p 360)] In other words, one way for authors to realize reflexivity is to state explicitly the claim-grounds-warrant connections. Doing so requires that the Results section transparently describes how the authors transformed data into findings and then into a coherent whole. Such transparency has occurred when the reviewer—and the reader—can easily find answers to questions such as

- What are the qualities and characteristics that define the identified themes or recurring pattern?
- How have the authors defined their key terms (i.e., the qualities and characteristics of themes and/or patterns)?
- What relationships, if any, exist between the identified themes and/or patterns?

## Blending the Results and Discussion sections

Finally, in addition to examining the coherence, clarity, and quality of the Results section itself, it is important for the reviewer to determine whether the format used in the Results section connects in a meaningful way both to the research questions articulated in the Introduction[19] and to the elaborations of the findings offered in the Discussion section. In this regard, a final note should be made about how, in certain qualitative traditions, the line that divides the Results section from the Discussion section of the report is porous. Given that qualitative findings are already the result of interpretation and contextualization, many qualitative traditions blend the Results and Discussion sections.[9] Therefore, in the presentation of qualitative results, incorporating theory, other literatures, and conceptual frameworks is certainly acceptable.[9]

## Summary

As qualitative research becomes increasingly common and valued in medical education,[9] qualitatively savvy reviewers will be essential for ensuring the publication of high-quality reports.

*Funding/Support:* None reported.
*Other disclosures:* None reported.
*Disclaimer:* The views expressed herein are the authors' own and do not necessarily represent those of the Uniformed Services University of the Health Sciences, the U.S. Department of Defense, or the federal government of the United States.

## References

1. Glaser BG, Strauss AL. The Discovery of Grounded Theory: Strategies for Qualitative Research. Chicago: Aldine de Gruyter; 1967.
2. Charmaz K. Constructing Grounded Theory. London: SAGE Publications; 2006.
3. Speigelberg H. The Phenomenological Movement: A Historical Introduction. The Hague, The Netherlands: M. Nijhoff; 1960.
4. Heidegger M. Being and Time. New York: Harper; 1962.
5. Husserl E. Phenomenology and the Foundations of the Sciences. Boston: M. Hijhoff Publishers; 1980.
6. Jørgensen M, Phillips L. Discourse Analysis as Theory and Method. London: SAGE Publications; 2002.
7. Jäger S, Maier F. Theoretical and Methodological Aspects of Foucauldian Critical Discourse Analysis and Dispositive Analysis. In: Wodak R, Meyer M, eds. Methods of Critical Discourse Analysis, 2nd ed. Thousand Oaks, CA: SAGE Publications; 2009:34–61.
8. Brown G, Yule G. Discourse Analysis. New York: Cambridge UP; 1983.
9. O'Brien B, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: A synthesis of recommendations. Acad Med. 2014;89:1245–1251.
10. Sandelowski M, Voils CI, Knafl G. On quantitizing. JMMR. 2009;3:208–222.
11. Varpio L, Bidlake E, Casimiro L, et al. Resident experiences of informal education: How often, from whom, about what, and how. Med Educ. 2014;48:1220–1234.

12. Rees CE, Ajjawi R, Monrouxe LV. The construction of power in family medicine bedside teaching: A video observation study. Med Educ. 2013:47;154–165.
13. Kitto S, Goldman J, Etchells E, et al. Quality improvement, patient safety, and continuing education: A qualitative study of the current boundaries and opportunities for collaboration between these domains. Acad Med. 2015;90:240–245.
14. Kolehmainen C, Brennan M, Filut A, Isaac C, Carnes M. Afraid of being "witchy with a 'b'": A qualitative study of how gender influences residents' experiences leading cardiopulmonary resuscitation. Acad Med. 2014;89:1276–1281.
15. Ziller E, Balmer D, Hermann N, Graham G, Charon R. Sounding narrative medicine: Studying students' professional identity development at Columbia University College of Physicians and Surgeons. Acad Med. 2014;89:335–342.
16. Toulmin SE. The Uses of Argument. Cambridge: Cambridge University Press; 1958.
17. Sandelowski M, Leeman J. Writing usable qualitative health research findings. Qual Health Res. 2012;22:1404–1413.
18. Giacomini MK, Cook DJ, 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;284:478–482.
19. Malterud K. Qualitative research: Standards, challenges, and guidelines. Lancet. 2011;358:483–488.
20. Sandelowski M, Barroso J. Finding the findings in qualitative studies. J Nurs Scholarsh. 2002;34:213–219.
21. Giacomini MK, Cook DJ, 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;284:357–362.

# Chapter 16
# Discussion and Conclusion: Interpretation

**Sonia J. Crandall, PhD, MS, and William C. McGaghie, PhD**

**S. J. Crandall** is professor, Department of Physician Assistant Studies, Wake Forest School of Medicine, Winston-Salem, North Carolina.

**W. C. McGaghie** was director, Ralph P. Leischner, Jr. MD Institute for Medical Education, Loyola University Chicago Stritch School of Medicine, Maywood, Illinois, at the time this work was done. He is now professor of medical education, Northwestern University Feinberg School of Medicine, Chicago, Illinois.

## Review Criteria

1. ***The conclusions are clearly stated; key points stand out.***
2. ***The conclusions follow from the design, methods, and results.***
3. ***The study limitations are discussed.***
4. ***Findings are placed in the context of relevant literature, and alternative interpretations are considered as needed.***
5. ***Practical significance or theoretical implications are discussed; guidance for future studies is offered.***

## Issues and examples related to the criteria

Research follows a logical process. It starts with a problem statement and moves through design, methods, and results. Researchers' interpretations, recommendations, and conclusions emerge from these four interconnected stages. Flaws in logic can arise at any of these stages, and, if they occur, the authors' interpretations of the results will be of little consequence. Flaws in logic also can occur at the interpretation stage. The researchers may have a well-designed and well-executed study but obscure the true meaning of the data by misreading the findings.[1]

The reviewer needs to have a clear picture of the meaning of the research results. He or she should be satisfied that

- the evidence is discussed adequately and appears reliable, valid, and trustworthy;
- the interpretations are justified given the strengths and limitations of the study; and
- the generalizability and practical significance of the researchers' conclusions are apparent given the architecture, operations, and limitations of the study.

The organization of the Discussion section should match the structure of the Results section (which itself may mirror that of the Methods section). This alignment helps structure the report such that it presents a coherent interpretation of methods and data. The reviewer needs to determine how the Discussion relates to the original problem and research questions. Most important, the Discussion must be clearly written, justified by the results, and illustrate key points. Broadly, important aspects for the reviewer to consider include the following:

- whether the conclusions, based on the description of the results, are reasonable;
- how the study results relate to other research in the field—that is, how the results fit within the context of relevant literature (e.g., Do the findings build consensus? Do they conflict with the outcomes of previous studies? Are they unexpected?);
- whether the authors consider alternative interpretations or proffer any practical or theoretical implications;
- how the study outcomes expand the knowledge base in the field and how they inform and guide future research; and
- whether the authors have described limitations in the design, procedures, and analyses of their study. Failure to discuss the limitations of the study should be considered a significant flaw.

On a subtler level, the reviewer must evaluate whether the authors have explicitly distinguished between (1) inferences drawn from the study's results, which are based on data- analysis procedures and (2) extrapolations to the conceptual framework used to design the study (see Chapter 6). Interpreting data within the confines of the methods rather than drawing broad conclusions about the implications of the data marks the difference between formal hypothesis-testing and theoretical discussion.

### Quantitative approaches

When interpreting any hypothesis-testing aspects of a quantitative study, authors should discuss the meaning of both statistically significant and nonsignificant results. A statistically significant result, given its *P* value and confidence interval, may have no practical implications.[2,3] For example, statistically significant findings are important when deciding the merits of an educational intervention, but practical significance (i.e., the effect of the intervention) may be more important. Results of a large cohort study designed to investigate differences in the outcomes of two different educational approaches for pharmacology instruction may reach statistical significance ($P < 0.01$), but the actual effect size of the differences in outcomes may be quite modest (Cohen's $d = 0.3$). The course director may decide that the difference in outcomes does not merit changing the curriculum.

The reviewer should confirm that authors have explained, using their data as support, whether each hypothesis as outlined in the problem statement or design is confirmed or refuted. The reviewer should also look for a discussion of the magnitude of the effect of the outcomes when appropriate, and whether each finding agrees or conflicts with previous research. Authors should not introduce

new data or analyses in the Discussion section, nor should they interpret data or findings unless they have first presented them in the Results section.

While authors may occasionally inadvertently misrepresent or misinterpret data, errors come more often from overinterpreting the data from a theoretical perspective. For example, a reviewer may see a statement such as, "The sizable correlation between test scores and 'depth of processing' measures clearly demonstrates that the curriculum should be altered to encourage students to process information more deeply." The curricular implication may be true, but it is not supported by the data. Although the data show that encouraging an increased depth of processing is associated with improved test scores, they neither indicate the nature or direction of the relationship nor demonstrate the need to change curricula. The intent to change the curriculum is a statement of values or opinion based on a judgment about the utility of high test scores and their implications for professional performance. The connection between test scores and professional performance does not directly denote the need for curricular change.

The language used in the Discussion needs to be clear and precise. For instance, in research based on correlations, the Discussion must address the goals and underlying assumptions of the research design (e.g., whether the correlations derive from data collected concurrently or longitudinally).[4,5] Correlations over time suggest a predictive relationship among variables, which may or may not reflect the investigator's intentions. The language used to discuss such an outcome must be unambiguous.

## Qualitative approaches

Qualitative researchers must convince the reviewer that their data are trustworthy. To describe the trustworthiness of the collected data, authors may use criteria such as *credibility* (analogous to or resembling internal validity), *transferability* (resembling external validity), and *dependability* (resembling reliability).[6,7] Authors must also explain how they determined or addressed each of these three criteria.[6,7] (See Giacomini and Cook for a thorough explanation of how to assess the evidence for validity in qualitative health care research.[8]) A reviewer may check for credibility, transferability, and dependability by examining what procedures or methods the authors used to verify their data, as explained below.

Authors may determine *credibility* through data triangulation, member checking, and/or peer debriefing.[6,9,10] Triangulation is a method by which researchers compare multiple data sources. For example, authors may perform a content analysis of curriculum documents, transcribed interviews with students and faculty members, patient-satisfaction questionnaires, and observations of standardized patient examinations. Member checking is the process of "testing," or verifying, any interpretations and conclusions with the individuals who provided the data (e.g., participants in interviews, focus group members).[6] Peer debriefing entails an "external check on the inquiry process" by engaging disinterested peers whose qualities parallel those of participants to confirm or expand upon interpretations and conclusions.[6]

*Transferability* implies that the authors or others may use the research findings in other contexts (i.e., resembling generalizability).[7,9] Notably, authors reporting qualitative research cannot establish evidence for external validity (or reliability) through the same methods used in quantitative research.[6,7] The reviewer's task is to judge whether the authors' conclusions could transfer to other settings based on the authors' presentation and interpretation of the results as they relate to the study's context.

Authors may provide evidence for *dependability* by reporting or describing the processes and methods they used in sufficient detail for "a future researcher to repeat the work."[7]

## Biases

The reviewer must be aware of the biases, inherent in all research designs (i.e., qualitative, quantitative, mixed methods), that affect the generalizability and transferability of outcomes. For example, a researcher using a quantitative design may select an appropriate random sample of members of a national professional organization as the target audience for a membership-satisfaction survey. The survey may garner a low response rate, or the respondents may not represent the selected sample (e.g., a disproportional number of responses from men or a disproportional representation of geographic regions or of public versus private intuitions).

When judging qualitative research, the reviewer should carefully consider the meaning and impact of the authors' personal perspectives and values. He or she should verify that the authors have clearly explained any potential biases that are likely to influence the analysis and presentation of outcomes. Those biases include

- the influence of the researcher on the study setting,
- the purposive selection of research participants,
- the selective presentation and interpretation of results, and
- the thoroughness and integrity of the interpretation.

Peshkin's work is an excellent example of announcing one's subjectivity and its natural influence on the research process.[11] Qualitative research, based on interpretation of nonquantitative data, is not objective. Authors retain responsibility for explaining how their values, perspectives, and experiences may affect research outcomes.[12] Reviewers of qualitative research need to be convinced that the authors have suitably addressed the influences of subjectivity.

Because of the intrinsic nature of bias within all research, authors must openly discuss the limitations of their conclusions. Important elements that the reviewer should look for are limitations within the following:

- the particular study design the authors have chosen,
- the authors' strategies for recruiting (i.e., sampling) and retaining study participants,

- any procedures the authors used to determine group assignment or identify cases,
- the quality of the data-collection instruments,
- the means the authors have used to maintain quality control of data (e.g., how missing data are managed and accounted for in analyses), and
- the types of analyses the authors have conducted (i.e., statistical and nonstatistical).

## Discussion vs. conclusions

Finally, a note on the distinction between discussion and conclusions: Some journals request separate sections for discussion and conclusions, while others combine these into a single section. Whether these appear under distinct headings or combined into one section does not matter, provided the authors have covered all the necessary information. The Discussion section includes interpretation of study findings and recommendations for practice. The reviewer should expect the authors to connect their findings to the original intent of study and to the research questions or hypotheses. The Conclusions section provides an overall "big picture" synopsis of the study's results, articulates the importance or educational significance of these outcomes, and may include personal, theoretical opinions and recommendations or directions for future research.

*Funding/Support:* None reported.
*Other disclosures:* None reported.

## References

1. Day RA, Gastel B. How to Write and Publish a Scientific Paper. 7th ed. Santa Barbara, CA: Greenwood; 2011.
2. Goodman S. A dirty dozen: Twelve p-value misconceptions. Semin Hematol. 2008;45:135–140.
3. Kazdin AE. Clinical significance: Measuring whether interventions make a difference. In: Kazdin AE, ed. Methodological Issues and Strategies in Clinical Research, 3rd ed. Washington, DC: American Psychological Association; 2003.
4. Fraenkel JP, Wallen NE III, Hyun HH. How to Design and Evaluate Research in Education. 8th ed. New York, NY: McGraw-Hill; 2012.
5. Osborne JW. Correlation and other measures of association. In: Hancock GR, Mueller RO, eds. The Reviewer's Guide to Quantitative Methods in the Social Sciences. New York: Routledge; 2010.
6. Lincoln YS, Guba EG. Naturalistic Inquiry. Newbury Park, CA: Sage; 1985.
7. Shenton AK. Strategies for ensuring trustworthiness in qualitative research projects. Education for Information. 2004;22:63–75.
8. Giacomini MK, Cook DJ. 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;284:357–362.
9. Grbich C. Qualitative Research in Health. London: Sage; 1999.
10. Grbich C. Qualitative Data Analysis: An Introduction. London: SAGE; 2007.
11. Peshkin A. Places of Memory: Whiteman's Schools and Native American Communities. Mahwah, NJ: Lawrence Erlbaum; 1997.
12. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: A synthesis of recommendations. Acad Med. 2014;89:1245–1251.

# Chapter 17
# Title, Authors, and Abstract

**Georges Bordage, MD, PhD, William C. McGaghie, PhD, and David A. Cook, MD, MHPE**

**G. Bordage** is professor of medical education, Department of Medical Education, University of Illinois at Chicago, Chicago, Illinois.

**W. C. McGaghie** was director, Ralph P. Leischner, Jr. MD Institute for Medical Education, Loyola University Chicago Stritch School of Medicine, Maywood, Illinois, at the time this work was done. He is now professor of medical education, Northwestern University Feinberg School of Medicine, Chicago, Illinois.

**D. A. Cook** is professor of medicine and medical education and director, Office of Education Research, College of Medicine, Mayo Clinic, Rochester, Minnesota.

## Review Criteria

1. ***The title is clear, informative, and representative of the content.***
2. ***The abstract contains essential details.***
3. ***The conclusions in the abstract are justified by the information in the abstract and the text.***
4. ***There are no inconsistencies in detail among the abstract, text, tables, and figures.***
5. ***All the information in the abstract is present in the text.***

## Issues and examples related to the criteria

When a manuscript arrives, the reviewer immediately sees the title and the abstract, and in some instances—depending on the policy of the journal—the names of the author or authors. This triad of title, authors, and abstract is both the beginning and the end of the review process. It orients the reviewer, but it can be fully judged only after the manuscript is studied thoroughly.

### Title

The title can be viewed as the shortest possible abstract. In addition to providing information about the manuscript's content, the title must have appeal, prompting readers to study the report. Consequently, it needs to be clear and concise and accurately reflect the content and scope of the study. The reviewer should judge whether the title

- is too general or misleading,
- lends appropriate importance to the study, and
- grabs the reader's attention without overstating the conclusions.

Huth describes two key qualities of good titles: they should be "indicative" and "informative." The indicative aspect of the title tells the reader about the nature of the study (what did the authors do?), while the informative aspect presents the message derived from the study results (what did the authors find or conclude?).[1] In their study of 110 articles reporting experimental studies in medical education, Cook, Beckman, and Bordage found that only 10 percent of titles were both indicative and informative, suggesting substantial room for improvement.[2] To illustrate, consider this title: "A Survey of Academic Advancement in Divisions of General Internal Medicine." It tells readers what the authors did (it is indicative) but fails to convey a message (it is not informative). An informative title would read, "A Survey of Academic Advancement in Divisions of General Internal Medicine: A Slower Rate and More Barriers for Women." The subtitle now conveys the message, and the whole title still remains concise. Conversely, a title might inform readers of the central message without indicating what the authors did to arrive at the message. Subtitles such as "A Randomized Trial," "A Grounded Theory Study," or "A Systematic Review" can succinctly communicate the type of study the authors conducted. Indeed, some journals encourage subtitles.

### Authorship

The reviewer is not responsible for setting criteria for authorship. This is a responsibility of the editor and editorial boards. Moreover, some journals may keep the name of the authors from reviewers to prevent biased judgments. When information about authorship is available, the reviewer can help detect possible authorship inflation (too many authors), unwarranted authors (authors ineligible for that title), or "ghost authors" (missing authors).

The International Committee of Medical Journal Editors' (ICMJE) *Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals*[3] (previously the *Uniform Requirements*) report covers a broad range of issues and contains the most influential single definition of authorship:

> authorship [is] based on the following four criteria:
>
> - Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND
> - Drafting the work or revising it critically for important intellectual content; AND
> - Final approval of the version to be published; AND
> - Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.[3]

More than 1,550 biomedical journals, including *Academic Medicine*, have voluntarily endorsed the ICMJE recommendations, although not all of them officially follow this strict definition

of authorship.[4] Some journals use other authorship conditions (criteria) or set limits on the number of authors.

The number of authors per manuscript has increased steadily over the years, both in medical education and in clinical research.[5–8] Author inflation may indicate the increasing size of research teams[9] or new topics of inquiry[7]; however, author inflation may also signal a less benign phenomenon. A disproportionate number of authors on a manuscript may indicate some researchers' efforts to pad their curriculum vitae for promotion.[10] From a publication standpoint, overly expanding the list of authors for promotion purposes is "unauthorized" authorship and can be difficult to detect.

Journals generally publish their specific criteria for authorship to help scholars decide who should be included in the author list. Many journals also require each author to complete and sign a statement of authorship indicating his or her contributions to the manuscript. These contributions are typically listed in the published article.

Huth argues that individuals do not merit authorship if their involvement comprises only acquiring funds, collecting data, administering the project, or proofreading or editing manuscript drafts for style and presentation, not ideas.[11–14] Such contributions can be recognized in a footnote or in an acknowledgment. Other limited or indirect contributions that do not warrant authorship include providing access to or recruiting subjects (research participants), participating in a pilot study, processing data without conceptualizing or planning the analyses, or providing materials or research space.[15] Finally, granting authorship for "contributions" that are honorary is not appropriate; for example, department chairs, division chiefs, laboratory directors, or senior faculty members should not receive credit as an author for work in which they have had limited involvement.[16,17]

Conversely, some author lists fail to recognize individuals who have made substantial contributions to the work, so-called "ghost authors." Omission of qualified authors can arise either through unintentional oversight or through deliberate efforts to hide the contributions of someone whose involvement might discredit the work (e.g., a contributor with a conflict of interest).

Wislar and colleagues, for example, found that 21 percent (95 percent confidence interval: 18.0 to 24.3 percent) of articles published in six large-circulation general medical journals in 2008 had honorary (unwarranted) authors, ghost (warranted but unrecognized) authors, or both.[17] The reviewer may be able to indicate to the editor any suspicion that an author might not meet authorship criteria or that an individual has made substantial contributions without receiving credit as an author.

## Abstracts

Abstracts serve a double function: they "act as a reference tool (for example in a library abstracting service)," and they help readers "decide whether or not to read the full text."[18] Accordingly, the abstract should contain a brief but complete summary of the full research report. The abstract "is published in isolation from the main text and should therefore stand on its own and be understandable without reference to the longer piece. It should report the latter's essential facts, and should not exaggerate or contain material that is not in there."[18]

Abstracts vary in length (the recommended or maximum number of words) and format (unstructured narrative or structured abstracts). Structured formats vary widely but have in common the use of distinct, labeled sections.[19] Simple structured formats include four or five sections with headings such as Background (or Context), Purpose (or Objectives, Aims), Methods, Results, and Conclusions.

In the mid-1980s, scholars proposed a more extensive list of subheadings to provide "more informative abstracts."[20–22] The headings they suggested, based on published criteria for the appraisal of medical literature, included the following:

1. *Objective* (the exact question[s] addressed by the article),
2. *Design* (the basic plan for conducting the study),
3. *Setting* (the location, including the dates of the study, and type of clinical care or the level of training [education]),
4. *Patients or Participants* (the manner of selection and the number of persons who both entered and completed the study),
5. *Interventions* (the exact treatment, if any),
6. *Main Outcome Measures* (the primary study outcome measure or dependent variable),
7. *Results* (key findings), and
8. *Conclusions* (including direct clinical or educational applications).[20–22]

Sometimes, a ninth heading was added about *limitations*. Leading clinical journals quickly joined in adopting these more informative abstracts using either all the subheadings or some variation.[23] Distinct subheadings have since been developed for specific study designs, such as the PRISMA abstract format for review articles, which employs seven subheadings: Context, Objective, Data Sources, Study Selection, Data Extraction, Data Synthesis, and Conclusions.[24,25]

There is evidence that structured abstracts provide more information than unstructured abstracts.[26–28]

A journal's instructions to authors will specify the length and type of abstracts, which may vary depending on the article type. Some journals may view abstracts as simply promotional trailers designed to entice potential readers to read further, while others aim to create abstracts that are as informative as possible within the specified word limit. Even when the journal does not require a structure with specific headings, authors should provide as much information as possible to maximize utility to readers. The reviewer may use the list of headings (above), which are required

in the more informative, structured abstracts as a reminder of what information the abstract should include.

Investigators have found repeatedly that both structured and unstructured abstracts frequently omit substantial amounts of information.[2,29,30] Cook and colleagues found that many abstracts omit, in particular, clear statements about the study design, setting, specific results, and comparison intervention (for studies with a comparison group).[2] A study by Pitkin and Branagan showed that merely giving authors specific instructions about three types of common defects in abstracts—(1) inconsistencies between abstract and text, (2) information present in the abstract but not in the text, and (3) conclusions not justified by the information in the abstract—was ineffective in lowering the defect rates.[31] The reviewer plays a key role in identifying these and other flaws in abstracts.

*Funding/Support:* None reported.
*Other disclosures:* None reported.

## References

1. Huth EJ. Chapter 13. The first draft: Titles and abstracts—Types of titles. In: Writing and Publishing in Medicine, 3rd ed. Baltimore, MD: Williams & Wilkins; 1999:131–132.
2. Cook DA, Beckman TJ, Bordage G. A systematic review of titles and abstracts of experimental studies in medical education: More informative elements needed. Med Educ. 2007;41:1074–1081.
3. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. Updated December 2014. http://www.icmje.org/icmje-recommendations.pdf. Accessed March 13, 2015.
4. International Committee of Medical Journal Editors. Journals Following the ICMJE Recommendations. http://www.icmje.org/journals-following-the-icmje-recommendations. Accessed March 13, 2015.
5. Dimitroff A, Davis WK. Content analysis of research in undergraduate education. Acad Med. 1996;71:60–67.
6. Drenth JP. Multiple authorship: The contribution of senior authors. JAMA. 1998;280:219–221.
7. Papatheodorou SI, Trikalinos TA, Ioannidis JP. Inflated numbers of authors over time have not been just due to increasing research complexity. J Clin Epidemiol. 2008;61:546–551.
8. Sukotjo C, Yuan JC, Bordage G. A content analysis of dental education research as reported in two journals. J. Dental Educ. 2010;74:1106–1112.
9. Wuchty S, Jones BF, Uzzi B. The increasing dominance of teams in production of knowledge. Science. 2007;316: 1036–1039.
10. Cronin B. Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? J Am Soc for Info Sci and Tech. 2001;52:558–569.
11. Huth EJ. Chapter 4. Preparing to write: Materials and tools. In: Writing and Publishing in Medicine, 3rd ed. Baltimore, MD: Williams & Wilkins; 1999:41–44.
12. Huth EJ. Appendix A: Guidelines on authorship. In: Writing and Publishing in Medicine, 3rd ed. Baltimore, MD: Williams & Wilkins; 1999:293–296.
13. Huth EJ. Appendix B: The "uniform requirements" document: An abridged version. In: Writing and Publishing in Medicine, 3rd ed. Baltimore, MD: Williams & Wilkins; 1999:297–315.
14. Huth EJ. Guidelines on authorship of medical papers. Ann Intern Med. 1986;104:269–274.
15. Hoen WP, Walvoort HC, Overbeke AJ. What are the factors determining authorship and the order of the authors' names? A study among authors of the Nederlands Tijdschrift voor Geneeskunde (Dutch Journal of Medicine). JAMA. 1998;280:217–218.
16. Flanagin A, Carey LA, Fontanarosa PB, et al. Prevalence of articles with honorary authors and ghost authors in peer-reviewed medical journals. JAMA. 1998;280:222–224.
17. Wislar JS, Flanagin A, Fontanarosa PB, DeAngelis CD. Honorary and ghost authorship in high impact biomedical journals: A cross sectional survey. BMJ. 2011;343:d6128.
18. Emerald Group Publishing. How to … write an abstract: What is an abstract? http://www.emeraldgrouppublishing.com/authors/guides/write/abstracts.htm. Accessed March 13, 2015.
19. National Institutes of Health. U.S. National Library of Medicine. What are structured abstracts? Updated October 2014. http://www.nlm.nih.gov/bsd/policy/structured_abstracts.html. Accessed March 13, 2015.
20. Ad Hoc Working Group for Critical Appraisal of the Medical Literature. A proposal for more informative abstracts of clinical articles. Ann Intern Med. 1987;106:598–604.
21. Altman DG, Gardner MJ. More informative abstracts (letter). Ann Intern Med. 1987;107:790–791.
22. Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. More informative abstracts revisited. Ann Intern Med. 1990;113:69–76.
23. Nakayama T, Hirai N, Yamazaki S, Naito M. Adoption of structured abstracts by general medical journals and format for a structured abstract. J Med Libr Assoc. 2005;93:237–242.
24. Huth EJ. Structured abstracts for papers reporting clinical trials. Ann Intern Med. 1987;106:626–627.
25. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Ann Intern Med.* 2009;151:W65–W94.
26. Comans ML, Overbeke AJ. De gestructureerde samenvatting: Een hulpmiddel voor lezer en auteur. [The structured summary: A tool for reader and author.] Ned Tijdschr Geneeskd. 1990;134:2338–2343.
27. Sharma S, Harrison JE. Specialist Orthodontist, Saltaire Orthodontics, Bradford, United Kingdom.
28. Structured abstracts: Do they improve the quality of information in abstracts? Am J Orthod Dentofacial Orthop. 2006;130:523–530.

29. Taddio A, Pain T, Fassos FF, Boon H, Ilersich AL, Einarson TR. Quality of nonstructured and structured abstracts of original research articles in the British Medical Journal, the Canadian Medical Association Journal and the Journal of the American Medical Association. Can Med Assoc J. 1994;150:1611–1614.
30. Hartley J. Current findings from research on structured abstracts. J Med Libr Assoc. 2004;92:368–371.
31. Froom P, Froom J. Deficiencies in structured medical abstracts. J Clin Epidemiol. 1993;46:591–594.
32. Pitkin RM, Branagan MA. Can the accuracy of abstracts be improved by providing specific instructions? A randomized controlled trial. JAMA. 1998;280:267–269.

# Chapter 18
# Reviewing a Review Manuscript

**Colin P. West, MD, PhD**

**C. P. West** is professor of medicine, medical education, and biostatistics, Departments of Medicine and Health Sciences Research, Mayo Clinic, Rochester, Minnesota.

Scientific reports that summarize previous research form a special category of manuscripts with a number of unique characteristics. Some journals may seek reviewers with specific expertise in evaluating reviews; however, the general approach to reviewing these reports is similar to the approach to reviewing any other scientific work. In particular, the question of internal validity, or whether the review's methods sufficiently limit bias, is central. The recently expanded appraisal criteria set forth in the users' guides to the medical literature[1,2] are a useful starting point for the review of such manuscripts. Other checklists are available as well,[3,4] and these may be helpful to reviewers who want to ensure that they have identified the key strengths and weaknesses of a review.

## Definitions

Several types of reports present reviews of the literature. A *narrative review* is any summary of scientific literature, ideally addressing a focused educational question. Narrative reviews can be very useful in presenting evidence in a consolidated form, but they do not follow specific methodological approaches (e.g., a comprehensive search strategy, a criterion-based assessment of the quality of included studies, and a demonstration of reproducibility of decisions made in the course of the review), which are intended to reduce bias. Compared with reviews that closely follow the structured steps of a more methodological approach, there is a greater chance with narrative reviews that the authors have missed important studies or that they have not uncovered all relevant evidence. Further, there is a risk that the authors have consciously or unconsciously selected sources of evidence to support a particular perspective that may or may not accurately reflect the totality of the existing literature.

A *systematic review* is a summary that attempts to address a focused question using well-defined, predetermined methods designed to reduce bias. As suggested above, these methods include attention to thorough, expansive searches of the literature; to the quality of studies comprising the systematic review; and to presenting the methods of the study protocol with enough detail for it to be replicated consistently. A systematic review does not require quantitative synthesis of the evidence, and in many situations, such a quantitative synthesis is inappropriate.

Finally, a *meta-analysis* is a systematic review that applies quantitative methods to summarize and characterize or compare the results. A meta-analysis need not be based on an underlying systematic review, and some meta-analyses are, in fact, based on a narrative or other review. The reviewer should, during the appraisal process, confirm the type of review that supports the meta-analysis. Regardless of the type of review the authors have written, the reviewer can take the same general approach to evaluating it.

## Potential impact

An initial general review criterion is the potential influence of the manuscript (see also Chapter 8). This criterion applies equally to narrative reviews, systematic reviews, and meta-analyses. Notably, a review may not meaningfully add to the literature even if its methods are highly rigorous. The review should address an educationally relevant question, and the authors should clearly identify this question in the manuscript. If the question under study is too narrow, the source literature may be sparse, and the scientific impact may be small. If the question is too broad, it may not be possible to offer a summary that is useful.

For example, a systematic review might address whether online teaching modalities aid learning, a question that will seem quite important to many in health professions education. If the authors limit this review to studies of online learning courses in a single discipline within a single university system, they may find extremely few relevant resources, and the larger relevance of any reported results may be questionable. On the other hand, if they review all articles on online learning modalities for learners across all ages and fields, their summary is unlikely to provide data relevant to any individual group.

Once the reviewer has assessed whether the question the review addresses is important and whether the review truly extends or consolidates knowledge in an area of inquiry, he or she should evaluate the review's validity, results, and educational applicability. The sections that follow address each of these topics.

## Internal validity

The reviewer should consider several criteria when evaluating the internal validity of a literature review—that is, how well the authors have controlled possible sources of bias.

The first criterion related to internal validity to consider is the thoroughness of the search strategy. Narrative reviews will generally fare poorly on this criterion, but at times, the field of study will be so new or the literature so clearly known that even without an exhaustive search strategy, the review may be appropriately inclusive. Consistent with the goal of completely representing the literature, a strong search might include multiple databases, reference lists, tables of contents of relevant journals, meeting abstracts, and even items published in foreign languages. Notably, smaller, negative studies are less likely to be published, and this publication bias may lead to overestimates of effect. Unfortunately, methods to assess for publication bias have limitations, commonly including a lack of power to identify missing studies. It is more important for authors to consider and

note the possibility of publication bias than to apply any specific approach to assess for its presence.

Second, the studies included in the review should themselves be of high methodological quality. Although systematic reviews are often thought of as providing a very high level of evidence, they can be only as strong as the source literature they include. A well-executed systematic review of high-quality studies provides a very different level of evidence than a well-executed systematic review of low-quality studies. Therefore, the authors should carefully evaluate and report the quality of the contributing literature. Well-established guides can help with the quality assessment of randomized trials (e.g., the *Cochrane Handbook for Systematic Reviews of Interventions*[5]), of observational studies (e.g., the Newcastle-Ottawa Scale[6]), and of studies of diagnostic test performance (e.g., QUADAS-2[7]).

Third, the report should provide evidence that the decisions the authors made in the course of conducting their review are reproducible. This potential reproducibility means, for example, that at least two authors should evaluate inclusion and exclusion decisions for a systematic review. Further, they should report their level of agreement regarding these decisions. A low level of agreement may suggest that the protocol lacked clarity. This criterion also means that a reviewer should view a systematic review by only a single author with skepticism since the reproducibility of the results emanating from methodological decisions made in the course of the review cannot be assessed.

## Results

For a narrative or systematic review, at least a descriptive summary of the evidence should be present. This summary should be educationally useful but should not extrapolate beyond the data from the individual contributing studies. A discussion of heterogeneity across studies—do results differ from study to study?—is particularly important. Discussing and speculating about why results differ is an essential element of a sound systematic review and can offer important insights into future research necessary to further advance a field.

If study results differ too widely or if the study question is too broad, pooling the study results quantitatively in a meta-analysis may not be reasonable. For example, as noted previously, a single pooled summary estimate of the impact of online learning interventions on knowledge likely makes little sense. Because online learning interventions are so diverse, their effects almost certainly differ across types of learners, and there are innumerable different types of knowledge and ways of measuring knowledge. Authors should clearly state the rationale underlying their decision either to perform or to defer a meta-analysis.

Authors should consider and explore results that do not align among studies, and many statistical measures of heterogeneity are available. Study results that differ greatly offer authors an opportunity to explore (often through formal subgroup analyses) possible explanations for this heterogeneity. Some of explanations will be evident before the systematic review is initiated (e.g., the type of online learning intervention affects the effectiveness of the learning) and should appear in the Methods section of the study, but even post hoc analyses can be useful in generating hypotheses for future research. Differences in results from studies of varying quality are particularly important to consider.

The reviewer should ensure that the authors of a meta-analysis have clearly reported the methods that support the pooling of study results, including pooling of outcomes from different study designs, intervention categories, learner types, and assessment measures and instruments. The authors must make many decisions in the course of a systematic review, and their rationale for each decision should be clear. Likewise, the authors should explicitly present the absolute magnitude of any summary effects rather than just measures of relative effect. Finally, the authors should provide confidence intervals to allow the reader (and the reviewer) to evaluate the precision of both the source literature and the review's summary estimates.

## Applicability

The audience to which a review and its conclusions are relevant is important. If the authors quantitatively pool their results, those results should be applicable to a clearly defined target audience. In addition, although a review should ideally provide a comprehensive discussion of all educationally important outcomes, it is common—given the complexity of a good systematic review—for authors to focus on only one key outcome. Narrative reviews may address a wider range of outcomes, but this expanded scope should be balanced against the potential validity limitations of the narrative review. Regardless, the reviewer should carefully consider the scope of the review and its conclusions.

## Summary

Reviews have become an increasingly important part of the scientific literature. Although the methods specific to sound reviews—and to systematic reviews and meta-analyses in particular—may be complex, the basic approach to appraising these studies parallels the approach to reviewing any scientific work. Unique considerations include attending to the quality of the individual contributing studies in the review and evaluating the heterogeneity of results from different studies. Still, impact, validity, results, and applicability remain central to the effective assessment of reviews.

*Funding/Support:* None reported.
*Other disclosures:* None reported.

## References

1. Guyatt G, Rennie D, Meade MO, Cook DJ. Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice, 3rd ed. New York, NY: McGraw-Hill; 2015.
2. Murad MH, Montori VM, Ioannidis JP, et al. How to read a systematic review and meta-analysis and apply the result to patient care: Users' guides to the medical literature. JAMA. 2014;312:171–179.

3. Department of General Practice, University of Glasgow. Critical appraisal checklist for a systematic review. http://www.gla.ac.uk/media/media_64047_en.pdf. Accessed March 12, 2015.
4. PRISMA 2009 Checklist. http://www.prisma-statement.org/2.1.2%20-%20PRISMA%202009%20Checklist.pdf. Accessed March 12, 2015.
5. Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. The Cochrane Collaboration. Updated March 2011. http://www.cochrane-handbook.org. Accessed March 12, 2015.
6. Wells GA, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of Nonrandomized studies in meta-analyses. Ottawa Hospital Research Institute. 2014. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp. Accessed March 11, 2015.
7. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155:529–536.

# Chapter 19
# Reviewing Descriptions of Innovations

**Christopher S. Candler, MD, EdD, and Michael T. Fitch, MD, PhD**

**C. S. Candler** is professor of medicine and senior associate dean for academic affairs, University of Oklahoma College of Medicine, University of Oklahoma, Oklahoma City, Oklahoma.

**M. T. Fitch** is professor and vice chair for academic affairs, Department of Emergency Medicine, Wake Forest School of Medicine, Winston-Salem, North Carolina.

Innovations are an important part of the future of academic medicine. They have the potential to change practice, inform future research, and, ultimately, improve the human condition. Technologies we take for granted today, such as high-fidelity patient simulation, once started as novel ideas that grew into innovations.

Virtually any innovation can be thought of as an attempt to solve a specific problem. In the academic medicine community, problems may arise from a variety of challenges in one or more missions of medical schools or teaching hospitals. An innovation to address any given challenge—such as inadequate student diversity, decreasing faculty retention, or ineffective resident-to-resident handoff of patients—may take the form of a program, a course, an activity, a process, or a policy.

Many faculty members working within medical education desire scholarly recognition for an innovation they have developed, or they may wish to share their potential solution with colleagues at other institutions. These faculty members may submit a manuscript describing their innovation to a health professions education journal for consideration. *Academic Medicine* has developed an article type—the innovation report—for just such manuscripts.[1] This chapter describes a general framework for evaluating the quality of descriptions of innovations in journals.

Reviewers who evaluate descriptions of innovations should recognize the fundamental difference between these manuscripts and research reports. While research reports evaluate a research question associated with a particular intervention (which may also consist of an innovation), descriptions of innovations explain the environmental context that led to the conceptualization, development, and implementation of the innovation. Unlike typical research reports, these descriptions are not required to prove success or present an analysis of outcomes-based data; however, the authors of reports describing innovations should still use a scholarly approach and reflect on the success and potential impact of the innovation.

Notably, reports or descriptions of innovations should not be limited to the novel technical characteristics of an innovation. Instead, the authors should systematically describe the entire process they used to identify, research, plan, implement, and evaluate the innovation. They must also include in the report an appropriate description of both the context in which the innovation was required and the perceived need it filled or the problem it addressed.

## A framework for reviewing descriptions of innovations

Because reports describing innovations are qualitatively different from research reports, a qualitatively different rubric is needed to evaluate these manuscripts. In a 2008 *Academic Medicine* editorial,[2] Dr. Steven Kanter offered a framework for authors to use when describing innovations. His editorial, "Toward Better Descriptions of Innovations," outlined nine criteria that authors should use to describe the planning, development, and implementation of their innovation so that it will be generalizable to other readers. With some subtle modifications, these criteria can form the basis of an evaluation rubric that reviewers can use to assess the scholarly value of descriptions of innovations, including any innovation reports submitted to *Academic Medicine*.

To enhance the potential application of these concepts by reviewers, we have condensed Kanter's nine criteria into five (see also Table 1):

1. description of the problem,
2. exploration of potential solutions,
3. implementation of the innovation,
4. critical analysis of the quality of the innovation, and
5. assessment of the innovation's potential impact.

The recommended format and word-count limit for some innovation reports may limit the authors' ability to comprehensively address all five of these criteria in every circumstance. In these cases, the reviewer must judge whether the authors have appropriately expanded their focus on some criteria and provided more limited information for others, depending on the stage of implementation.

## Description of the problem

Compared with traditional research manuscripts, reports describing innovations do not require a specifically constructed hypothesis or research question. Instead, descriptions of innovations should include a brief but complete account of the problem or challenge that the authors sought to address through the development of the innovation. The authors may explicitly state the desired outcome (e.g., higher standardized exam scores, sufficient faculty diversity), or the outcome may be implicit in the problem description itself (e.g., decreasing scores on the U.S. Medical Licensing exam, inadequate representation of Native Americans among the basic science faculty). While problem

Table 1

## A Framework for Reviewing Descriptions of Innovations

| Criteria for scholarly reports describing an innovation | Elements in a comprehensive description |
|---|---|
| A description of the problem the innovation addresses | • A brief but complete description of the problem<br>• A clearly articulated expected outcome of the innovation<br>• Information on the context, "age," importance, and generalizability of the problem |
| An exploration of potential solutions | • An explanation of the process for identifying potential solutions<br>• A literature review<br>• A description of the conceptual framework<br>• An explanation of the process for selecting the final chosen solution |
| A description of the implementation of the innovation | • Details about preparing the innovation and the setting<br>• A list of identified stakeholders<br>• A description of communication and promotion strategies<br>• A note on required resources and financial considerations<br>• A description of identified barriers<br>• The timeline for full implementation |
| A critical analysis of the innovation and its implementation | • Reflections on quality<br>• A summary of any available outcomes data<br>• Lessons learned<br>• A description of methods for assessing success<br>• A description of unexpected failures |
| An assessment of the impact of the innovation | • A speculation about the potential impact of the innovation beyond its first initial application<br>• A consideration of adoption by other programs<br>• Ideas about potential research studies using the innovative methods<br>• A consideration of direct implications for practice improvement or health care policy<br>• A discussion of scalability and sustainability |

descriptions may take many forms, the authors should, at a minimum, help readers (and the reviewer) understand the context in which the problem has occurred, how long the problem has existed, the importance of the problem, and how common the problem is within the academic medicine community.

Authors should begin by summarizing the local or national conditions that led to both the development and the identification of the problem. In addition, they should describe the scope and limits of the problem within their organization: Which individuals, departments, processes, or outcomes are affected? How does the problem affect stakeholders? Have the authors received internal or extramural funding to resolve the problem?

As mentioned, the authors should help readers understand how long the problem has existed. Is it a chronic problem that has persisted for years or decades? Has the problem been only recently identified, or is this a future problem that the community will soon encounter?

The authors should convince the reader that the problem is important and generalizable to the larger medical education community. Innovations that address minor or inconsequential

problems will be of little or no interest to readers. To make the case that the problem is important, the authors should summarize relevant data that convey the extent and severity of the problem. Further, the authors may speculate on the consequences of *not* solving the problem. The authors' case for generalizability should include a summary of any national or multi-institutional data demonstrating that the problem is applicable to institutions beyond the authors' setting. Such data show that the innovation described is applicable to other institutions, individuals, or conditions.

## Exploration of potential solutions

Many problems can be addressed by a variety of innovative solutions; however, an innovative solution to a problem at one institution may not be suitable for a similar problem at another institution. The authors should describe the process by which they identified potential solutions. This process typically involves a literature review, a description of any theoretical or conceptual framework they used, and an explanation of any assumptions and constraints that limited the available solutions. The description of how the authors explored possible solutions should be followed by a description of the decision-making process they used to select the innovation that they ultimately applied to address the problem.

**Literature review.** An appropriate literature review should reflect the authors' evaluation of published materials and their understanding of how others previously addressed comparable problems in similar settings. The authors may summarize the search terms they used and publication databases they accessed. Descriptions of innovations should report the specific methods, interventions, and techniques—including any theoretical or conceptual frameworks—that others have used to address the problem at hand. The authors should describe the strategies others used to evaluate the success of similar innovations, and they should compare these with their own assessment tools.

**Theoretical or conceptual frameworks.** A thorough description of an innovation includes information about the theoretical or conceptual framework or frameworks the authors used to organize or characterize potential solutions (see also Chapter 6.) Such frameworks can help readers understand the conceptual basis for the innovation, including any theories that guided the early exploration process.

**Assumptions and constraints.** The authors should briefly describe any assumptions or constraints that limited the types of potential solutions they considered. Because certain assumptions may otherwise go unrecognized, the authors should discuss how they identified all the significant assumptions that guided the identification of potential solutions. The reader should also understand how constraints affected the number and types of available solutions. For instance, did financial, human resource, or technological constraints limit the types of innovations that could be used to address the problem?

**The decision-making process.** Finally, the reader should understand the decision-making process that led to the selected innovation. Who were the decision makers, and which criteria did they use to select the most appropriate innovation? Why did they exclude other potential solutions? What generalizable and local factors affected the decision? What influence did organizational dynamics or political perspectives have on the decision-making process?

## Implementation of the innovation

Once the authors have adequately explained the potential solutions they explored, they should describe the implementation of their chosen approach, their innovation. This description should include details about not only the specific actions taken but also the stakeholders involved, the communication and promotion strategies deployed, and the resources (including number and type of faculty, software, and other resources) required. The authors should summarize the implementation timeline.

The level of detail needed in a descriptive report of an educational innovation should be tailored to the type of innovation, how novel the approach is, and how difficult it may be for readers to consider implementing similar initiatives at their own institutions. Truly novel approaches may require very specific descriptions of each step while a lower level of specificity may be acceptable for more familiar approaches.

Importantly, the authors should be clear about any barriers they encountered when they were implementing their innovation. The challenges associated with putting any new or different approach into practice are well recognized; an honest description of any barriers and effective strategies for overcoming these are important for readers who may want to adopt the innovation. The authors should identify any specific features, administrative supports, or services that supported implementation and mitigated barriers to implementation.

Although financial considerations are typically not the primary focus for most descriptions of innovative approaches, authors should acknowledge that certain solutions may be associated with additional costs. In some cases, costs may take the form of actual monetary investments, and these expenditures are important to report. Descriptions of innovations should report other potential costs, such as reallocations of faculty effort, investments of staff time, and other intangible costs.

Some innovative approaches to solving complex problems will be accompanied by new curricular materials. The authors should discuss the development, review, and deployment of new materials, courses, educational modules, videos, or other resources that were essential for the success of the innovation. Such products may be suitable for peer review and wide dissemination through online resources such as MedEdPORTAL Publications.[3] (The non-peer-reviewed MedEdPORTAL iCollaborative[4] could be another venue for some materials.)

## Critical analysis of the quality of the innovation

An important aspect of demonstrating a scholarly approach to launching a new educational innovation includes critical reflection on the quality of the innovation and what was learned as a result

of implementation. This sort of reflection includes lessons learned along the way and the processes used to determine whether the innovation successfully addresses the stated problem. In many circumstances, initial pilot projects will reveal aspects of an idea that worked well in addition to elements that did not go as planned. It is important for the authors to describe the unexpected failures of their new technique and what steps they took to improve the innovation or its implementation for any future iterations. A critical reflection of the implementation process often uncovers institutional factors or other preexisting conditions that either facilitated or impeded an innovation's initial success; descriptions of these factors will allow readers to plan for a successful implementation in their own environment.

The critical analysis should describe the authors' approach to assessing the success of the innovation. As mentioned, a formal analysis of outcomes-based data is unnecessary for descriptions of an innovation; nonetheless, the authors should present any preliminary outcomes data (e.g., learner performance gains) they have obtained either during or after the implementation. Some descriptions of innovations will not include any data; instead, they may provide an explanation of alternative criteria (e.g., learner feedback, peer evaluation) for measuring success. The authors may also wish to consider how the innovation can be formally studied as the basis of future research reports.

Some innovations may lead to broad changes after successful deployment, either immediately or over a period of time. In these circumstances, the authors should take the opportunity to reflect on how the implementation affected the stakeholders and the institutional environment. Such reflection allows readers to understand the potential generalizability of the authors' innovation.

An essential component of critical evaluation and scholarly reflection is demonstrating an understanding of how any new innovation relates to other established methods. Authors should provide appropriate citations to published articles and reports that may be similar or relevant to the current description. An appropriate literature review or discussion of other resources and educational approaches that have been used previously will help establish the context for the new innovation. The authors should reflect on the key technical or environmental features that distinguish their innovation from similar approaches.

## Assessment of the innovation's potential impact

Although explaining the implementation and evaluation criteria for an innovation is essential, it is even more important for the authors to report the potential impact of their innovation beyond their own uses of it at the time of publication. Some innovations may lead to only limited changes outside of the local environment, while others may readily facilitate widespread transformation. Impact may be considered in terms of

- potential adoption by other programs,
- research studies that use the innovative methods to evaluate educational (or other) outcomes,
- direct implications for improvements in practice, or
- changes in health care policy.

The authors should speculate on the extent to which the described innovation may potentially change practice.

Local impact or regional adoption of a new innovation not only supports its potential for widespread use, but also allows insights into any preconditions needed for successful deployment in another setting. Reporting such insights may include providing information on how others may need to use modified approaches at different institutions with different stakeholders. It may be helpful for the authors to consider whether the innovation would be most successfully deployed in other environments as an incremental improvement or as the basis of a more transformative change.

Finally, the authors should address the scalability and sustainability of their innovation. They should describe how the innovation will be or has been institutionalized locally, and they should list any resources required for its ongoing success. Suggesting next steps for implementing the innovation on a broad scale (e.g., among larger groups of learners, among additional learner types, or with increased frequency) is important, as is anticipating any necessary modifications.

Innovative approaches to problem solving in academic medicine are essential for continuing to improve education in the health professions. Publishing scholarly descriptions of such novel ideas, tools, and programs is an effective way to disseminate these successful ventures throughout the academic community. Authors and reviewers can use the guidelines discussed here for evaluating (or writing) manuscripts that describe such innovations.

*Funding/Support:* None reported.
*Other disclosures:* Dr. Candler is editor-in-chief and Dr. Fitch is deputy editor of MedEdPORTAL.

## References

1. Sklar, DS. Sharing new ideas and giving them wings: Introducing innovation reports. Acad Med. 2013:88:1401–1402.
2. Kanter SL. Toward Better Descriptions of Innovations. Acad Med. 2008;83:703–704.
3. Association of American Medical Colleges. MeEdPORTAL Publications. 2015. www.mededportal.org. Accessed March 12, 2015.
4. Association of American Medical Colleges. MedEdPORTAL iCollaborative. 2015. www.mededportal.org/icollaborative. Accessed March 12, 2015.

# Chapter 20
# Presentation and Documentation

**C. Jessica Dine, MD, MSHPR, and Judy A. Shea, PhD**

**C. J. Dine** is assistant professor of medicine, Division of Allergy, Pulmonary, and Critical Care, and associate program director, Internal Medicine Residency Program, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

**J. A. Shea** is professor of medicine and associate dean, Medical Education Research and Assessment, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

## Review Criteria

1. ***The text is well written and easy to follow.***
2. ***The manuscript is well organized.***

## Issues and examples related to the criteria

*Presentation* refers to the clarity and effectiveness with which authors communicate their ideas. In addition to evaluating how well the researchers have constructed their study, collected their data, and interpreted important patterns in the information, the reviewer needs to evaluate whether the authors have successfully communicated all these elements. Ensuring that ideas are properly presented is another factor the reviewer must consider when assessing papers for publication. Clear, effective communication takes different forms. This chapter will discuss several elements that the reviewer should consider in assessing the clarity of the manuscript.

## The text is well written and easy to follow

The reviewer should be able to grasp the substance of the manuscript without having to work any harder than necessary. Of course, some ideas are quite complex and require both intricate explanation and great effort to comprehend, but too often, simple ideas are dressed up in complicated language without good reason. The reviewer must both consider how well the authors have matched the level of communication to the complexity of the substance in their presentation and ensure that even complex ideas are presented clearly.

Poor presentation may, in fact, reflect poor content. A description of a study's method that is incomprehensible to the reviewer may hint at the authors' own confusion about the elements of their study. Jargon-filled conclusions may reflect authors' inability to apply their data to the real world. Poor writing and jargon do not always reflect confusion; rather, some excellent researchers are simply unable to transfer their thoughts to paper without assistance. Sorting these latter authors from the former is a daunting task, so the reviewer should consider and evaluate *both* the presentation of the study and the elements related to methodology, analyses, and interpretation.

Notably, many international authors submit manuscripts to journals, and the reviewer might note when the language in a report seems to indicate that English is not a first language for the authors. The reviewer may recommend that the authors have a native English speaker review the manuscript for usage, grammar, and vocabulary before they submit a revision.

## The vocabulary is appropriate

The writing should not be complicated by inappropriate vocabulary such as excessive jargon, inaccurately used words, undefined acronyms, or new, controversial, or evolving vocabulary. Special terms should be defined when they first appear in the text, and the vocabulary chosen for the study and presentation should be used consistently. Authors should also avoid informal language (contractions, trendy words) and trite expressions. The vocabulary should be appropriate for the audience of the journal; authors of reports in more niche journals may use more specialized language (without explaining terms), while the language in journals with broader audiences should be more general.

## The content is complete and fully congruent

All information contained in the text should be clearly related to the topic. The sum of the sections should allow the reader not only to understand how the findings add to the current literature but also to be able to reproduce the study. Each section of the manuscript should be comprehensive and help explain why the topic is important and relevant to the scientific community. Each section should relate to the study question—by explaining the design used to seek an answer to the question, outlining the specific methods and analyses used to garner data related to the question, exploring or interpreting new data in light of the question, and relaying how the findings fit into the context of the broader literature.

## The manuscript is well organized

Clarity is also a function of a manuscript's organization. In addition to following a required format, such as IMRaD (Introduction, Methods, Results, and Discussion), a manuscript's internal organization (sentences and paragraphs) should follow a logical progression that supports the topic. The Introduction typically provides a review of the relevant literature (including, preferably, an explicit description of a conceptual model) and explains the purpose of the study. The Methods section should clearly delineate the steps the authors took to complete their study, including how they collected, analyzed, and interpreted data (although, in general, the actual interpretation should be saved for the Discussion section). The information (i.e., the findings) relayed in the Results section may align with the order of the methods presented and may be displayed through graphs or tables. Finally, the authors should summarize and explain their findings in the Discussion/Conclusion section of their manuscript, and this section may mirror the reporting of findings. Each section should complement, relate to, and lead logically to the next.

### The data reported are accurate (e.g., the numbers add up) and appropriate; tables and figures are used effectively and agree with the text and the abstract

The reviewer's task includes checking to see that the authors have presented their data accurately across the text and abstract and in all exhibits (e.g., tables, figures, lists, charts). A reviewer should note any contradictions or errors. In addition to simply confirming that the data are correct and corresponding throughout, the reviewer should assess whether all the results presented are relevant and necessary to address the proposed question(s). A reviewer may suggest removing text, tables, or figures if the information presented is irrelevant or has already been presented in a different format.

The reviewer should be prepared to evaluate exhibits or graphic representations of information. When well done, these present complex information succinctly, communicating ideas that would take too many words to tell; in fact, visual displays may be the most efficient way to convey complex ideas. Exhibits should be self-explanatory, so that the reader does not need to refer to the text to understand them. Ideally they require only short descriptive legends or titles. Tables, lists, and figures should not simply repeat information that is already presented in the text, nor should they introduce data that are not accounted for elsewhere in the text. The information in exhibits should complement, never contradict, information given in the text and abstract. The reviewer may also identify exhibits or data in exhibits that are not necessary and, in these cases, should recommend not including these in the final manuscript. Importantly, the authors (and the reviewer) should be alert to any specific instructions (e.g., the number of exhibits allowed, the use of color) about tables and figures provided by the journal.

Some journals now accept manuscripts that use only or mainly graphics to illustrate important ideas in medical education—for example, *Academic Medicine*'s Last Page column. In the August 2014 issue of that journal, Xierali and colleagues review the current data on the diversity of the current physician workforce within the United States, using primarily graphics to illustrate the potential impact of increasing this diversity.[1] This one-page infographic illustrates the idea that graphics can supplement other material—or even, on their own, illustrate important concepts.

Finally, as the use of electronic formats has increased, many journals allow authors to provide supplemental digital files to complement the text. The reviewer should access these and provide any comments about the content and appropriateness of this material.

### Reference citations are complete and accurate

The reviewer's evaluation of the presentation of the manuscript should extend to the presentation of references. Proper documentation ensures that the source of material cited in the manuscript is accurately and fully acknowledged. Further, accurate documentation allows readers to quickly retrieve the referenced material. Finally, proper documentation allows for citation analysis, a measure of, for example, the times a published article is cited in subsequent articles. Journals describe their documentation formats in their instructions to authors. The International Committee of Medical Journal Editors' (ICMJE) *Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals* also details suggested formats.[2]

Reviewers should not concern themselves with the specific details of a reference list's format; instead, they should check whether the in-text citations seem to match the appropriate references noted in the References list and whether the bibliographic information (e.g., authors' names, title, journal, date, volume, page number) for each reference appears accurate, complete, and current. The reviewer is not expected to actually check each reference.

### Summary

The reviewer should assess the manuscript not only on the substance of the content but also on how the content is presented. Because ideas are necessarily communicated through words and pictures, presentation and substance often seem to overlap. (As much as possible, the substantive aspects of the criteria for this section are covered in other sections of this guide.)

The extent to which a reviewer must judge presentation depends on the journal. As mentioned in Chapter 5, "Manuscript Revision and Final Editing," some journals (e.g., *Academic Medicine*) employ editors who work closely with authors to clearly shape text and tables; the reviewer, then, can concentrate on the substance of the study. Other journals publish articles more or less as authors have submitted them; in those cases, the reviewer's burden is greater. The reviewer is not expected to edit the papers, but his or her comments on the narration and exhibits may help authors address any presentation problems before final acceptance.

*Acknowledgments:* The current authors would like to acknowledge Gary Penn and Ann Steinecke as contributors to previous versions of this work.
*Funding/Support:* None reported.
*Other disclosures:* None reported.

### References

1. Xieral IM, Castillo-Page L, Zhang K, Gampfer KR, Nivet MA. AM Last Page: The urgency of physician workforce diversity. Acad Med. 2014;8:1192.
2. International Committee of Medical Journal Editors. Preparing for Submission: References. Updated September 2014. http://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html#g. Accessed March 17, 2015.

# Chapter 21
# Scientific Conduct

**Louis Pangaro, MD, and William McGaghie, PhD**

**L. Pangaro** is professor and chair, Department of Medicine, F. Edward Hébert School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland.

**W. C. McGaghie** was director, Ralph P. Leischner, Jr. MD Institute for Medical Education, Loyola University Chicago Stritch School of Medicine, Maywood, Illinois, at the time this work was done. He is now professor of medical education, Northwestern University Feinberg School of Medicine, Chicago, Illinois.

## Review Criteria

1. ***Ideas and materials of other authors are correctly attributed. (There are no instances of plagiarism.)***
2. ***Prior publication by the author(s) of substantial portions of the data or study is appropriately acknowledged.***
3. ***Any apparent conflict of interest is appropriately disclosed.***
4. ***There is an explicit statement of ethical review and approval (e.g., by an institutional review board [IRB]) for studies directly involving human subjects or data about them.***

The reviewer provides an essential service to editors, journals, scholars, and society by identifying issues of ethical conduct, which apply either to editorial concerns (such as the behavior of authors) or to the protection of human subjects (participants), that are implicit in manuscripts.[1]

## Editorial concerns

The following issues fit into the first major area for consideration, what may be called "editorial concerns": authorship, plagiarism, the falsification of data, the misrepresentation of publication status, the deliberate omitting of pertinent results (including negative or nonconfirming findings) or of relevant work by others, and conflicts of interest.

**Issues of authorship**. Issues of authorship cover whose name appears in the byline. Many journals follow the four criteria defining "Who is an author?" put forth by the International Committee of Medical Journal Editors (ICMJE), which are

1. contributing substantially to the "conception or design of the work; or the acquisition, analysis, or interpretation of data for the work,"
2. composing or "revising [the work] for important intellectual content,"
3. agreeing to the final draft of the manuscript, and
4. "be[ing] accountable for all aspects of the work," specifically its integrity and accuracy.[2]

Defining who is responsible for the material in the manuscript is often reflected in the explicit statement of each author's role that some journals require; otherwise, the authors' roles may not be apparent to the reviewer. (See also Chapter 17.)

**Plagiarism**. Plagiarism is attributing others' words or ideas to oneself. A related editorial concern is not correctly attributing ideas and insights—even if the authors have not explicitly or implicitly attributed these to themselves. The reviewer may detect, and should point out to the editor in the space for confidential comments, any specific ideas and specific words that the authors have not properly credited to others or cited. Being familiar with the topic covered in the manuscript and the pertinent literature is helpful. A sudden shift in the character of the narration (in the vocabulary, tone, verb tenses, etc.) may also signal plagiarism. While the plagiarism-detection software now used in some editorial offices may facilitate the detection of quotations taken verbatim or almost verbatim from references, the reviewer should still remain alert for text or ideas that may not be those of the authors.

**Falsifying data**. Falsifying data to make the study seem more successful, such as reporting results inaccurately or incompletely, or even fabricating results, is hard for the reviewer to detect unless the results strain credulity. Reviewers may be most effective at looking for inconsistencies or contradictions, as discussed in the three chapters on presenting data (Chapters 13–15).

**Misrepresenting publication status.** Misrepresenting publication status occurs when authors submit for publication the same or largely the same research that they have either previously published or have submitted to another journal.[3] Duplicate publication of the same material is easier to detect when the manuscript authors' names are revealed to the reviewer. The reviewer cannot usually tell whether parts of the study under review have already been published or detect when part or all of the study is also under consideration or "in press" with another journal. Some reviewers conduct an Internet search of the manuscript topic and/or, when authorship is not masked, of the authors themselves. A scan of the search results may uncover prior or duplicate publication and also aid in a general review of citations. The reviewer should mention any concerns to the editor, who then can decide how to proceed.

Dual publication of the same data set in a different journal, which may be called "self-plagiarism," is unethical. However, publication of the same paper in another language may be acceptable if the original article is cited and the appropriate permissions are obtained.

**Omitting pertinent results**. The deliberate removal or withholding of relevant results from a study so as to describe them in another manuscript and, thereby, achieve multiple publications ("salami-slicing") is not always apparent. The reviewer should look for explicit phrasing in the Methods and Results sections that make

clear whether the authors have reported all observations obtained in the study in the manuscript. Doing so allows the reviewer to judge whether displacement of data from the present manuscript is actually appropriate, either due to the length of the article or due to the distinctiveness of the questions being addressed.[4,5] Notably, it is not unusual for authors to cite their own work in a manuscript's list of references; it is the reviewer's responsibility to determine the extent and propriety of these citations earlier.

**Suppression of negative results.** Just as the reviewer should be aware of any inadvertent omission of relevant, similar results, he or she should be alert to authors' conscious omission, or suppression, of negative results. While a negative study—that is, one with conclusions that do not confirm the authors' hypothesis—may be valuable if the research question is important and the study design rigorous, authors may not have the confidence to include results that do not support their hypothesis or current conventional wisdom. The reviewer should be alert to this possibility and read carefully to detect the omission of expected data. A "negative" study may warrant publication, and the reviewer should not quickly dismiss such a paper without full consideration of the study's relevance and methods.[6]

(Notably, nowhere in this guide for reviewers is there a criterion that labels a "negative study" as flawed because it lacks a "positive" conclusion.)

**Omission of others' work.** Important and relevant research previously published by others may have been unconsciously omitted. Authors are prone to honest omissions in their reviews of prior literature and may have gaps in their awareness of others' work, so omissions do not always result from a conflict of interest. However, authors might suppress previous research to enhance the importance of their own work or to make it seem more innovative. The reviewer should point out missing citations and attributions to the authors and mention any concerns to the editor.

**Conflict of interest.** Finally, the reviewer should be alert to several sources of conflict of interest. Likely, the most familiar one is hoped-for material gain for the authors if the outcomes of a study lead to a specific conclusion. In their scrutiny of methods (see Chapters 9–12), the reviewer should be aware of any factors or circumstances that may interfere with the integrity of research. Financial interest in an educational project may not be apparent, but the reviewer should look for an explicit statement concerning financial interest when any marketable product (such as a software program or simulation model) is either the subject of an investigation or a means of conducting the study. Such an "interest" does not preclude publication. However, the reviewer should expect a clear statement, explaining that there is no commercial interest or how such a conflict of interest has been handled.

In addition to material gain, conflicts of interest may include competition for funding, position, or credit. Such a conflict may be manifested in a failure to acknowledge the prior work of others on the same subject (as mentioned above) or even in the misrepresentation of others' work to characterize it as misguided, or as a fallacy to be refuted by the authors of the manuscript at hand.

## The protection of human participants

The principles of human subject protection, as judged by the local institutional review boards (IRBs), have recently been reviewed.[7] The ethics of studies using human subjects (hereafter, "participants") has become a national issue.[8] Interpretations of regulations for the protection of human participants have expanded such that they apply to areas of research at universities and academic medical centers that had not received prior attention.[9] For instance, a study investigating a new educational experience based on a "clinical research" model by comparing trainees who experience the clinical model against an appropriate control group might reveal that one of the two groups received a less valuable educational experience. Hence, obtaining informed consent of participants, putting into place other protections (e.g., anonymizing and safeguarding personal data), and allowing participants to experience the other instructional method after the formal study is over would be expected of the authors. These are often required for approval by an IRB,[10] and the reviewer should make a point to look for an explicit statement in the manuscript that IRB approval has been obtained.

**Quality improvement vs. research.** Curriculum managers often collect data systematically about student performance to judge the success of specific instructional methods. These data are collected for quality improvement (QI), so the collection may be exempt from IRB review or some consent requirements. However, once investigators decide to disseminate these data as scholarship, they do not themselves have the authority to declare their work exempt from IRB review.[11] Investigators may decide to disseminate their results post hoc—that is, after the data have already been collected—and ask later for their study to be exempt from IRB review. Seeking IRB approval retroactively may represent a form of deceit. In health care QI, post hoc IRB approval remains an unsettled issue; however, some courts have concluded that QI projects do, in fact, constitute research.[12]

Thus we recommend that investigators proactively ask IRBs to consider whether the studies described in journal articles may be regarded as QI projects or research. The reviewer should look for a statement of IRB review anytime a manuscript details the participation of human participants—and mention the absence of such a statement in his or her comments.

**Recruiting learners as participants.** The procedures for recruiting learners (e.g., medical students, house officers, trainees in allied health professions) for educational research protocols deserve special attention because educator-investigators may not be aware of their conflicts of interest.[13] The investigator who is recruiting learners or other participants for a study should not be the potential participants' immediate supervisor (e.g., the clerkship director who is responsible for student assessment and grading, the department chair who performs faculty evaluations). Participants must not be at risk if they decline to participate in a study. It is, of

course, the IRB that should attend to this possibility, but different IRBs may approach the same educational research protocol problem in different ways.[14] For example, if investigators hoped to change the curriculum and to disseminate the results of the change, they should have approached the IRB *pro*spectively for guidance,[7] and this should be apparent to the reviewer.

Additionally, some qualitative research methodologies, such as structured interviews, could place a study participant at risk if unpopular opinions could be attributed to him or her. Here again, the researcher retains the ethical and legal responsibility to protect participants and seek IRB review. We anticipate that most health professions journals require statements about IRB approval in all research papers.

## Summary

Manuscript authors should meet standards of ethical behavior in both the processes of conducting and publishing their research. Any field that involves human participants in investigations—particularly fields in the health professions—should meet the ethical standards for such research, including the new requirements for education research. The reviewer should be alert to both editorial concerns and protecting human participants, and should highlight any potential issues for the editor. In so doing, the reviewer fulfills an essential function in maintaining the integrity of academic publications.

*Funding/Support:* None reported.
*Other disclosures:* None reported.
*Disclaimer:* The views expressed herein are the authors' own and do not necessarily represent those of the Uniformed Services University of the Health Sciences, the U.S. Department of Defense, or the federal government of the United States.

## References

1. Caelleigh AS. Role of the journal editor in sustaining integrity in research. Acad Med. 1993;68(9 suppl):S23–S29.
2. International Committee of Medical Journal Editors. Defining the Role of Authors and Contributors. 2015. http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html. Accessed March 18, 2015.
3. LaFolette MC. Stealing into Print: Fraud, Plagiarism, and Misconduct in Scientific Publishing. Berkeley, CA: University of California Press, 1992.
4. Karlsson J, Beaufils P. Legitimate division of large data sets, salami slicing and dual publication, where does a fraud begin? Knee Surg Sports Traumatol Arthrosc. 2013;21:751–752.
5. Norman G. Data dredging, salami-slicing, and other successful strategies to ensure rejection: Twelve tips on how to not get your paper published. Adv Health Sci Educ Theory Pract. 2014;19:1–5.
6. Chalmers I. Underreporting research is scientific misconduct. JAMA. 1990;263:1405–1406.
7. Johansson AC, Durning SJ, Gruppen LD, Olson ME, Schwartzstein RM, Higgins PA. Perspective: Medical education research and the institutional review board: Reexamining the process. Acad Med. 2011;86:809–817.
8. Office of the Secretary, U.S. Department of Health and Human Services. The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. April 1979. http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html. Accessed March 18, 2015.
9. Department of Health and Human Services. Code of Federal Regulation, Title 45, Public Welfare, Part 46—Protection of Human Subjects. 2009. http://www.hhs.gov/ohrp/policy/ohrpregulations.pdf. Accessed March 18, 2015.
10. Casarett D, Karlawish J, Sugarman J. Should patients in quality improvement activities have the same protections as participants in research studies? JAMA. 2000;284:1786–1788.
11. Henry RC, Wright DE. When do medical students become human subjects of research? The case of program evaluation? Acad Med. 2001;76:871–875.
12. Lynn J. When does quality improvement count as research? Human subject protection and theories of knowledge. Qual Saf Health Care. 2004;13:67–70.
13. Dyrbye LN, Thomas MR, Papp KK, Durning SJ. Clinician educators' experiences with institutional review boards: Results of a national survey. Acad Med. 2008;83:590–595.
14. Dyrbye LN, Thomas MR, Mechaber AJ, et al. Medical education research and IRB review: An analysis and comparison of the IRB review process at six institutions. Acad Med. 2007;82:654–60.

# Part 3

# Chapter 22
# Reviewer's Recommendation

**C. Jessica Dine, MD, MSHPR, and Judy A. Shea, PhD**

**C. J. Dine** is assistant professor of medicine, Division of Allergy, Pulmonary, and Critical Care, and associate program director, Internal Medicine Residency Program, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

**J. A. Shea** is professor of medicine and associate dean, Medical Education Research and Assessment, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

Here, we discuss the reviewer's task of summarizing the overall suitability of a manuscript for a journal. No one formula for developing an overall judgment exists and, in the absence of a fatal methodological flaw, arriving at a judgment means each reviewer must weigh the manuscript based on a wide range of review criteria, including those presented in this guide. In the end, the reviewer is expected to provide a constructive reply to the authors and a clear, consistent, and useful recommendation to the editor (who makes the ultimate decision about whether to go forward with a manuscript).

Part 2 of this guide presents many criteria for reviewers to use as guideposts or reminders as they evaluate a research manuscript. For some of the criteria, the reviewer may assess the manuscript at hand through either a dichotomous approach (e.g., yes/no, has it/does not have it) or a rating scale that incorporates similar terms as anchor points (e.g., 1 = explains clearly and thoroughly, 5 = does not include). Using the criteria is somewhat akin to scoring a test and marking each item "right" or "wrong," or assigning "partial credit" or "not answered."

Once the evaluation phase is done, the task that remains is to use the responses to the criteria to make a recommendation for or against publication of the manuscript. Making this recommendation is also similar to scoring a test—only in this task, instead of comparing the score against a rubric, the reviewer is setting the standard, deciding how much is enough. Did the authors meet enough of the criteria, or the *right* criteria, to lead to a recommendation either to publish or to ask for revisions? Or, conversely, did the work contain certain flaws or show an accumulation of shortcomings that collectively mandate a recommendation to reject?

## Aspects of the overall recommendation to the editor

At least two features of the reviewer's overall recommendation deserve explicit attention: it is a suggested or proposed decision, and it is a judgment. Chapter 4, "Publication Decision," describes processes used by the editor for making the final decision. The main point is that the reviewer's recommendation is just that: a recommendation. The final decision belongs to the editor. The reviewer's recommendation may not be followed. (Notably, when the editor's decision is contrary to the reviewer's recommendation, it seldom means that the reviewer's assessment was poor.)

The editor almost always wants a clear, reasoned recommendation from the reviewers—and a recommendation requires a judgment.

## Deciding on a recommendation

Reviewers are invited to review a manuscript on the basis of their expertise and prior work (see Chapter 2). They are expected to know enough to address the criteria (or to seek outside help for technical issues) and then to meld all their thoughts together into an overall recommendation. Determining the recommendation is not as easy as counting the favorable and unfavorable marks on the list of criteria and deciding whether the count (i.e., score) is high enough for a "publish" recommendation. Almost never do—or should—all items count equally. Compiling a list of minor and major flaws, along with a list of possible fixes, may help the reviewer determine whether any particular flaw is amenable to revisions and whether the manuscript warrants publication. However, in general, a research report represents a noncompensatory system—that is, a series of strengths cannot offset a fatal flaw or a series of important shortcomings.

For example, a new and highly creative way to assess medical students' clinical skills should have a very strong Methods section so that the reviewer can glean a good understanding of what the authors did. The reviewer cannot ignore other parts of the manuscript, but he or she might not weigh these (either implicitly or explicitly) so heavily. On the other hand, a technical manuscript that focuses on modeling the decision-making processes of residents and novices who are facing new versus familiar tasks would likely need very strong Introduction and Data-Analysis sections. How a reviewer puts all the pieces together and makes a final judgment is a very individual process.

## A recommendation to accept the manuscript

A recommendation to "accept as is" is relatively rare but does happen—it is the logical outcome when the reviewer gives high ratings on most or all the review criteria. A reviewer's job is not to be unduly critical, and when an excellent manuscript is spotted, the reviewer should mark it accordingly and point out to the editor the strong features that make it so appealing. Reviewers should not restrict themselves to negative comments!

## A recommendation to seek revisions

When the decision is not so clear, a list of minor and major flaws can be quite helpful. Typically, the balance of the scores on the criteria should roughly align with the final recommendation. It is not helpful to the editor when the majority of the criteria are marked high and the recommendation is not to publish. As mentioned, many journals give reviewers the option of writing confidential comments to the editor that will not be shared with the authors. These comments are where the reviewer should explain the lack of congruency between criteria scores

and the recommendation. The opposite scenario is probably more common: a reviewer provides a low rating for many criteria, provides well-grounded constructive criticism, and then recommends publishing or asking for revisions.

It appears to be very difficult for a reviewer to actually recommend "reject." This reluctance is understandable from the perspective of the reviewer, who is likely to have received his or her own share of reject letters and may be quite familiar with the angst such letters cause. But the primary commitment of the reviewer is to help the editor (and, by extension, the reader and the literature). Often, the recommendation itself is not passed on to the authors. Even when it is, if the recommendation aligns with the individual marks about the manuscript, the authors may be better able to understand the recommendation.

A middle ground of recommending revisions is certainly appropriate for many manuscripts, but the reviewer must use this recommendation cautiously—it should not be used to soften or postpone what will almost certainly be a final rejection decision. A "revise" recommendation is most helpful to the editor and to the authors when it is accompanied by explicit comments about what needs to be addressed to improve the manuscript based on the *existing* methods and data. In other words, it usually is not realistic to ask the authors to change what has already occurred during the study phase. Manuscripts can be rewritten for clarity, different analyses can be done, more details on methods can be given, conclusions and interpretations can be amplified or reduced, and the authors can be directed to previous literature that was omitted. But, in general, the authors—and reviewers—need to work with what is there.

## A recommendation to reject the manuscript

Occasionally, a manuscript will have a fatal flaw. When this happens, the recommendation is easy. But fatal flaws do not happen often; usually, the reviewer will, as discussed, have to weigh the flaws against the strengths.

If, after careful consideration, the reviewer is convinced that major weaknesses cannot be fixed, (e.g., using an uncommon, nongeneralizable experimental case), then a reject recommendation is in order. Aside from the presence of readily apparent fatal flaws, or even multiple nonfatal flaws, there are many reasons to make a reject recommendation. For example, a study might be very well designed, executed, and described but fail to add anything to the existing literature. Or it might be inappropriate (or not interesting) to the journal's readership. A rejection is also appropriate when there is something wrong with the manuscript (e.g., the study objectives and the study methods do not align; the results do not match the conclusion; the authors have systematically ignored disconfirming literature; the outcomes seem too perfect given the design, methodology, analyses, and limitations).

## Confidential comments

The reviewer needs to be clear about which of his or her comments may be passed on to the authors. As noted, many journals give reviewers the option of writing confidential comments to the editor. It is an option, not a requirement. If the reviewer has summarized the major positive and negative issues in a manuscript, there may not be a need to write anything else. Occasionally, however, the reviewer will have remaining comments or insights that he or she does not wish to share with the authors. As with other communications to the editor, comments are most helpful when they are clear and explicit. Ideally, these confidential comments expand on or complement points already made to the authors so that the editor is not left with a review in which the comments to the authors do not match the recommendation.

## Summary

In the end, the recommendation requires a judgment with which the reviewer is comfortable. It is a judgment, but it is not capricious. It requires the integration of many evaluations and reflections; intuition and gestalt are both valid parts of the formula. Reviewing becomes easier with experience, but it is rarely straightforward. If a reviewer goes into the process with the understanding that the goal is to be helpful to the editor and fair to the authors without being unduly critical, she or he will be poised to succeed.

*Funding/Support:* None reported.
*Other disclosures:* None reported.

# Part 4

# Chapter 23
# Reviewer's Etiquette

**Elizabeth S. Karlin, MA**

**E. S. Karlin** is staff editor, *Academic Medicine*, Association of American Medical Colleges, Washington, D.C.

The reviewer's assigned task is to assess the quality of the manuscript—from the title to the references—but reviewing goes beyond recommending additional analyses, expanding the discussion, or suggesting new references. The reviewer has an obligation to the editor and to the authors to fulfill his or her duties in an ethical, respectful, and timely manner. Reviewing scientific research before it is available to the public is both a service and a privilege.[1] Some argue that scholars in a particular field have a duty to contribute to that field however they are able; one such contribution is providing critical assessments of new research. Reviewing requires those who undertake the task to uphold confidentiality and follow the procedures set forth by the journal, but it also affords them an opportunity: to shape science and to gain insight into the latest research.

## Confidentiality, plagiarism, and intellectual property

The reviewer must recognize that the works submitted to journals are under embargo. They should maintain the strictest confidence about the research they are reviewing. They should not contact the authors of the work.[1] They should not share—with colleagues, students, patients, family, or others—the manuscript itself, data or information from the manuscript, or even the fact that the manuscript exists.[1] There are rare exceptions to this rule: a more senior reviewer may want to mentor a colleague who is new to reviewing, or—vice versa—a junior faculty may want a colleague's expertise on reviewing or on a topic. In such cases, the reviewer must contact the editorial office and seek formal permission. The reviewer who originally received the invitation should work with the additional reviewer and approve the review submitted to the editorial office.

The reviewer should not retain, copy, or in any way disseminate manuscripts he or she has reviewed or is reviewing. He or she must respect the intellectual property of the authors and the embargo of the journal and must never plagiarize the language of the manuscript or in any way use any information gleaned from it in advance of its publication.

## Accepting an invitation to review

The invited reviewer should review all the information in an invitation thoroughly and accept or decline it within two or three business days. Most reviewer invitations include the title and abstract of the manuscript to be reviewed; some also include the byline. Reviewers should read each of these carefully to make sure they have the expertise (methodological or subject matter, or both) to review the content well and to make sure they do not have a conflict of interest (i.e., that they do not know the authors or the work personally; see below). Most invitations also include a deadline. Reviewers should look at the deadline and their calendars before agreeing to review. They should respond to invitations honestly. The editor would much prefer to spend time finding another appropriate reviewer early in the review process than to find out weeks after a reviewer has accepted an invitation that the he or she does not have enough time or expertise to complete the review, after all.

Of course, once a reviewer has accepted an invitation to review, an unexpected conflict, emergency, or additional obligation may arise. The reviewer should simply alert the editorial office as soon as possible if this happens. If the reviewer cannot accept an invitation or complete a review, providing another potential name to the editorial office and explaining that he or she remains willing to review in the future are appreciated.

## The content of the review

Most journals offer reviewers the opportunity to provide comments for the authors as well as confidential comments for the editorial office. The reviewer should aim to give the authors and the editor substantive feedback. If the reviewer recommends revisions, he or she should offer clear, specific, actionable guidelines for improving the work or a specific aspect of it (e.g., the methodology, literature review, or discussion). Line-item or sentence-level edits are far less helpful, especially if the authors eventually decide to substantially rewrite or delete the sentences involved.

In general, the majority of the comments should be written for the authors. Comments will help the editor make decisions and may sometimes help the authors understand those decisions. Confidential comments to the editor should never be contrary to the comments for the authors, but they provide an opportunity for the reviewer to share concerns privately, offer informal impressions, elaborate on a recommendation to publish or not to publish, remark on the appropriateness of the manuscript for the journal, and disclose personal biases (see below).

Many journals offer reviewers an opportunity to recommend a decision to reject a manuscript, reconsider it after revisions, or accept it in its current form. Typically, the reviewer is asked to indicate their recommendation separately from their comments to the authors in case it differs from the final publication decision, which is strictly the purview of the editor. A reviewer should not explicitly state his or her views about whether a manuscript should be rejected, reconsidered after revisions, or accepted in the comments to the authors; these comments can be made, as mentioned, in the confidential comments to the editor.

If the reviewer recommends a rejection of the report, the reviewer should focus his or her comments to the authors on clearly, and respectfully, explaining what aspects of the manuscript were most troublesome or irreparable. Why is the reviewer not in favor of

publishing the manuscript as is, or in favor of giving the authors a chance to revise? The reviewer should consider providing commendations for the authors as well: What aspects of the work could the authors use in a new or revamped study? Where should the authors go from here?

If the reviewer is recommending a full accept and not recommending any changes, he or she should, again, explain why. What is important or new about the authors' work? What should readers or the editor gain from it? What about the work makes it so important, so high-quality, so worthy of publication?

The reviewer should remember—even when he or she feels that the manuscript does not add to the literature, the methods are flawed, or the project is derivative—that an author or group of authors has invested substantial time, effort, and, often, heart into the project. The tone and vocabulary of the review should be academic and—while not necessarily extremely formal or staid—certainly collegiate and courteous. The purpose of the review is to improve the work at hand, the content of the journal, and the literature and science overall, never to embarrass or assail authors or their work. As a general rule, the reviewer should record only comments for the authors in the same tone and vocabulary he or she would use in a face-to-face conversation with a colleague.[1] Some journals require the reviewer to sign his or her review in part to reduce the chances of conveying inappropriate tone and vocabulary to the authors.

The reviewer should never misrepresent the authors' work to the editorial office.[1] For example, if the authors' outcomes or conclusions contradict the findings or opinions of the reviewer, the reviewer should not denigrate the manuscript to keep the findings from being published.

## Conflict of interest and bias

Conflicts of interest arise from professional and personal affiliations. Minimally, reviewers must disclose conflicts of interest to the editor (this is an excellent use of the confidential comments-to-the-editor box) and, in many cases, must recuse themselves from reviewing. Reviewers should decline an invitation to review if they

- work in the same institution as any of the authors,
- are family or personal friends of the authors,
- have been involved in some aspect of the work covered in the manuscript,
- are in some way competing with the authors (for a publication, promotion, position, or funding), or
- have conflicting financial or legal relationships (e.g., with industry, consultancies, stock ownership, honoraria, expert testimony).[1]

As a general rule, if a reviewer is in doubt about a potential conflict, he or she should decline the invitation to review or should seek guidance from the editorial staff.

Reviewers should disclose personal biases they have for or against the authors' work, theories, cited literature, analyses, or methodology. If the reviewer thinks it will not be possible to provide a fair review based solely on the merit of the research, he or she should decline the invitation. Reviewers should be aware of subtle biases they may have for or against the authors' institution or nation,[2] for positive or affirming (as opposed to negative) findings,[3] or for impressive names and institutions. Reviewers should never use their reviews as a medium for espousing their own beliefs, theories, published works, or agenda.[1]

*Acknowledgments:* The author of this chapter is indebted to the authors who wrote the parallel chapter in the 2001 first edition. Their ideas remain relevant, and their words are hard to improve upon.
*Funding/Support:* None reported.
*Other disclosures:* None reported.

## References

1. Bland C, Calleigh A, Steinecke A. Chapter 4: Reviewer's etiquette. Acad Med. 2001;76:954–955.
2. Link AM. US and Non-US submissions: An analysis of reviewer bias. JAMA. 1988:280:246–247.
3. Callaham ML, Wears RL, Weber EJ, Barton C, Young G. Positive-outcome bias and other limitations in the outcomes of research abstracts submitted to a scientific meeting. JAMA. 1998:280:254–257.

# Afterword
# Peer Review Now and in 2030?

**Jan D. Carline, PhD, and Steven J. Durning, MD, PhD**

**J. D. Carline** is professor, Biomedical Informatics and Medical Education, Departments of Family Medicine and Pharmacy, University of Washington, Seattle, Washington.

**S. J. Durning** is professor, Medicine and Pathology, Department of Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland.

The basic criteria and procedures in this updated guide essentially follow the thrust of the previous edition; they represent guidelines for providing fair and high-quality reviews of research manuscripts in the growing community of scholars in the health professions. Through a series of conversations with many of the authors and with members of the "Review Criteria for Research Manuscripts" task force, we have simplified and shortened the list of specific criteria; yet, in many ways, the task of reviewing has actually increased in complexity. For example, specific research methods in health professions education have expanded in scope, some methodologies have changed, new methodologies have emerged, and the variety of topics covered in the field has increased since the original publication of these criteria in 2001.

The demands for publication have also increased with the burgeoning of health professions faculty in medical and allied health schools, residency programs, and hospitals. The sheer number of faculty has grown, but so has the diversity of skills and expertise. Further, new venues (e.g., on-line-only journals) and new types of reports (e.g., of innovations, of quality-improvement initiatives) have made new types of publications possible. Despite changes and improvements in research methods, and in the face of the diversification of the avenues for publication, the reviewer still faces the basic task of assessing the quality, relevance, and value of manuscripts.

The explicit criteria for review of manuscripts in this guide convey the same message to authors, reviewers, and publishers; these are the standards, the backbone, upon which a submission will be judged. All parties are informed of the expectations, and all parties should anticipate that any review will be based on these shared criteria. Of course, the extent to which any manuscript reaches or exceeds these accepted standards may be debated among these parties within and across journals, and the final decision to publish a report likely rests on other factors beyond these criteria (including suitability for the journal or timeliness of the work). Still, meeting the criteria herein remains essential, the basis on which any decision should depend.

What will be the status of reviewing in 2030? We suspect that a third edition will be needed by then (if not earlier). That edition will likely have to address challenges such as peer review following publication (e.g., crowd sourcing), global relevance (journals with wider audiences, publication in multiple languages), and work that increasingly builds on a conceptual and theoretical base. The third edition, like this one, will have to cover more numerous (and more refined) methods, and it will have to be relevant to new members entering our community with more diverse skill sets and areas of expertise. Importantly, we believe, the need for such criteria and for peer review will be just as relevant in 15 years as it is today and was in 2001.

As with the first edition, we hope the community will use this one to help train and support new peer reviewers. New reviewers are continuously needed not only to assess the increasing number of submissions to scholarly journals, but also to help evaluate the innovative techniques and new issues that scholars are exploring. As we developed this new edition of the "Review Criteria for Research Manuscripts," we explicitly sought to pair authors of the 2001 edition with faculty who are emerging as experts. If this work succeeds in providing a set of commonly understood criteria that can be shared with all reviewers, including those who are just beginning their careers, it will have been a successful effort.

As Georges Bordage stated in the conclusion to the first edition of this guide,[1] these criteria have not been "cast in stone"; rather, they are living, they change as our community also changes. This second edition of the criteria, revised for 2015, is not identical to the first, and there will be modifications to the criteria in the future as the state of scholarship and publication evolves. As research evolves, so will the criteria by which it is judged.

*Funding/Support:* None reported.
*Other disclosures:* None reported.
*Disclaimer:* The views expressed herein are the authors' own and do not necessarily represent those of the Uniformed Services University of the Health Sciences, the U.S. Department of Defense, or the federal government of the United States.

## Reference

1. Bordage G. Conclusion. Acad Med. 2001;76:956.


**Association of**
**American Medical Colleges**
655 K Street, N.W., Suite 100, Washington, D.C. 20001-2399
**T** 202 828 0400
www.aamc.org