# Table of Contents

## Forewords

- **Reaching the Front of a Reader’s Mind Through the Journal’s Back Page**  
  Steven L. Kanter, MD  
  1

- **Conducting Research in Health Professions Education: From Study Idea to Peer-Reviewed Publication**  
  Anthony R. Artino Jr, PhD  
  2

## General

- **Guidelines for Multi-Institutional/Collaborative Research**  
  Society of Directors of Research in Medical Education  
  4

- **OneHealth: A Paradigm for Interdisciplinary Collaboration**  
  Sam Shomaker, MD, JD  
  5

- **A Comparison of Three Health Care Education Collaboration and Publication Portals**  
  Libby Gordon Cohen, EdD, and Youmna Ashraf Sherif  
  6

- **AM Last Page: Generalizability in Medical Education Research**  
  Anthony R. Artino, Jr., PhD, Steven J. Durning, MD, and John R. Boulet, PhD  
  7

- **AM Last Page: Reliability and Validity in Educational Measurement**  
  Anthony R. Artino, Jr., PhD, Steven J. Durning, MD, and Alisha H. Creel, PhD  
  8

- **Threats to Internal and External Validity in Health Professions Education Research**  
  Dario M. Torre, MD, MPH, PhD, and Katherine Picho, PhD  
  9

- **Responsible Conduct of Research Education (What, Why, and Does It Work?)**  
  Michael Kalichman, PhD  
  10

- **Knowledge Translation and Implementation Science in Health Professions Education: Time for Clarity?**  
  Aliki Thomas, PhD, OT, and André Bussières, PhD, DC  
  11

- **Towards a Greater Understanding of Implementation Science in Health Professions Education**  
  Aliki Thomas, PhD, OT, and André Bussières, PhD, DC  
  12
Before

Conceptual Frameworks to Guide Research and Development (R&D) in Health Professions Education
Georges Bordage, MD, MSc, PhD, Matthew Lineberry, PhD, and Rachel Yudkowsky, MD, MHPE

Generating Good Research Questions in Health Professions Education
C. Jessica Dine, MD, MSHP, Judy A. Shea, PhD, and Jennifer R. Kogan, MD

Generating Research Questions Appropriate for Qualitative Studies in Health Professions Education
Bridget C. O’Brien, PhD, Victoria J. Ruddick, and John Q. Young, MD, MPP

Planning the Literature Review
Justin L. Sewell, MD, MPH, Lauren A. Maggio, PhD, and Anthony R. Artino Jr, PhD

AM Last Page: How to Perform an Effective Database Search
Lauren A. Maggio, MS(LIS), MA, Nancy H. Tannery, MLS, and Steven L. Kanter, MD

AM Last Page: A Guide to Research Paradigms Relevant to Medical Education
Esther Bergman, MSc, Jeantine de Feijter, MD, Janneke Frambach, MA, MSc, Merijn Godefrooij, MD, Irene Slootweg, Renée Stalmeijer, PhD, and Jonne van der Zwet, MD

AM Last Page: Understanding Qualitative and Quantitative Research Paradigms in Academic Medicine
Laura Castillo-Page, PhD, Sue Bodilly, PhD, and Sarah A. Bunton, PhD

AM Last Page: Quality Criteria in Qualitative and Quantitative Research
Janneke M. Frambach, MA, MSc, Cees P.M. van der Vleuten, PhD, and Steven J. Durning, MD, PhD

AM Last Page: Avoiding Five Common Pitfalls of Experimental Research in Medical Education
Mariëtte H. van Loon, MSc, Ellen M. Kok, MSc, Rachelle J.A. Kamp, MSc, Katerina Bohle Carbonell, MSc, Jorrick Beckers, MSc, Janneke M. Frambach, MA, and Anique B.H. de Bruin, PhD

AM Last Page: Common Evaluation Designs in Medical Education I
Dario M. Torre, MD, MPH, PhD, and Barbara Daley, PhD

Common Evaluation Designs in Medical Education II
Dario M. Torre, MD, MPH, PhD, Allison Ferris, MD, Barbara Daley, PhD, and Steven J. Durning, MD, PhD

Common Evaluation Designs in Medical Education III
Dario M. Torre, MD, MPH, PhD, Allison Ferris, MD, Barbara Daley, PhD, and Steven J. Durning, MD, PhD

Using a Logic Model to Assist in Program Planning, Implementation, and Evaluation of Educational Programs
Elaine Van Melle, PhD
P Value: Pitfalls and Solutions
Timothy J. Wood, PhD

How to Calculate a Survey Response Rate: Best Practices
Andrew W. Phillips, MD, MEd, Benjamin T. Friedman, MD, NREMT-P, and Steven J. Durning, MD, PhD

After

Presenting Data From Health Professions Education Research in Tables and Figures
Jennifer Campi and Elizabeth S. Karlin, MA

It’s a Story, Not a Study: Writing an Effective Research Paper
Lorelei Lingard, PhD, and Chris Watling, MD, PhD

The Ethics and Etiquette of Research Dissemination
Holly Meyer, PhD, Lara Varpio, PhD, Larry Gruppen, PhD, and Gurjit Sandhu, PhD

Ten Tips to Move From “Revisions Needed” to Resubmission
Holly S. Meyer, PhD, Jan Carlile, PhD, and Steven J. Durning, MD, PhD

Congratulations! Your Article Has Been Accepted. Now What? Media, Social Media, and Other Outlets for Promoting Your Work
Toni Gallo, MA

A Medical Educator’s Guide to #MedEd
Andrew Micieli, MMI, Jason R. Frank, MD, and Alireza Jalali, MD
A community of scholars should be in continuous conversation about the concepts and methods, theories and assumptions, and approaches and new ideas that form a field of inquiry. They should refresh, renew, and refine this ongoing dialogue by continually revisiting what is known and how they have come to know it.

One of my goals as editor-in-chief of Academic Medicine from 2008 to 2012 was to explore new ways that the journal could catalyze this conversation. This objective gave birth to the notion of using the journal’s last page to promote a broader understanding of issues of importance for medical schools and teaching hospitals, to illustrate basic principles and processes in medical education, and to make the journal’s content more accessible to more people by explaining—with graphs, diagrams, and data on a single page—fundamental concepts and ideas in academic medicine.

When the journal began publishing this feature in November 2008, the editorial staff and I decided to call it “AM Last Page,” until such time as someone could think of a better name. During my tenure, we never did, and now the name seems to have caught on.

Although the entire editorial team contributed valuable ideas as we fleshed out the notion of AM Last Page, Liza Karlin deserves special mention as the editor who agreed to make this feature her special project. Liza’s sharp eye and keen mind ensured that Last Pages were accurate, relevant, and useful. In 2011, Dr. Steven Durning joined Academic Medicine in the role of assistant editor, AM Last Page, and worked with Liza to develop the feature into the remarkable resource it is today. And now, Dr. Anthony Artino has edited this collection on conducting research in health professions education.

In one way, AM Last Page is a kind of “spaced learning” about topics of importance to those who work and learn at medical schools and teaching hospitals. It offers a dose of learning “q month” that, over time, refocuses the reader’s attention on a range of important topics—from fundamental issues like generalizability in medical education research, to newer concepts, like OneHealth, which strives to disrupt intellectual barriers between human and animal health. These and many others—all focused on conducting research in health professions education—are reprinted in this eBook to provide readers with a robust set of AM Last Pages in one place.

The original prescription for AM Last Page was to take one each month and call the editor if you were not learning. With this eBook, the dose can be increased. You can add breadth to your knowledge base of research methods every day.

I hope you enjoy this excellent collection of AM Last Pages.

Funding/support: None reported.
Other disclosures: None reported.
Ethical approval: Reported as not applicable.
In November 2008, *Academic Medicine* published its first AM Last Page (LP). As described in the journal’s Instructions for Authors, the one-page LP “is designed to make the journal’s content more accessible to more people by promoting a general understanding of important issues that affect medical schools and teaching hospitals. This feature tells a story, visually and succinctly, through images or figures and complementary text.” Since 2008, *Academic Medicine* has published more than 90 LPs that are publicly accessible on the journal’s Web site; they have become a favorite among readers, especially our international audience. Over the last three years, about a third of LPs have attained an Altmetric score of at least 10 (the average score for articles in *Academic Medicine* is 8.4). The Altmetric score, a potential alternative to other indicators of scholarly content. Moreover, we agree to daily, online conversations about Altmetric data, although limited, suggest Twitter, Facebook, and Mendeley. These measures of how an article is shared in the journal's Web site; they have become a favorite among readers, especially our international audience. Over the last three years, about a third of LPs have attained an Altmetric score of at least 10 (the average score for articles in *Academic Medicine* is 8.4). The Altmetric score, a potential alternative to other indicators of scholarly content. Moreover, we agree with Twitter, Facebook, and Mendeley. These metrics of how an article is shared in scholarly communication, including social media and other outlets, such as Twitter, Facebook, and Mendeley. These Altmetric data, although limited, suggest that the journal’s LPs are contributing to daily, online conversations about scholarly content. Moreover, we agree with prior editor-in-chief Steven L. Kanter, MD, who dreamt up the LP idea and noted that the feature has gone a long way toward helping the journal advance the story of medical schools and teaching hospitals, the story of educating the next generation of physicians and scientists, the story of biomedical research, and the story of the resources required to do these things in the most ethical, efficient, and effective ways possible.”

Many months ago, the staff and editors of *Academic Medicine* undertook a project to compile a series of LPs, some previously published and some new, centered on the topic of research in health professions education and written by content experts from around the globe. The purpose of this collection is to provide readers, particularly junior faculty, with a set of resources that introduce the principles of educational scholarship. We imagine that this collection might be used in introductory courses and faculty development workshops. Clearly, however, these LPs are only the beginning of the story; they are not meant to be thorough treatments of their respective topics.

The fruits of our efforts over the past year are presented here in a series of 44 LPs designed to take readers from the beginning of a research project, the study idea, to the final product, a published manuscript in a high-quality, peer-reviewed journal. All of the LPs in this collection were reviewed by at least two internal reviewers (i.e., members of the *Academic Medicine* editorial team), and most were reviewed by one or more external reviewers as well. Although deceptively simple, all of these LPs have gone through multiple iterations and have been edited by the *Academic Medicine* staff. Furthermore, I suspect that most of the LP authors would agree with the sentiment that it is far from easy to condense such complex ideas into a single, readable page.

In just a few decades, the field of health professions education has grown tremendously. Our most respected journals are no longer filled with descriptions of local curriculum improvements or simple correlational studies with no conceptual framework. Today, the field has matured considerably, and so, too, have the tools of the research trade. To be taken seriously by our colleagues both inside and outside the health professions, we must ensure that our research continues to be conducted thoughtfully, responsibly, and with the utmost rigor. Although not a comprehensive solution to these priorities, we sincerely hope this collection of LPs will be a step in the right direction.

The LPs included here are organized around the before, during, and after stages of an educational research study. As anyone who has completed at least one study can attest, much of the work of conducting an empirical investigation—particularly the research team’s most difficult, cognitive work—occurs long before the first participant is ever enrolled. As such, it is no accident that the before section of this collection contains the bulk of the LPs. For it is the study’s design and planning stage where researchers often spend the most time thinking about their research question(s), considering conceptual frameworks, and obtaining ethical approval, to name just a few of the important before activities. Once this foundation has been built, the remaining activities frequently run more smoothly. This is not too say that the during and after stages are easy. They, too, take considerable time and effort, particularly the final acts of writing and rewriting the manuscript. I believe, however, that if the before stage is done well, the remaining two stages have a tendency to fall into place.

Although several scholars have lamented the problems associated with health professions education research, we have seen in our work as editors and reviewers that, over the past decade, the quality of research in the field has improved markedly. We hope the LPs presented in this collection help to further facilitate this positive trend by serving as accessible, high-quality resources for young scholars in the health professions.

**Funding/Support:** None reported.

**Other disclosures:** None reported.

**Disclaimer:** The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Uniformed Services University of the Health Sciences, 4301 Jones Bridge Rd., Bethesda, MD 20814; telephone: (301) 295-3693; e-mail: anthony.artino@usuhs.edu.
Foreword

References

Guidelines for Multi-Institutional/Collaborative Research
Society of Directors of Research in Medical Education

Many research funding agencies are looking for multi-institutional proposals, and given advances in technology, conducting collaborative research has become less cumbersome. Further, generalizable results have more rigor and are the first step in the chain of outcomes research.\textsuperscript{1,2} Additionally, institutions can maximize their resources by working together. Thus, in order for results of medical education studies to be generalizable, schools need to collaborate and conduct research across institutions. There has been one medical education paper on collaboration advice,\textsuperscript{3} but a one-page tip sheet does not exist. To that end, the Society of Directors of Research in Medical Education have developed guidelines for conducting multi-institutional/collaborative research. An example from the American Medical Association’s (AMA’s) 27-school Learning Environment Collaborative is provided to illustrate each suggestion.

### Multi-Institutional/Collaborative Research Guidelines

<table>
<thead>
<tr>
<th>Planning Stage</th>
<th>Implementation Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Achieve local buy-in, interest, and ownership</strong></td>
<td><strong>Example From the AMA Learning Environment Collaborative</strong></td>
</tr>
<tr>
<td>Understand the culture of all institutions</td>
<td>Representatives of each school (n = 27) negotiated permission for multiple data collections and sharing of anonymized data with the collaboration.</td>
</tr>
<tr>
<td>- Consult institutional officials about the need for a data sharing agreement</td>
<td></td>
</tr>
<tr>
<td>- Discuss personal goals and possible conflicts</td>
<td></td>
</tr>
<tr>
<td>Define roles needed to advance work</td>
<td>The differences among school cultures, although necessary to this study, required accommodation and adaptation in the protocol.</td>
</tr>
<tr>
<td>- Agree on and document leadership and authorship roles</td>
<td></td>
</tr>
<tr>
<td>- Consider opportunities to mentor more junior colleagues</td>
<td></td>
</tr>
<tr>
<td>Set clear expectations of collaboration</td>
<td>AMA staff provided logistic and administrative leadership. Scholarly leadership in study design and writing came from individual school representatives.</td>
</tr>
<tr>
<td>Set clear goals for the study</td>
<td>Several initial meetings were needed to define the study protocol and the key outcomes.</td>
</tr>
<tr>
<td>Obtain institutional review board (IRB) approval</td>
<td>The goal of this prospective cohort study was to explore the relationship between individual-level and institutional-level factors and the medical education learning environment.</td>
</tr>
<tr>
<td>- Understand differences among IRBs across institutions and countries</td>
<td></td>
</tr>
<tr>
<td>Gather and distribute resources</td>
<td>The Data Management Center obtained overall IRB review for the collaboration, but each school obtained IRB review for its own data collection. Several meetings were required to understand the differences in institutional cultures and to negotiate agreement on a shared study protocol.</td>
</tr>
<tr>
<td>Conduct research</td>
<td>The AMA provided centralized resources for group meetings and a Data Management Center, which served all the schools.</td>
</tr>
<tr>
<td>Schedule regular communication</td>
<td>The agreed-upon protocol was implemented at each school by local personnel. Data were aggregated and managed by the Data Management Center.</td>
</tr>
<tr>
<td>- Set an agenda</td>
<td></td>
</tr>
<tr>
<td>- Use viable technology</td>
<td></td>
</tr>
<tr>
<td>- Take minutes</td>
<td></td>
</tr>
<tr>
<td>Maintain momentum</td>
<td>The AMA scheduled periodic conference calls and face-to-face meetings.</td>
</tr>
<tr>
<td>- Develop a timeline with deadlines</td>
<td></td>
</tr>
<tr>
<td>- Hold one another accountable</td>
<td></td>
</tr>
<tr>
<td>- Celebrate small victories</td>
<td></td>
</tr>
<tr>
<td><strong>Note:</strong> The Society of Directors of Research in Medical Education is a nonprofit organization (501c3) dedicated to enhancing the quality of education in medical schools. It is composed of directors of units with responsibility for educational research, evaluation, and program development.</td>
<td></td>
</tr>
</tbody>
</table>

### References

**Author contact:** jone.colbert-yet@hsc.utah.edu
**OneHealth: A Paradigm for Interdisciplinary Collaboration**

Sam Shomaker, MD, JD, immediate past dean of medicine and immediate past vice president for clinical affairs, Texas A&M Health Science Center College of Medicine

**Definition and History of OneHealth**

OneHealth, “the collaborative effort of multiple disciplines—working locally, nationally, and globally—to attain optimal health for people, animals, and our environment,” originated as a collaboration between human and veterinary medicine but has expanded to include many disciplines.

**Broad Reach of OneHealth**

**Government organizations involved in OneHealth:**
- Centers for Disease Control and Prevention
- U.S. Department of Agriculture
- World Bank
- World Health Organization

**Members of the OneHealth Commission:**
- American Medical Association
- American Veterinary Medical Association
- American Public Health Association
- Association of American Medical Colleges
- Association of American Veterinary Medical Colleges
- Infectious Diseases Society of North America

**Disciplines involved in OneHealth:**
- Human medicine
- Veterinary medicine
- Agricultural science
- Public health
- Environmental science
- Bioengineering
- Climatology
- Architecture
- Wildlife biology
- Economics

**Benefits of OneHealth**

- Bringing the expertise of multidisciplinary partnerships to the solution of complex problems (for an example, see supplemental digital content at http://links.lww.com/ACADMED/A231).
- Developing new scientific relationships and teams that would not have otherwise come together
- Exposing students from different disciplines to interprofessional collaboration and teamwork
- Creating centers of excellence to work on the most important global health problems

**OneHealth and the Tripartite Academic Medicine Mission**

| Medical education: | 
| Trainees need to be aware that diseases can be transmitted from animals to humans and vice versa. The health of a family’s animals can and does impact human health. | Trainees must appreciate that zoonotic diseases remain a major health problem around the world; they can be rapidly transmitted to populations anywhere in the world. | Climate change will have major health impacts which trainees must be ready to address wherever they practice. | 

| Biomedical research: | 
| Many diseases manifest in human and animal species in similar ways, making animal models an effective way to study human diseases. | Diseases regularly jump from animals to humans; predicting which diseases might do so (and where) could head off future pandemics. | Compounds found in wild plants and animals can be used to treat human disease. | 

**One World**

**OneHealth**

**AHCs and OneHealth: Next Steps**

The beauty of OneHealth is that universities do not need a particular set of programs or faculty to make significant contributions.

**Best practices:**
- Commitment from senior leadership is a necessary catalyst.
- Designation of one or more leaders for the initiative makes it someone’s job to move the project forward.
- Identification of potential collaborators is an important process; no particular set of programs, disciplines, or faculty are imperative, but those who do participate need to be willing to work together.
- Institutional policies that encourage and reward collaboration are important.
- Agreement on how funds will flow and credit will be apportioned are best worked out in advance.
- Having designed research space to allow interprofessional interaction is ideal.

**Challenges:**
- Resources are necessary to pay for project leadership and staff support.
- OneHealth is such a big, all-encompassing idea that some may struggle to operationalize it in tangible projects.
- Interprofessional educational collaborations face logistical and scheduling constraints.
- Researchers working on projects that cut across multiple disciplines may have trouble identifying funding agencies to support their work.

**References:**

**Author contact:** shomaker@me.com

First published online July 29, 2014
A Comparison of Three Health Care Education Collaboration and Publication Portals
Libby Gordon Cohen, EdD, and Younna Ashraf Sherif, Duke—National University of Singapore Graduate Medical School

Health care education collaboration and publication portals are digital gateways for accessing, sharing, and consuming health education materials, resources, tools, and research. These portals create constantly evolving online communities that allow users to interact and collaborate with others who share similar interests. They are enhanced by rapid content updates and core services such as newsletters, announcements, personal accounts, user profiles, social media, and efficient search capacity. After surveying available health care education collaboration and publication portals, three—MedEdPORTAL, MedEdWorld, and MERLOT (compared below)—were identified as dynamic portals that promote scholarship and the dissemination of resources in health care education.

<table>
<thead>
<tr>
<th></th>
<th>MedEdPORTAL</th>
<th>MedEdWorld</th>
<th>Multimedia Educational Resource for Learning and Online Teaching (MERLOT)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Major sponsor</strong></td>
<td>Association of American Medical Colleges (AAMC)</td>
<td>Association for Medical Education in Europe (AMEE)</td>
<td>California State University System</td>
</tr>
<tr>
<td><strong>Purpose</strong></td>
<td>To foster innovation in health care education, research, and delivery</td>
<td>To promote interprofessional collaboration and scholarship</td>
<td>To create an international online community</td>
</tr>
<tr>
<td><strong>Scope</strong></td>
<td>Health professions education including prehealth and undergraduate, graduate, and continuing education</td>
<td>Health professions education</td>
<td>Health sciences education including allied health</td>
</tr>
<tr>
<td><strong>Access</strong></td>
<td>Free registration allows submitting and downloading materials</td>
<td>Paid membership allows browsing, downloading, and submitting materials</td>
<td>Free registration allows browsing and downloading articles</td>
</tr>
<tr>
<td><strong>Submissions</strong></td>
<td>Templates provided for simulations, standardized patients, assessments, and team-based learning modules</td>
<td>Any medical education resources including PowerPoint slides, videos, Web sites, animated diagrams, mini modules, etc.</td>
<td>Previously unpublished articles</td>
</tr>
<tr>
<td><strong>Screened to ensure compliance with copyright laws and patient privacy standards</strong></td>
<td>Screened to ensure compliance with copyright laws and patient privacy standards</td>
<td>Screened for relevance to MedEdWorld’s mission, usefulness to health professions educators, and fulfillment of minimum writing standards</td>
<td>Screened for relevance to MedEdWorld’s mission and fulfillment of minimum writing standards</td>
</tr>
<tr>
<td><strong>Reviewer</strong></td>
<td>AAMC staff conducts the screening</td>
<td>AAMC staff conducts the initial screening</td>
<td>MedEdWorld administrators conduct the screening</td>
</tr>
<tr>
<td><strong>Content updates</strong></td>
<td>Authors can review their materials for relevancy and accuracy after three years. Minor changes can be made. If major changes are required, the authors can resubmit their materials.</td>
<td>Authors of time-sensitive materials can reexamine their submissions after three years to determine if any updates or revisions are needed</td>
<td>Resources remain posted indefinitely unless identified as out of date</td>
</tr>
</tbody>
</table>

Author contact: lcohent71@gmail.com
AM Last Page: Generalizability in Medical Education Research

Anthony R. Artino, Jr., PhD, assistant professor of preventive medicine and biometrics, Steven J. Durning, MD, professor of medicine, Uniformed Services University of the Health Sciences, and John R. Boulet, PhD, associate vice president for research and data resources, Foundation for Advancement of International Medical Education and Research

**Generalizability**, also referred to as external validity, is the extent to which the conclusions of a study would hold for variations in persons, settings, treatments, and outcomes.¹

**A depiction of the concept of generalizability.** The conclusions of a particular study are represented by the red box. These conclusions likely generalize to persons, settings, treatments, and outcomes that are similar to those found in the study. However, as one considers other persons, settings, treatments, and outcomes that are more dissimilar to those in the study (i.e., as one moves out along the blue arrows), it is less likely that the study conclusions will generalize.²

The following restrict the generalizability of a study’s conclusions:

- A sample that is homogenous or somehow unusual
- An atypical setting
- A peculiar time in history
- An atypical treatment (e.g., an intervention applied in an abnormally small classroom)
- Inappropriate or unimportant outcomes (e.g., using only medical knowledge to assess clinical competence)

The following improve the generalizability of a study’s conclusions:

- Using random sampling
- Selecting multiple study sites (e.g., multi-institutional studies)
- Analyzing the similarities and differences between the study context and other contexts
- Replicating the study with different people, in a variety of settings, and with variations in the independent and dependent variables

When considering generalizability, researchers should ask, “To what populations, settings, treatment variables, and measurement variables can this effect be generalized?”³

**The tension between generalizability and internal validity.** Internal validity is the extent to which an observed association between variable A and B reflects a causal relationship from variable A to variable B.¹ Generalizability and internal validity are often in conflict within a given study (as illustrated below). That is, design features that improve internal validity, such as controlling certain independent variables, may also reduce the generalizability of the study’s findings. Thus, a study with high internal validity (e.g., a tightly controlled laboratory experiment) often has low generalizability, and vice versa.

**Generalizability**

A qualitative study conducted in multiple hospitals may have high generalizability, but it likely has low internal validity.

**Generalizability and internal validity are often in conflict within a particular study.**

**Internal Validity**

An experimental study conducted in a single, tightly controlled laboratory may have high internal validity, but it likely has low generalizability.

References

AM Last Page: Reliability and Validity in Educational Measurement

Reliability is the extent to which the scores produced by a particular measurement tool or procedure are consistent and reproducible. Reliability answers the question, “Does the assessment yield the same scores at different times, from different raters, or from different items?”

Validity is the degree to which an assessment measures what investigators want to measure, all of what they want to measure, and nothing but what they want to measure. Validity answers the question, “Does the assessment provide information that is relevant to the inferences that are being made from it?” An assessment, such as a test or questionnaire, does not have validity in any absolute sense. Instead, the scores produced are valid for some uses and not valid for others.

A target provides a metaphor for the relationship between reliability and validity. The true score (or value) for the concept the researcher is attempting to measure is at the center of the target, and the observed score the investigator gets from each person assessed is a shot at the target.

Neither reliable nor valid

Reliable, but not valid

Both reliable and valid

Reliability is a necessary but insufficient condition for validity. To be valid, scores must first be at least moderately reliable. However, scores that are reliable may be devoid of validity for the application the researcher has in mind.

Many methods of assessing reliability and validity are available. Each method provides the researcher with slightly different information about the reliability and validity of the assessment.

### Assessing reliability

- **Test-retest**
- **Equivalent forms**
- **Interrater**
- **Internal consistency**
  - Split-half
  - Kuder-Richardson
  - Cronbach alpha

### Assessing validity

- **Construct**
  - Convergent
  - Discriminant
  - Known-groups

- **Criterion-related**
  - Predictive
  - Concurrent
  - Postdictive

- **Content-related**
  - Content
  - Face

### Method definition

- **Assessing reliability**
  - Assesses agreement between the same assessment given on two separate occasions.
  - Assesses agreement between similar forms of an assessment given at separate times.
  - Assesses agreement between two or more coders or raters.
  - Assesses correlation between the different items on an assessment.

- **Assessing validity**
  - Determines whether the assessment captures the concept. Example: Factor analysis reveals that the survey items “hang together” as expected, relate to a single factor, and are unrelated to another, different set of items.
  - Compares the assessment to a criterion that the researcher thinks is important. Example: MCAT scores should be related to medical school performance.
  - Assesses whether the items measure all the important aspects of the construct. Example: Items on an exam should assess all of the learning objectives.

### References

Threats to Internal and External Validity in Health Professions Education Research

Dario M. Torre, MD, MPH, PhD, associate professor of medicine, and Katherine Picho, PhD, assistant professor, Department of Medicine, Uniformed Services University of the Health Sciences

- Internal validity refers to the degree to which inferences can be made about the causal relationship between two variables.
- External validity pertains to whether study outcomes can be generalized across different persons, treatments, outcomes, and settings.
- Internal and external threats to validity can occur during any stage of the research process.
- Below, we describe some of the most common threats to internal and external validity.

Example: Consider a pre/posttest control group research design to evaluate an Internet-based, multimedia cardiac auscultation teaching program (intervention) compared with a traditional lecture-based program (control). The intervention is aimed at enhancing cardiac auscultation skills of medical students.

<table>
<thead>
<tr>
<th>Threat</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal Validity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>History</td>
<td>Events occurring concurrently with treatment may cause the observed effect.</td>
<td>Students participate in a series of grand rounds on cardiac auscultation while taking the multimedia cardiac auscultation program.</td>
</tr>
<tr>
<td>Maturation</td>
<td>Naturally occurring changes over time may be confused with a treatment effect.</td>
<td>Students’ psychological development and confidence in the use of technology naturally grow during the time of the intervention and positively affect the cardiac auscultation skills of the intervention group.</td>
</tr>
<tr>
<td>Regression to the mean</td>
<td>When units are selected for their extreme scores, subsequent scores are likely to be less extreme.</td>
<td>Students with extremely high (or low) scores on a previous cardiac auscultation test are selected for the study. Their subsequent scores after the intervention are likely to be much lower (or much higher for low scores) than on the previous test; thus, scores regress toward the mean.</td>
</tr>
<tr>
<td>Attrition</td>
<td>Loss of respondents to treatment or measurement can produce artifactual effects if loss is systematically correlated with conditions.</td>
<td>Fifty percent of students assigned to the Internet-based multimedia program could not access the Web site and did not complete the program.</td>
</tr>
<tr>
<td>Testing</td>
<td>Exposure to a test can affect scores on subsequent exposures to that test, and this change in test scores can be confused for a treatment effect.</td>
<td>By taking the pretest, students are sensitized to a number of murmurs that are later included in the posttest. Their performance in the posttest may be affected by the content of the pretest.</td>
</tr>
<tr>
<td>Instrumentation</td>
<td>The nature of a measure may change over time/conditions in ways that may be confused for treatment effects.</td>
<td>A change in the pretest is made during the experiment by including more difficult and complex murmurs and by decreasing the time given to complete the test. The instrument change, rather than the intervention, may affect students’ performance.</td>
</tr>
<tr>
<td><strong>External Validity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple treatment interference</td>
<td>The results of one kind of treatment may not hold when that treatment is combined with other treatments.</td>
<td>Students are exposed to real patients with cardiac murmurs in a clinical rotation, and subsequently to a multimedia cardiac auscultation program. The carryover effects between the two interventions may be difficult to separate.</td>
</tr>
<tr>
<td>Reactive effects of experimental arrangement</td>
<td>A result that occurs in one kind of setting may not hold in other settings.</td>
<td>The cardiac auscultation proficiency of medical students in a simulated setting may not translate into the same level of proficiency in a real patient setting.</td>
</tr>
<tr>
<td>Interaction of selection bias and experimental treatment</td>
<td>The effect of selection bias interacting with the experimental treatment has an effect on the outcome of the intervention.</td>
<td>The multimedia program is given to students who have poor cardiac auscultation skills as opposed to students who have excellent skills (e.g., a group of residents in training vs. a group of experienced cardiologists). The results of the intervention may not be generalizable; they may be biased, according to the different level of skills of the group selected to receive the intervention.</td>
</tr>
</tbody>
</table>

Disclaimer: The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Uniformed Services University of the Health Sciences, the Department of the Navy, the Department of Defense, or the U.S. Government.

Additional resources:

Author contact: dario.torre@usuhs.edu
Responsible Conduct of Research Education (What, Why, and Does It Work?)

Michael Kalichman, PhD, director, Research Ethics Program, University of California, San Diego

Responsible conduct of research (RCR) education is often required and described as important.1,2 One of the major reasons is concern about research misconduct, but arguably even more important is simply the desire to foster good research practices. In either case, providing RCR education is not only good for science but consistent with an obligation to the society served by science.

**What is RCR?** RCR is often considered synonymous with the list of topics recommended by the National Institutes of Health (NIH) for researchers supported by NIH training or career development awards, although the scope of that requirement has varied over time.1

<table>
<thead>
<tr>
<th>Current NIH guidelines list nine RCR topics:</th>
<th>Depending on research domains and experience, additional topics of as much or more importance to RCR might include:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Conflict of interest</td>
<td>Conflicts of conscience</td>
</tr>
<tr>
<td>2. Human and animal subjects</td>
<td>Sabotage</td>
</tr>
<tr>
<td>3. Mentoring</td>
<td>Use of statistics</td>
</tr>
<tr>
<td>4. Collaboration</td>
<td>Image manipulation</td>
</tr>
<tr>
<td>5. Peer review</td>
<td>Reproducibility</td>
</tr>
<tr>
<td>6. Data management</td>
<td>Censorship</td>
</tr>
<tr>
<td>7. Research misconduct</td>
<td>Scientists as activists</td>
</tr>
<tr>
<td>8. Authorship and publication</td>
<td>Deception</td>
</tr>
<tr>
<td>9. Scientists and society</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difficult conversations</td>
</tr>
<tr>
<td></td>
<td>Big data</td>
</tr>
<tr>
<td></td>
<td>Embryonic stem cells</td>
</tr>
<tr>
<td></td>
<td>Dual use technology</td>
</tr>
<tr>
<td></td>
<td>Weapons research</td>
</tr>
<tr>
<td></td>
<td>Managing a research group</td>
</tr>
<tr>
<td></td>
<td>Managing budgets</td>
</tr>
</tbody>
</table>

**Why teach RCR?** Without explicit goals for teaching RCR, questions such as “How should educators teach RCR?”—much less “Is RCR teaching effective?”—are impossible to answer. Unfortunately, no easy answers as to what the goals are, or even what they should be, exist. A nominal answer is that RCR should be taught because it is required, at least for some people by federal agencies. However, that begs the question, “Why should RCR education be required?”

**Examples of the range of possible outcomes6 for teaching RCR include**:  
- Decrease research misconduct  
- Decrease RCR disputes or misunderstandings  
- Increase particular areas of RCR knowledge  
- Develop or improve skills that will promote RCR  
- Foster positive attitudes about RCR and continued RCR learning  
- Promote a culture of RCR

**Does teaching RCR work? It depends on the desired outcome8:**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Effectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decreased research misconduct</td>
<td>No evidence,1 but access to data is limited by what information becomes public. Even if more data were available, a significant impact is unlikely since the classroom appears to be less important than the research environment.3</td>
</tr>
<tr>
<td>Decreased RCR disputes or misunderstandings</td>
<td>No evidence, but it is plausible that increased awareness of issues will diminish the risk of disputes and misunderstandings.</td>
</tr>
<tr>
<td>Increased knowledge and/or skills</td>
<td>Evidence is nominal. Although statistically significant improvements have been reported (e.g., for ethical decisionmaking), the magnitude of the impact of teaching RCR on knowledge and skills is typically modest, absent, or negative.1,4</td>
</tr>
</tbody>
</table>
| Positive attitudes, continued learning, and culture of RCR | **Individual impact:** Evidence is promising, but must allow for different individuals experiencing different impacts (e.g., improved ethical decision-making skill or increased awareness of authorship standards).5  
**Group impact:** Some evidence suggests successful fostering of a culture of integrity based on the extent to which trainees continue conversations outside the classroom (Kalichman and colleagues, unpublished observations). |

---

References:  

Author contact: mkalichman@ucsd.edu
Knowledge Translation and Implementation Science in Health Professions Education: Time for Clarity?
Aliki Thomas, PhD, OT, School of Physical and Occupational Therapy, Centre for Medical Education, McGill University, Centre for Interdisciplinary Research in Rehabilitation of Greater Montreal; and André Bussières, PhD, DC, School of Physical and Occupational Therapy, McGill University, Centre for Interdisciplinary Research in Rehabilitation of Greater Montreal, Department of Chiropractic, Université du Québec à Trois-Rivières

- A central objective of health professions education (HPE) is to provide learners with state-of-the-art education that will prepare them for their future practice as health care professionals.
- There is a growing evidence base in HPE regarding individual learners (teaching and assessment strategies), organizations (curricular review, program evaluation), and policy (accreditation, licensure).1-3
- High-stakes decisions such as selection, assessment, and licensure demand a critical and evidence-informed approach to HPE.
- Knowledge translation aims to promote the uptake and application of research evidence to improve educational practices and ultimately patient care.
- Efforts to inform and improve HPE should be underpinned by the science (theories, models, methods) of knowledge translation, also known as implementation science.

### End Goal: Knowledge Utilization to Inform and Improve HPE

<table>
<thead>
<tr>
<th>TYPE OF UTILIZATION</th>
<th>DEFINITION</th>
<th>EXAMPLES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research utilization</td>
<td>Specific kind of knowledge utilization whereby the knowledge has a research base to substantiate it. It is a complex process through which knowledge, in the form of research, is transformed from the findings of one or more studies into instrumental, conceptual, or persuasive/symbolic utilization.</td>
<td>Findings from a BEME systematic review that increase the use/application of an assessment method</td>
</tr>
<tr>
<td>Instrumental utilization</td>
<td>A concrete application of research from one or more studies, which is normally translated into a material and usable form, such as a protocol or set of guidelines</td>
<td>Findings from research that influence someone's attitudes and beliefs toward or intention to use a particular assessment</td>
</tr>
<tr>
<td>Conceptual utilization</td>
<td>Research findings from one or more studies that may change someone's thinking but not necessarily his/her observable actions</td>
<td>Findings from research that influence accreditation standards</td>
</tr>
<tr>
<td>Persuasive/symbolic utilization</td>
<td>The use of research findings from one or more studies as a persuasive (or political) tool to legitimize a position or practice</td>
<td></td>
</tr>
</tbody>
</table>

Numerous terms have been used interchangeably to refer to the dissemination and uptake of research findings that inform and change educational practices (e.g., diffusion, knowledge translation, dissemination, translational research, implementation). Below, we differentiate and clarify these terms to facilitate discussion within HPE and to help highlight the relationship between knowledge utilization and educational practices and policy.

### Meeting the End Goal of Knowledge Utilization: Three Distinct but Interrelated Processes

- **Diffusion** *(Let it happen)*
  - A passive process by which new evidence is communicated to researchers, educators, and educational policy makers using traditional vehicles
  - Examples: Conference presentations, Peer-reviewed publications, Non-peer-reviewed publications, Social media

- **Dissemination and Knowledge Translation** *(Make it happen)*
  - A process by targeted and tailored data and information (main messages or key implications) are transmitted to specific relevant audiences to increase the application and uptake of evidence as well as to bridge research–practice gaps
  - Examples: End-of-grant reports to funders, Faculty development activities, Summaries and briefs to stakeholders, Creation of knowledge tools such as guidelines and BEME systematic reviews

- **Implementation Science** *(Use robust methods)*
  - A process that uses robust scientific methods underpinned by theories, models, and frameworks to (1) identify research–practice gaps, (2) identify supports and barriers to the uptake of educational evidence, (3) design interventions to reduce research–practice gaps, and (4) evaluate the impact of the interventions on educational practices
  - Examples: Theory of planned behavior and knowledge-to-action process frameworks used both to assess supports and barriers associated with teachers’ use of effective feedback strategies and to design tailored, theory-driven interventions to promote uptake of effective feedback strategies

### Key Messages
- A growing emphasis is placed on evidence-informed approaches to HPE.
- Diffusion (raising awareness), dissemination (use of knowledge), and implementation science (theories and methods that underpin knowledge translation) are distinct but interrelated concepts.
- To promote the uptake and application of evidence in HPE, researchers and educators need to identify research–practice gaps, implement knowledge translation strategies, and evaluate impacts iteratively.

### References

Author contact: ali.thomas@mcgill.ca

BEME indicates the Best Evidence Medical Education Collaboration

First published online September 20, 2016
Towards a Greater Understanding of Implementation Science in Health Professions Education

Aliki Thomas, PhD, OT, School of Physical and Occupational Therapy, Centre for Medical Education, McGill University. Centre for Interdisciplinary Research in Rehabilitation of Greater Montreal, and André Bussières, PhD, DC, School of Physical and Occupational Therapy, McGill University. Centre for Interdisciplinary Research in Rehabilitation of Greater Montreal, Department of Chiropractic, Université du Québec à Trois-Rivières

In a previous AM Last Page, we advocated an evidence-informed approach to health professions education (HPE). Here we examine implementation science (IS).

- Educators are faced with the responsibility of ensuring that current best evidence in HPE is routinely used to inform decision-making processes.
- Knowledge translation (KT) is a process used to facilitate the uptake and application of best evidence.
- IS is the scientific study of KT; it encompasses all aspects of research relevant to the study of the methods, theories, and models to promote the uptake of research findings into educational and policy contexts. IS seeks to answer questions such as:
  - Why are some teachers more likely than others to adopt a new practice?
  - Why do certain faculty development programs lose effectiveness over time?
  - How can multiple educational interventions be effectively packaged to capture cost efficiencies and reduce suboptimal practices?

We present IS as a four-step process: (1) identify research-practice gaps; (2) identify facilitators and barriers to the uptake of new knowledge/feedback; (3) design interventions to promote uptake; and (4) implement and evaluate impact. For each step, we describe the purpose, methods, and expected deliverables/outcomes. The implementation process should consider the context (e.g., school, clinic, community, emergency department, surgery) and identify the target audience and stakeholders (e.g., learners, faculty, program directors, administrators) early and involve them throughout all stages of the process.

Four-Step Implementation Process

<table>
<thead>
<tr>
<th>Step</th>
<th>Purpose</th>
<th>Methods</th>
<th>Outcome</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Identifying research-practice gaps</td>
<td>Describe current practice</td>
<td>Knowledge syntheses, Portfolio, Surveys, Guided interviews, Focus groups, Curriculum and accreditation document reviews</td>
<td>List of important gaps, List of current teaching, assessment, and program development activities</td>
<td>Review evidence on strategies for giving residents effective feedback, Identify current feedback practices in residency training programs using questionnaires and focus groups, Confirm presence and nature of the gap between current feedback practices and best practice strategies</td>
</tr>
<tr>
<td>2. Identifying facilitators and barriers</td>
<td>Identify level of the facilitator/barrier: Individual knowledge, attitudes, motivation, skills, etc. Organizational availability of resources, culture, readiness to change, etc. System: health care reforms, regulations and laws, etc. Identify theoretical framework to explain reasons for the gaps: Theoretical Domains Framework (TDF), Consolidated Framework for Implementation Research, etc.</td>
<td>Use theories to identify and understand facilitators and barriers: Motivational, social-cognitive, action theories, etc. Data sources: Qualitative interviews, focus groups, Quantitative surveys, Mixed approaches</td>
<td>List of facilitators and barriers with explanatory components Data to inform the design of targeted strategies to improve educational practices and policies</td>
<td>Interviews among clinical teachers underpinning the TDF to identify the individual and organizational supports (e.g., readiness to change; residency training program with resources to support uptake of new practices, protected time to read and discuss evidence on feedback and barriers (e.g., lack of knowledge on effective feedback strategies; heavy patient caseloads) to effective feedback practices</td>
</tr>
<tr>
<td>3. Designing interventions</td>
<td>Design interventions that are: Theory-based and aligned with facilitators and barriers Target to appropriate audience Contextualized to local learning environment Feasible, acceptable, sustainable Developed and implemented in partnership with relevant stakeholders Select intervention components: Map practice change techniques to facilitators and barriers (modeling, self-monitoring, guided task, skill rehearsal, etc.) Use evidence supporting the effect of the intervention: Individual, feedback outreach visits, faculty development Organizational System Operationalize the intervention (targeted to whom, why, when, where, what, how often, and by whom) Select mode of delivery (must be feasible, acceptable, guided by local context)</td>
<td>Theory-based tailored intervention ready for implementation</td>
<td>Consider who needs to do what differently, why, when, and how? Involve teachers, department chairs, and residents in designing the KT interventions to promote uptake of new feedback strategies For example, intervention (feedback) mapped to previously identified barrier (a specific knowledge gap) delivered online biweekly over four months by (supervisory clinician) to a new group of residents</td>
<td></td>
</tr>
<tr>
<td>4. Implementing and evaluating impact</td>
<td>HPE researchers, implementation scientists, and other stakeholders evaluate intervention outcomes at three levels: Individual: learners, teachers, etc. Organizational: school, hospital ward, etc. System: education, health, etc. Pre-post studies Quasi-experimental Controlled trials Case studies Cohort studies Mixed methods</td>
<td>Individual outcomes Organizational outcomes System outcomes</td>
<td>Measurable changes in: Knowledge, attitudes, skills, and behaviors regarding effective feedback strategies in residency training programs Cost-effective and streamlined residency programs, improved learner outcomes, etc. Accreditation, reputation, quality of care, safety, etc.</td>
<td></td>
</tr>
</tbody>
</table>

Key messages:
- KT and IS are iterative processes targeted at specific populations, settings, and contexts to promote the systematic uptake of research findings and other evidence-based practices into HPE.
- KT and IS can foster environments conducive to building teaching and assessment capacity and students’ lifelong learning.
- Added value of medical education must be proven via robust scientific methods employed in IS.

References:

Author contact: ali Thomas@mcgill.ca
Conceptual Frameworks to Guide Research and Development (R&D) in Health Professions Education

Georges Bordage, MD, MSc, PhD, professor of medical education, Matthew Lineberry, PhD, assistant professor of medical education, and Rachel Yudkowsky, MD, MHPE, associate professor of medical education, Department of Medical Education, College of Medicine, University of Illinois at Chicago

Conceptual frameworks (CFs) are ways of...
- Thinking about a problem or question e.g., Thomas et al.'s six steps to curriculum development.²
- Representing how complex things work e.g., Dual-Process Cognition Theory.³

Each CF is inherently limited, focusing on specific operational elements while leaving others out.¹

To find CFs...
- Critically review the literature for similar initiatives.
- Note the CFs used.
- Be open-minded to the many frameworks from which to choose.
- Select the one(s) that best fits your needs.

When reporting educational research and development projects, state the CFs clearly so that others know your assumptions.

Why CFs?
- CFs are pervasive; they underlie, explicitly or not, all our educational choices and actions.
- CFs offer a variety of perspectives from which to look at educational problems or research questions.
- CFs provide a solid foundation, with standardized vocabulary and well-grounded principles, on which to build educational R&D projects and interpret outcomes and results.
- CFs allow researchers to build on one another’s work, leading to an ever greater understanding that moves the field forward.¹

<table>
<thead>
<tr>
<th>Dimensions of a project or study</th>
<th>Content</th>
<th>Variables and their interrelatedness</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Basic elements</td>
<td></td>
<td>Educational</td>
</tr>
<tr>
<td>Key questions addressed</td>
<td>&quot;What are the important elements to consider for this topic or issue?&quot;</td>
<td>&quot;How are the variables related?&quot;; &quot;What's our model or theory?&quot;</td>
<td>&quot;How might I design instruction or assessment for this project?&quot;</td>
</tr>
</tbody>
</table>

Example study Stefaniadis et al.²
- Problem: Learners are making limited gains from simulation-based surgical skills training and they struggle to transfer that learning into practice under stress and distractions in the operating room.

Authors’ CFs
- Fundamentals of Laparoscopic Surgery: Five Basic Skills
- Dual-Process Cognition Theory²

How each CF influenced the authors’ study from the beginning
- Suggested a skill to focus on, laparoscopic suturing, which is standardized and familiar internationally.
- Clarified what the authors did not choose to study (e.g., precision cutting or ligating loop).

Highlighted that whether learners have learned something to the point of automaticity (unconscious, effortless actions) is not evident solely by their strong performance of a task, but also by their having spare cognitive resources to multitask.

Suggested that learners should practice the skill until they reach a deliberately chosen performance standard, rather than that all learners simply practice for a fixed amount of time.

Suggested a research design for criterion measurement: Observe learners’ performance on the main task under two conditions: (1) without distraction vs. (2) while also performing a secondary task that requires similar cognitive processes.

Major insight gained from the use of CFs
- Interpretation: To help learners reach automaticity for a task (e.g., suturing) to a particular standard, clinical educators should require that they continue practicing the task until they can perform it well while substantially distracted.

References:

Author contact: bordage@uic.edu

First published online September 20, 2016
### Generating Good Research Questions in Health Professions Education

C. Jessica Dine, MD, MSHP, associate program director, Internal Medicine Residency Program, Judy A. Shea, PhD, associate dean, Medical Education Research, and Jennifer R. Kogan, MD, assistant dean, Faculty Development, Perelman School of Medicine at the University of Pennsylvania

- Generating a specific research question is an integral part of the overall research design.
- It lays the foundation for the research study and informs each step of the study design.

<table>
<thead>
<tr>
<th>Steps</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Identify an idea or problem</strong>&lt;br&gt;A general research question lays the foundation for the entire study design.</td>
<td>Use local problems or ideas to formulate a general question&lt;br&gt;- Problems, interests, or changes at your institution&lt;br&gt;- Ideas from reading medical education journals&lt;br&gt;- Lapses during handoffs have been occurring at your institution. You are wondering how to improve handoff practices.</td>
</tr>
<tr>
<td><strong>Perform a literature review</strong>&lt;br&gt;A literature review helps identify what has already been published and a conceptual framework.</td>
<td>Review prior publications&lt;br&gt;- Identifies prior methodology, gaps in understanding, and areas for elaboration&lt;br&gt;- Identify a conceptual framework&lt;sup&gt;1&lt;/sup&gt;&lt;br&gt;- Organizes related ideas into an overarching theme&lt;br&gt;- Informs your research including the selection of study variables and the interpretation of results&lt;br&gt;- Few studies have looked at whether simulated handoffs improve the quality of handoffs (literature review). You begin to read more about simulation-based mastery learning&lt;sup&gt;1&lt;/sup&gt; (conceptual framework).</td>
</tr>
<tr>
<td><strong>Generate a specific research question</strong>&lt;br&gt;The general research question is narrowed to state the specific goal of the study.</td>
<td>Narrow your general research question to a more specific question&lt;br&gt;- FINER question&lt;sup&gt;2&lt;/sup&gt;&lt;br&gt;- Needs to be answerable&lt;br&gt;- Your first question: Does handoff simulation reduce unnecessary test ordering? You are unable to determine what is necessary ordering, and refine your question to whether handoff simulation decreases handoff errors on call.</td>
</tr>
<tr>
<td><strong>Develop a study design</strong>&lt;br&gt;The specific research question and conceptual framework identify study variables and inform the study design.</td>
<td>Consider common medical education study designs&lt;sup&gt;3,4&lt;/sup&gt;&lt;br&gt;- Experimental&lt;br&gt;- Quasi-experimental&lt;br&gt;- Nonexperimental&lt;br&gt;- Qualitative&lt;sup&gt;5&lt;/sup&gt;&lt;br&gt;- You plan a quasi-experimental design. Interns on half of the services will participate in the handoff simulation. You plan to use a validated tool identified in the literature to assess quality of handoffs in the two groups.</td>
</tr>
</tbody>
</table>

The “FINER” criterion is an example of available frameworks that can be used to “test” the specific research question.<sup>3</sup>

<table>
<thead>
<tr>
<th>FINER research question</th>
<th>Feasible</th>
<th>Interesting and Important</th>
<th>Novel</th>
<th>Ethical</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is this question answerable with the resources you have available to you?</td>
<td></td>
<td>Is this question interesting to you as the investigator, as well as to the general health professions education community?</td>
<td></td>
<td>Can you answer this question without putting anyone at risk?</td>
<td>Does the answer to the question matter not only at your institution but also at others?</td>
</tr>
</tbody>
</table>

References:

Author contact: jessica.dine@uphs.upenn.edu
Generating Research Questions Appropriate for Qualitative Studies in Health Professions Education

Bridget C. O’Brien, PhD, University of California, San Francisco School of Medicine; Victoria J. Ruddick, University of California, San Francisco School of Medicine; and John Q. Young, MD, MPP, Hofstra Northwell School of Medicine

In qualitative studies, the process of crafting and refining research questions is similar to the process for other types of studies (quantitative, mixed methods, and reviews); the differences are in the nature of the questions, the underlying assumptions, and the treatment of the questions as evolving entities. Qualitative researchers often begin with broad questions, then hone their questions or identify new ones through an iterative process of collecting and analyzing data, reviewing relevant literature, and revising the purpose statement. The components of qualitative research design come together in a reflective process through which the components occur “more or less simultaneously, each influencing all the others.”1(p2) As shown below, the research question is the core of the research process.

- The research components that interact with the research question appear in the corners.
- To help readers formulate appropriate questions, we provide examples of initial and revised questions.

The Research Process: Using an Interactive Model to Develop Research Questions

PURPOSE AND GOALS
Identify a problem, dilemma, or phenomenon that...
- sparks your curiosity
- is not well explained in the literature
- is researchable, feasible, significant, and relevant

METHODS
Include key elements such as:
- approaches with epistemological and ontological assumptions
- data collection and analytic techniques
- participants and setting
- presentation of findings
- ethical considerations

My research question is qualitatively oriented if the study...
- aims to explore, understand, describe, discover, or generate
- focuses on process. How and why do things happen?
- uses open-ended or semi-structured approaches and techniques

A “good” qualitative research question...
- identifies the central phenomenon to be studied. Usually seeks to understand a process, the meaning of activities and events from the participant’s perspective, or the physical or social context of activities and events.
- explores rather than assumes. Typically does not impose a conceptual framework upfront; does not assume that participants’ experiences and perspectives match those of the researcher.
- is clear, concise, focused. Frames the study, clarifies scope and context, strikes a balance between general topic and specific focus.
- is significant and relevant. Has implications for fundamental concepts in education or important educational decisions.

CONCEPTUAL FRAMEWORK
Apply theories, beliefs, and prior research findings that...
- are based on the literature, preliminary studies, or personal experiences
- explain your thinking about the problem or phenomenon
- guide study design and analysis

TRUSTWORTHINESS
Demonstrate rigor in approach and methods, such as:
- reflexivity
- credibility

Initial Research Questions
Do more communication errors occur when sign-out is given by a junior trainee to a senior trainee, or vice versa?

Feedback on Research Questions
This question may be better suited to a quantitative approach. A qualitative question would focus on process rather than outcomes.

What are the characteristics of an effective handoff?

Refined Research Questions
This question is too broad. Focus on a particular setting, perspective (giver or receiver), and type of handoff. Make sure the study addresses a gap in the literature.

How do faculty respond to errors in clinical reasoning?

References:

Author contact: Bridget.ORobin@ucsf.edu Twitter: @obrien_15
Planning the Literature Review

Justin L. Sewell, MD, MPH, assistant professor of medicine, University of California, San Francisco; Lauren A. Maggio, PhD, associate professor of medicine, Uniformed Services University of the Health Sciences; and Anthony R. Artino, Jr, PhD, professor of medicine, Uniformed Services University of the Health Sciences

A literature review provides a synthetic summary and synthesis of what is known and unknown regarding a scholarly topic. A rigorous literature review may encourage the use of appropriate methods, enhances generalizability, places research within a broader context, accelerates progress in health professions education, and improves the likelihood of publication. 2–4. This Last Page highlights the steps for planning and executing a literature review within health professions education.

**Define scope**

- Identify accepted standards for articles, and review model examples of the type you plan to write that have been published in target journals.
- Map out general topics of interest.
- Assess your existing knowledge and expertise. Novices may need to spend significant time searching and reading the literature.

**Design search**

- Select databases. Consider medical (PubMed, CINAHL), educational (ERIC), psychology (PsycINFO), and general databases (Scopus, Google Scholar).
- Review database indexing terms for a few relevant articles.
- Examine published reviews for potential search terms and strategies.
- Consult a medical librarian or another colleague for assistance.

**Locate articles**

- Dedicate time for iterative rounds of searching.
- Consult AM Last Page: How to perform an effective database search. 5
- Review reference lists of key articles for additional relevant references.

**Organize & assess results**

- Select a reference manager (e.g., Zotero, Endnote, Refworks, Mendeley), considering issues of access and cost.
- Develop a plan to systematically review titles, abstracts, and full text.
- Check search results for presence of key articles. If absent, revise your search.

**Refine & repeat**

- Carefully select references to cite and ensure reference accuracy, as citation problems contribute to rejection. 1,4
- Stop searching when the same citations appear repeatedly and new relevant citations are not seen.
- Repeat your search at critical time points, such as just before manuscript preparation, submission, and revision.

Disclaimer: The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Uniformed Services University of the Health Sciences, the Department of the Navy, the Department of Defense, or the U.S. Government.

References:

Author contact: justin.sewell@usfcl.edu

First published online July 5, 2016
AM Last Page: How to Perform an Effective Database Search

Lauren A. Maggio, MS(LIS), MA, medical education librarian, Stanford University School of Medicine, Nancy H. Tannery, MLS, associate director for User Services, University of Pittsburgh Health Sciences Library System, and Steven L. Kanter, MD, vice dean, University of Pittsburgh School of Medicine

1. Choose a database

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage</td>
<td>Biomedical literature, journal articles</td>
<td>Diverse disciplines, journal articles, book chapters, dissertations, abstracts</td>
<td>Education literature, journal articles, book chapters, Association of American Medical Colleges reports</td>
<td>Nursing and allied health literature, journal articles, book chapters, dissertations, audiovisuals</td>
<td>Scientific, technical, medical and social sciences literature, citation searching, journal articles, conference papers</td>
</tr>
<tr>
<td>Controlled vocabulary (See 2A)</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Access</td>
<td>Open</td>
<td>Open</td>
<td>Open</td>
<td>Subscription</td>
<td>Subscription</td>
</tr>
</tbody>
</table>

2. Select search terms

A. If available, use the database’s controlled vocabulary:
- A controlled vocabulary provides one term for a concept that may have different names.
- Controlled vocabularies help create exhaustive and unambiguous searches.
- MEDLINE’s controlled vocabulary is Medical Subject Headings (MeSH). For example, *myocardial infarction* is the MeSH for *heart attack*. Using the term *myocardial infarction* retrieves articles on both heart attack and myocardial infarction in MEDLINE.

B. Include synonyms and use truncation:
- Include synonyms and abbreviations to broaden your search and to help ensure comprehensiveness for ideas not yet identified in a standard thesaurus.
- Use truncation to search for alternate endings of search terms.

3. Use Boolean operators to combine search terms

- **reform** OR **revision** AND **gme** OR **reform**
  - “OR” broadens the search
  - “AND” narrows the search

4. Limit results

- Use limits to narrow the search.
- Apply limits one at a time to control search results.
- Popular limits include English language and date ranges.

5. Explain the search process in the methodology section of any report. Include the following:

- Search terms (indicate if controlled vocabulary was used)
- Boolean operators
- Databases searched
- Any limits applied
- Date of search

For additional information, consult your medical librarian.
AM Last Page: A Guide to Research Paradigms Relevant to Medical Education

Esther Bergman, MSc, PhD student, Jeanette de Feijter, MD, PhD student, Jannike Frambach, MA, MSc, PhD student, Merijn Godefrooij, MD, PhD student, Irene Slootweg, PhD student, Renée Stalmeneier, PhD, assistant professor, Jonne van der Zwa, MD, PhD student, Maastricht University

In order to design or interpret qualitative and quantitative research, one should have some understanding of the assumptions that underpin them. Below, we provide an overview of some of the concepts underlying four philosophical paradigms in medical education research and illustrate the relationships between them.

**Paradigm: A philosophical framework that underlies and affects research activities. What are the assumptions underlying one’s views on reality and knowledge? (Synonyms: theoretical or epistemological stance, world view)**

- **Positivism**: There is one truth, and it can be observed.
- **Post-Positivism**: There is one truth, but it can never be truly observed.
- **Critical Theory**: Multiple truths exist, and they are influenced by power relations among people.
- **Constructivism**: Multiple truths are constructed by and between people.

**Ontology: Theory of the view on reality. What is the nature of physical and social reality?**

- **Realism**: Reality is objectively observable and exists independently of the human knower. The world is operated by laws of cause and effect. Variables can be observed, measured, and predicted.
- **Critical Realism**: Reality is assumed to exist, but evidence in research is fallible due to the complexity of the enquiry.
- **Historical Realism**: Reality is shaped by structures of social, political, cultural, economic, ethnic, and gender factors.
- **Relativism**: Reality is socially and experientially based; multiple realities exist, change, conflict, and/or become more crystallized.

**Epistemology: Theory of knowledge. What are the origin, nature, and limits of knowledge about reality?**

- **Radical Objectivism**: Knowledge is independent of the human knower. People can provide an objective, value-free description of reality.
- **Relative Objectivism**: Knowledge is conjectural and based on hypotheses that have not yet been falsified. Objective knowledge about reality is the ideal, which cannot be achieved.
- **Relative Subjectivism**: Knowledge is value-dependent, is influenced by power relations, and is the result of interaction between researcher and participants.
- **Radical Subjectivism**: Knowledge consists of constructions that arise from interaction between researcher and participants.

**Methodology: Strategic approach to answer the research question and to gain knowledge. What is the research design?**

- **Verification**: Knowledge is gained through hypothesis generation and testing (induction). It focuses on prediction and control of phenomena. The aim is to produce generalizable data.
- **Falsification**: Knowledge is gained by testing if hypotheses can be disproved, using a deductive approach. Outcomes are never totally objective.
- **Transformation**: Knowledge is gained by raising participants to a different level of consciousness and thereby empowering them.
- **Interaction**: Knowledge is gained by an inductive approach: recognizing, understanding, developing, and contrasting constructions through dialogue.

As illustrated below, an understanding of research paradigms can guide researchers in designing and performing medical education research. Each step invites the researcher to consider underlying assumptions about knowledge and reality within the field of medical education and related disciplines.

Suggestions for further reading:
# Understanding Qualitative and Quantitative Research Paradigms in Academic Medicine

Laura Castillo-Page, PhD, senior director, Diversity Policy and Programs, Sue Bodilly, PhD, senior director, Research and Data Programs, and Sarah A. Bunton, PhD, research director, Organization and Management Studies, Association of American Medical Colleges

Qualitative research is becoming more prominent in academic medicine and health care fields, and an increasing number of publications using qualitative methods are featured in prominent journals. Thus, recognizing the different available approaches can benefit researchers of all types. While a debate may wage between proponents of qualitative versus quantitative research, both sets of methods—and often a blend of the two—offer important insights into the problems the academic medicine community faces.

<table>
<thead>
<tr>
<th>Qualitative paradigm</th>
<th>Quantitative paradigm</th>
</tr>
</thead>
<tbody>
<tr>
<td>How and why events or behaviors occur in complex settings where context is important to understanding:</td>
<td>Nature of the research question</td>
</tr>
<tr>
<td>Examples: How do a diverse student body and faculty affect teaching and learning? How does a resident make the transition to attending physician? What characterizes the phenomenon of a mentor-mentee relationship?</td>
<td>How many, how often, what level, and what direction of relationships between defined variables in settings that can be decontextualized:</td>
</tr>
<tr>
<td>Inductive by researchers (e.g., normative or transcribed text analyzed thematically for patterns)</td>
<td>Examples: What is the relationship between student grades and graduation rates? What type and amount of monetary incentive or financial reward affects a medical student’s specialty choice?</td>
</tr>
<tr>
<td>• Case study: An in-depth study of a particular case, which can be descriptive, explanatory, or exploratory</td>
<td>Types of designs</td>
</tr>
<tr>
<td>• Ethnography: Research intended to provide descriptions of systems, processes, or phenomena within their specific context; stems from anthropology</td>
<td>• Experimental: The researcher manipulates all variables including the assignment to treatment and control groups in order to discern causality</td>
</tr>
<tr>
<td>• Grounded theory: A theory developed based on the examination of data (rather than applying a predetermined theory)</td>
<td>• Quasi-experimental: Research using an experimental variable with groups not formed through random assignment or selection</td>
</tr>
<tr>
<td>• Historiography: Research directed at the study of a past event, issue, or problem that uses information from the past</td>
<td>• Surveys: Measurement procedures that involve asking questions of respondents</td>
</tr>
<tr>
<td>• Phenomenology: The study of individuals’ perspectives on particular phenomena</td>
<td>• Mixed methods: A combination of quantitative and qualitative approaches including triangulation design, embedded design, explanatory design, and exploratory design</td>
</tr>
<tr>
<td>• Action research: A reflective and team-based approach led by those involved in solving a particular problem</td>
<td></td>
</tr>
<tr>
<td>• Mixed methods: A combination of quantitative and qualitative approaches including triangulation design, embedded design, explanatory design, and exploratory design</td>
<td></td>
</tr>
<tr>
<td>Normative data from interviews, documents, focus groups, and/or observations</td>
<td>Data sources</td>
</tr>
<tr>
<td>• Thematic analysis</td>
<td>Ordinal or cardinal data from surveys, financial reporting, census reports, test scores, demographics, and/or observations</td>
</tr>
<tr>
<td>• Content analysis</td>
<td>Analytic techniques</td>
</tr>
<tr>
<td>• Analysis of frequency</td>
<td>• Descriptive statistics</td>
</tr>
<tr>
<td></td>
<td>• Regression</td>
</tr>
<tr>
<td></td>
<td>• Regression discontinuity</td>
</tr>
<tr>
<td></td>
<td>• Hierarchical linear modeling</td>
</tr>
<tr>
<td>• Internal validity (e.g., through triangulation, member checking, coding check)</td>
<td>Assessment of rigor</td>
</tr>
<tr>
<td>• External validity (e.g., through representativeness check)</td>
<td>• Internal validity (e.g., through study design and procedures)</td>
</tr>
<tr>
<td>• Reliability (e.g., through chain of evidence and interrater reliability check)</td>
<td>• External validity (e.g., through criterion measurement)</td>
</tr>
<tr>
<td>• Provides valid and dense information about real situations and contexts, including interactions of variables in context</td>
<td>• Reliability (e.g., through test–retest, internal consistency)</td>
</tr>
<tr>
<td>• Allows an in-depth and comprehensive understanding of motives and social or behavioral processes</td>
<td>Strengths</td>
</tr>
<tr>
<td>• Provides an understanding and description of people’s personal experiences of phenomena</td>
<td>• Delineates relationships among variables</td>
</tr>
<tr>
<td>• May produce findings that are not easily generalizable to other settings</td>
<td>• Provides generalizable research findings when the data are based on sufficiently sized random samples</td>
</tr>
<tr>
<td>• May be of limited scope due to the in-depth data-gathering approaches used</td>
<td>• Provides generalizable results when research has been replicated in different populations/subpopulations</td>
</tr>
<tr>
<td>• May take more time to collect and analyze data</td>
<td>• Is useful for large populations</td>
</tr>
<tr>
<td>• May be more difficult to test theories with large participant pools</td>
<td>Weaknesses</td>
</tr>
<tr>
<td></td>
<td>• Narrow variables might not be valid</td>
</tr>
<tr>
<td></td>
<td>• Knowledge produced might be too general for direct application to specific contexts or individuals</td>
</tr>
<tr>
<td></td>
<td>• Phenomena may be missed if analysis focuses on hypothesis testing rather than hypothesis generation</td>
</tr>
</tbody>
</table>

References

AM Last Page: Quality Criteria in Qualitative and Quantitative Research

Janneke M. Frambach, MA, MSc, PhD student, Cees P.M. van der Vleuten, PhD, professor of education, Maastricht University, Steven J. Durning, MD, PhD, professor of medicine and pathology, Uniformed Services University of the Health Sciences

Good research in medical education is characterized by evidence that is trustworthy, applicable to (multiple) practical settings, consistent, and neutral (unbiased)—regardless of whether a qualitative or a quantitative approach is used. However, while qualitative and quantitative research share similar standards for good evidence (quality criteria), the conception and operationalization of these quality criteria differ between the two. Below, we provide an overview of these criteria and a number of techniques that researchers can use to meet them. In addition, we note that the criteria are interlinked, and that some of the techniques contribute to multiple criteria at the same time.

Techniques to enhance quality in quantitative research

- Calculate the sample size that is needed for sufficient statistical power (power calculation)
- Describe details of the educational context and intervention
- Avoid loss of participants or provide information on non-responses
- Standardize treatment conditions
- Use control groups (controlled design)

Quality criteria in quantitative research

- Internal validity: The extent to which observed effects can be attributed to the independent variable
- External validity: The extent to which the results can be generalized from the research sample to the population

Quality principles

- Truth value of evidence
- Credibility: The extent to which the study’s findings are trustworthy and believable to others

Quality criteria in qualitative research

- Applicability of evidence
- Transferability: The extent to which the findings can be transferred or applied in different settings

- Reliability: The extent to which the results are consistent if the study would be replicated
- Consistency of evidence
- Dependability: The extent to which the findings are consistent in relation to the contexts in which they were generated

- Objectivity: The extent to which personal biases are removed and value free information is gathered
- Neutrality of evidence
- Confirmability: The extent to which the findings are based on the study’s participants and settings instead of researchers’ biases

Techniques to enhance quality in qualitative research

- Use multiple data sources (data triangulation), methods (methodological triangulation), researchers (investigator triangulation) and theories (theory triangulation)
- Collect data for an extended period of time (prolonged engagement)
- Ask feedback from participants on the data or interpretation of the data (member checking)
- Make the findings meaningful to others by describing them and their context in detail (thick description)
- Explain the sampling strategy (e.g., typical case sampling or maximum-variation sampling)
- Discuss the findings’ resonance with existing literature from different settings

- Collect data until no new themes emerge (saturation)
- Continuously analyze the data to inform further data collection (iterative data collection)
- Continuously re-examine the data using insights that emerge during analysis (iterative data analysis)
- Be flexible and open towards the process and topic (flexible emergent research design)

- Search the data and/or literature for evidence that disconfirms the findings
- Discuss the research process and/or findings with peers/experts (peer debriefing)
- Keep a diary to reflect on the process and the researcher’s role and influence (reflectivity)
- Document the steps and decisions taken in the research, and their motives (audit trail)

Suggestions for further reading:


Author contact: jframbach@maastrichtuniversity.nl

Disclaimer: The ideas expressed in this Last Page are solely those of the authors and do not represent an endorsement by or the views of the Uniformed Services University of the Health Sciences, the Department of Defense, or the U.S. government.

Dr. Durning is an associate editor for Academic Medicine, but he was not involved in the review of, or decision to publish, this Last Page.
# Avoiding Five Common Pitfalls of Experimental Research in Medical Education

Mariëtte van Loon, MSc, PhD student; Ellen M. Kok, MSc, PhD student; Rachelle J.A. Kamp, MSc, assistant professor; Katerina Bohle Carbonell, MSc, PhD student; Jonrick Beckers, MSc, PhD student; Janneke M. Frambah, MA, PhD student; and Anique B.H. de Bruin, PhD, associate professor, Maastricht University

Experimental research is a scientific method that aims to provide evidence for cause-and-effect relations. One or more independent variables are systematically manipulated to determine the effect(s) on a dependent variable while controlling other relevant factors. Often, the goal is to gain insight into underlying factors of an educational intervention. However, pitfalls are numerous in medical education experiments. Below, we present five common pitfalls and ways to avoid them.

<table>
<thead>
<tr>
<th>Pitfall</th>
<th>Explanation of the problems and examples</th>
<th>Recommended solutions</th>
<th>Good example from the literature</th>
</tr>
</thead>
</table>
| Using an inappropriate control condition | When you compare an experimental condition with a control condition, then you can attribute differences in outcome to differences between the conditions. If these conditions vary on too many elements, it is impossible to attribute outcomes to a specific element.  
- If you compare Web-based learning with lectures, which differ in many aspects (e.g., learning pace, interaction with peers and teachers), you won’t know which aspect(s) of these learning modes caused differences in outcomes. | • Identify the crucial element of your intervention.  
• Make the experimental and control conditions as similar as possible, except for the crucial element. | Issa et al. compared a lecture that was designed according to multimedia principles with a lecture that was not designed according to these principles, but similar in every other aspect. |
| Failing to align your outcome measures to your research questions | Outcome measurements should reflect the dependent variable(s) stated in your research question(s). If your outcome measures do not match your theory, your results do not answer your research questions.  
- If you expect that students learn communication skills better when they have contact with real instead of simulated patients, you should measure communication skills rather than knowledge or perceptions about communication. | • When designing a study, first clarify expected effects.  
• Next, define how you can observe these effects.  
• Then decide which instruments measure these effects. | Cook et al. operationalized their dependent variable (learning outcomes) with two test types: a post-test after each module and a cumulative test. |
| Ignoring possible reactive effects of a pretest | A pretest could provide information on baseline differences between participants. However, a pretest can cause participants to acquire relevant information. Therefore, the pretest can reinforce your intervention or have a direct effect on the dependent variable(s) that you measure with the posttest.  
- If you ignore effects of a pretest that assesses prior knowledge, you won’t know whether your results can be attributed solely to your intervention. | In a nonrandomized design:  
• Let students do an irrelevant task between the pretest and intervention.  
• Use existing data (e.g., grades) as a pretest.  
In a randomized design:  
• Don’t use a pretest. | Hatala et al. randomly allocated students to one of two instructional approaches and didn’t use a pretest to investigate the superiority of one of these approaches. |
| Not taking time-on-task into account | It is likely that increased time spent on learning tasks yields increased learning outcomes. If you do not take this into account, it is impossible to attribute your outcomes solely to the variables you measured because they might be explained by differences in time-on-task as well.  
- If you compare Web-based learning with lectures, the time-on-task is the actual time spent on the study activities. | • Design conditions so that participants spend the same amount of time on the task.  
• Control for time-on-task in statistical analyses if there are differences between conditions. | In Mamede et al. the time participants were allowed to spend on each study case was the same for all conditions. |
| Confusing ecological and external validity | Ecologically valid experiments do not necessarily have high external validity. Ecological validity is the extent to which your study approximates the real world. It often introduces elements (e.g., teacher characteristics, motivation) that mask or change effects, which, in turn, may compromise the external validity or generalizability of your study.  
- If you investigate effects of an individual assignment in a classroom setting, student interaction can influence the effect and thus compromise external validity. | • Focus on external validity instead of ecological validity.  
• Achieve high ecological validity by conducting a well-controlled experiment that is repeated in different settings and conditions. | Marquard et al. investigated patient identification errors, controlling for the number and type of errors identified, during medication administration. |

References:

Author contact: m.vanbon@maastrichtuniversity.nl
# AM Last Page: Common Evaluation Designs in Medical Education I

Dario M. Torre, MD, MPH, PhD, associate professor of medicine and vice chair, Education, Drexel University College of Medicine, and Barbara Daley, PhD, interim associate dean, College of Nursing, and professor of Adult and Continuing Education Program, University of Wisconsin—Milwaukee

Evaluation design is important to establish that an instructional program produced the measured effects and learning outcomes. Evaluation design should help isolate extraneous factors so that differences or outcomes can be safely attributed to the instructional program. The goal of this Last Page is to describe three common program evaluation designs, along with benefits, drawbacks, and examples.

Internal validity of an evaluation measures the certainty with which educators can ascertain whether the program actually caused the effects they find. Educators should always consider threats to validity in designing program evaluation. Some of these internal validity threats (history, maturation, testing) are included in the description of some of the evaluation designs below.

### One-Group Pre-Post Test Design

**Example:**

One group of students rotating through a medicine clerkship is given an 8-station objective structured clinical examination before and after the administration of a month-long, Web-based instructional program about the clinical presentation and diagnosis of ten common medical diseases.

**Pros:**
- Easy to implement
- May be helpful for formative evaluation (e.g., to gather information that will guide program improvement), particularly if the interval of time is short

**Cons:**
- Internal validity threats, including both history (within the time that passes before and after the intervention, events may occur that influence the outcome) and maturation (learners may naturally grow during the time of the experiment), should be considered
- Absence of comparison group makes it difficult to assess whether extraneous variables affected the outcome

### Non-Equivalent Control Group Pre-Post Test Design

**Example:**

Students from hospital X (group E) and students from hospital Y (group C) are selected (not randomly) and given a multiple-choice-question (MCQ) knowledge-based pretest on asthma and chronic obstructive pulmonary disease (COPD). Group E experiences a self-directed reading program while Group C experiences a seminar-based program, involving discussion of asthma and COPD. Both groups are tested again one month later through a MCQ knowledge-based test that has the same content, but different questions, as the pretest.

**Pros:**
- Feasible when randomization is not possible
- Allows for the comparison of two educational interventions
- The use of a pretest allows researchers to assess the comparability of the groups (e.g., are pretest scores the same or different between groups?) at the beginning of the program, since it is important to ensure that the 2 groups are similar at the beginning of the intervention

**Cons:**
- The use of a pretest can lead to testing effect; that is, students may identify certain content topics that will be on the posttest, based on test items in the pretest (e.g., the use of beta blockers in the treatment of congestive heart failure), regardless of question items being different between pre-and posttest
- Selection bias and dissimilar initial groups may be misleading and influence the outcome

### True Control Pre-Post Test Design

**Example:**

A group of students are randomly assigned to either the E group or the C group at the beginning of the academic year (or at the beginning of a rotation). Group E experiences a small-group discussion instructional program and Group C experiences a video-based program with facilitators. Both groups learn about the diagnosis and treatment of osteoarthritis. Both groups are tested through a MCQ, knowledge-based test before and after administration of the programs.

**Pros:**
- Randomization assures group equivalence and eliminates selection bias
- Eliminates many of the internal threats to validity, thus yielding stronger conclusions about the outcome

**Cons:**
- Randomization may be challenging in medical education settings, particularly when classes and rotations are predetermined
- If the pretest is reactive (i.e., the content of the pretest may cause students to focus their study on specific program material), it may influence the outcome of the evaluation

### References:


**Author contact:** Dario Torre@DrexelMed.edu
Common Evaluation Designs in Medical Education II

Dario M. Torre, MD, MPH, PhD, associate professor of medicine and associate director; Evaluation and Long Term Outcomes, Graduate Programs in Health Professions Education, Department of Medicine, The Uniformed Services University of the Health Sciences (USUHS); Allison Ferris, MD, associate program director, Internal Medicine, Drexel University College of Medicine; Barbara Daley, PhD, interim associate dean, College of Nursing, and professor, Adult and Continuing Education Program, University of Wisconsin–Milwaukee; and Steven J. Durning, MD, PhD, professor of medicine and pathology and director, Intro to Clinical Reasoning Course, USUHS

The goal of this Last Page, a follow-up of a previous Last Page,1 is to describe two additional evaluation designs, including their advantages and challenges, and to provide an example for each. We also mention any relevant internal threats to the validity of each design such as the effect of reactive testing.

**True control posttest only**

<table>
<thead>
<tr>
<th>Time</th>
<th>Rando</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>T</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>T</td>
</tr>
</tbody>
</table>

**Example:** A group of students is randomly assigned to either the E or C group at the beginning of a rotation. Group E is given a small-group discussion instructional program (I 1), and group C is given a video-based program with facilitators (I 2). Both groups are tested with an MCQ knowledge-based test after administration of the programs.

**Pros:**
- Randomization can assure group equivalence and eliminates selection bias.
- Lack of pretest eliminates the threat of reactive testing (i.e., the content of the pretest may alert students to focus on specific topics).

**Cons:**
- Randomization may be challenging in medical education since denying a potentially effective educational program to some students may influence learning and final grading (However, the modified true control posttest-only design [see below] can address this issue).
- Without a pretest, researchers cannot be sure that the two groups have the same level of knowledge or skills at the beginning of the program (i.e., pretest scores may be different between groups).

**Legend**
- E = experimental group; C = control group; I = instructional program; T = test, measurement, or observation; Rando = randomization

**Modified true control posttest-only design**

<table>
<thead>
<tr>
<th>Time</th>
<th>Rando</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>T</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>T</td>
</tr>
</tbody>
</table>

**Example:** A group of students is randomly assigned to either the E or C group at the beginning of a rotation. Group E is given a Web-based instructional program (I 1), and group B is given a lecture-based program (I 2). Both groups are tested with an MCQ knowledge-based test after 1 month. After the test, the group (E) that received the Web-based program is given the lecture-based program, and the other group (C) is given the Web-based program for 1 additional month.

The true control posttest-only design can be modified by creating a design wherein after the posttest, the instructional programs are switched between the two groups, so that each receives the program they had not previously received. This design ensures that no students are denied the opportunity to potentially learn from a different program.

**Pros:**
- All students, regardless of which group they are assigned to, receive the experimental program; thus, no student is denied the learning experience.

**Con:**
- The true control posttest-only design is more resource intensive and requires more time to complete.

**Disclaimer:**
The views in this Last Page are those of the authors alone and do not necessarily represent those of the US military, the Department of Defense, or the federal government of the United States.

**Reference:**

**Additional resources:**

**Author contact:** dario.torre.cm@usuhs.edu

MCQ indicates multiple-choice question.
Common Evaluation Designs in Medical Education III

Dario M. Torre, MD, MPH, PhD, associate professor of medicine and vice chair of education, Drexel University College of Medicine; Allison Ferris, MD, associate program director, Internal Medicine, Drexel University College of Medicine; Barbara Daley, PhD, interim associate dean, College of Nursing, and professor, Adult and Continuing Education Program, University of Wisconsin-Milwaukee; and Steven J. Durning, MD, PhD, professor of medicine and pathology and director, Intro to Clinical Reasoning Course, The Uniformed Services University of the Health Sciences

The goal of this Last Page, a follow-up of two previous Last Pages,1,2 is to describe one more evaluation design—known as interrupted time series—and to explain its advantages and challenges. We also provide an example and mention the internal threats to the design’s validity, such as testing, maturation, and reactive testing.

Interrupted Time Series Design (Longitudinal and Successful)

**Longitudinal**

<table>
<thead>
<tr>
<th>E</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
</tr>
</thead>
</table>

**Successful**

<table>
<thead>
<tr>
<th>Ea</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
</tr>
</thead>
</table>

**Examples:**

**Longitudinal:** A group of medical students enrolled in a 14-week rotation is tested with an MCQ knowledge-based test every 2 weeks for 6 weeks on the topic of CHF. At 6 weeks, a Web-based interventional instructional program is administered to the group of students for 2 weeks; after the program is no longer administered, a similar MCQ knowledge-based test is administered every 2 weeks to the same group on the topic of CHF for an additional 6 weeks. Test results are then compared and analyzed for trends pre and post program implementation.

**Successful:** A group of medical students enrolled in a 6-week rotation is tested with an MCQ knowledge-based test every 2 weeks on the topic of CHF. At 6 weeks, a Web-based interventional instructional program is administered to the group of students for 2 weeks; after the program is no longer administered, a different yet equivalent group (randomization may be needed to ensure the two groups are equivalent) begins a new 6-week rotation, and a similar MCQ knowledge-based test administered every 2 weeks to the new group on the topic of CHF for an additional 6 weeks. Test results of the two equivalent groups of students are then compared and analyzed for trends before and after the intervention.

**Legend**

E = experiment; C = control group; I = instructional program; T = test, measurement, or observation

1. Longitudinal: same group measured multiple times at regular intervals
2. Successful: different yet equivalent groups (Ea and Eb) measured multiple times at regular intervals

**Pros:**

- Very helpful to study the longitudinal effects of a program
- Useful if the effect of the intervention is occurring across time
- Helpful when treatment has to be given to all students and cannot be withheld for periods of time
- The successive time series design allows longitudinal evaluation of a program when courses are of short duration and new students start the rotation every 4 to 6 weeks (e.g., for a subinternship) as part of their curriculum

**Cons:**

- Requires time
- Need for multiple equivalent measurements across time
- Reactive testing (i.e., the content of the pretest may alert students to focus on specific topics) is a threat as a result of the repeated testing
- More complex statistical procedures may be needed to correctly interpret results
- History (occurrence of other events during the time of intervention that may affect outcome) and maturation (natural growth of learners independent of the intervention) may be difficult to assess
- Randomization may be needed in the successive time series design to ensure group equivalency

**Disclaimer:**

The views in this Last Page are those of the authors alone and do not necessarily represent those of the US military, the Department of Defense, or the federal government of the United States.


**Additional resources:**


**Author contact:** dario.torre@usuhs.edu

CHF indicates congestive heart failure; MCQ, multiple-choice question.
Using a Logic Model to Assist in the Planning, Implementation, and Evaluation of Educational Programs

Elaine Van Melle, PhD, education scientist, Royal College of Physicians and Surgeons of Canada, and education researcher, Department of Family Medicine, Queen’s University

Logic Model Basics

- A logic model is a diagram that shows how a program is thought to work; that is, how resources (inputs) produce key processes (activities), and how the products (outputs) produce desired results (outcomes).
- Logic models can be drawn in numerous ways; they range from fairly simple to more complex.
- A needs assessment sets the scene for educational program design and determination of expected outcomes.

Example of a Logic Model

A Program to Develop Family Medicine Resident Competency in Comprehensive Health Checks for Individuals With a DD

Tips for Using and Applying Logic Models

- Involve all key stakeholders (e.g., program leaders, developers, implementers, participants) in creating a draft.
- Develop additional iterations to help create consensus amongst the key stakeholders.
- Use the logic model to identify areas where further program planning may be required.
- Use the final logic model to determine the focus for program evaluation.
- Use the logic model to discover whether an established program is producing the desired outcomes.
- Use the logic model to determine whether a newer program is being implemented as intended.

A good logic model facilitates a common approach to program planning, implementation, and evaluation.

Resources


Author contact: e-mail: vanmelle@queensu.ca; Twitter: @elanavannelle
AM Last Page: A Snapshot of Three Common Program Evaluation Approaches for Medical Education

Rebecca D. Blanchard, PhD, assistant professor of education research, Baystate Medical Center, Laura Torbeck, PhD, vice chair of education, Surgery, Indiana University, and Whitney Blondeau, PhD, medical education specialist, Maine Medical Center

Program evaluation, or programmatic assessment, is the application of defensible criteria to determine the worth or merit of a program, project, or curriculum. ’ ’Decision-oriented,’’ ’’outcomes-oriented,’’ and ’’expert-oriented’’ are three common approaches to program evaluation in medical education. This snapshot provides a brief review of program evaluation within each approach along the following dimensions:

- **Perspective:** Do internal stakeholders or external stakeholders drive the evaluation?
- **Sensitivity:** Is the evaluation examining a broad scope of the program or delving deeply into one or two particular aspects in greater detail?
- **Feasibility:** Is this evaluation reasonably straightforward to complete, or does it require specialized knowledge and resources?
- **Utility:** Do the evaluation results benefit local decision makers or those in a broader audience?
- **Integration of theory:** Does the evaluation rely on a theory, or is it largely atheoretical?

**Decision-oriented approach**
The evaluation results help program personnel make effective decisions. The type of data included in the research design of, and the focus of evaluation are selected to maximize the evaluators’ utilization of evaluation results.

- **Perspective:** Informed by the needs of the program personnel
- **Sensitivity:** In-depth look into data at each stage of the educational process
- **Feasibility:** Can be completed with local resources but may be limited by the availability and sufficiency of the data
- **Utility:** Results largely favor local context and local decision makers, as program data are structured to reflect stages of the program (e.g., input, process, output)
- **Integration of theory:** Not directly theoretical, but the process of evaluation may draw out a theory underlying the program by identifying data points which personnel believe represent the input, process, and output of the program

**Outcomes-oriented approach**
Objectives are solidified so that specific outcome measures can be established and tracked. The evaluation determines whether the program objectives have been met.

- **Perspective:** Primarily for internal feedback to explore educational processes which lead to selected outcomes
- **Sensitivity:** Evaluation results reflect the breadth of a program’s process
- **Feasibility:** Evaluating the relationship between input (e.g., students’ knowledge, skills, and attitudes prior to participation in the program) and output and outcomes (e.g., proximal or distal curriculum objectives or students’ postparticipation knowledge, skills, and attitudes) may require specially trained educationalists
- **Utility:** Results are generally highly contextual and useful for local program and curriculum planning and development; however, results may provide broader utility by explaining educational effectiveness of programs across a spectrum of outcomes
- **Integration of theory:** Helpful for drawing out underlying assumptions and for framing the activities of the program

**Expertise-oriented approach**
The evaluator relies on an external expert to determine the value of various program criteria and data points, and the program evaluation results are judged by an expert.

- **Perspective:** Externally driven process for identifying which data points (e.g., duty hours, types of surgical cases observed, or number of publications) represent quality
- **Sensitivity:** Often a broad look across the program, but results could trigger in-depth analysis of some aspects
- **Feasibility:** Generally approachable with local resources, though some elements, such as gathering and analyzing qualitative data, may require additional specialization
- **Utility:** Results generally framed to meet external requirements
- **Integration of theory:** Generally atheoretical, as data are included to demonstrate standards of performance

**References**

**Author contact:** Rebecca.Blanchard@baystatehealth.org
Common Qualitative Methodologies and Research Designs in Health Professions Education

H. Carne Chen, MD, PhD, professor of clinical pediatrics, and Arianne Teherani, PhD, professor of medicine, University of California, San Francisco

Qualitative research includes many methodological approaches or research designs; we present the five most commonly used in health professions education. Choice of methodology will depend on the focus of inquiry and the framing of the research question. Each methodology has a specific goal. While data collection strategies (e.g., interview, focus group, observation, documentary reviews) often overlap, the approach to data analysis varies for each methodology, resulting in different research outcomes. Understanding the key features of each methodology will help researchers choose the best methodological fit for their research question.

Sample scenario: The pediatrics ward has implemented family-centered ward rounds. Concerns arise regarding didactic teaching in front of patients. Some suggest using role modeling of patient care as a teaching strategy. You wonder how students perceive teaching on family-centered rounds, if role modeling is perceived as teaching, and whether teachers deliberately role model.

<table>
<thead>
<tr>
<th>Grounded Theory</th>
<th>Phenomenology</th>
<th>Ethnography</th>
<th>Case Study</th>
<th>Narrative</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Goal/purpose</strong></td>
<td><strong>Goal/purpose</strong></td>
<td><strong>Goal/purpose</strong></td>
<td><strong>Goal/purpose</strong></td>
<td><strong>Goal/purpose</strong></td>
</tr>
<tr>
<td>Develop a theoretical model for how a process or action works</td>
<td>Understand the nature of a phenomenon, incident, or circumstance through those who experienced it</td>
<td>Describe and interpret a group’s culture/process by examining its behaviors</td>
<td>Develop an in-depth understanding of one or a small number of cases</td>
<td>Explore, in depth, one or more individuals’ longitudinal experience(s)</td>
</tr>
<tr>
<td><strong>Unit of analysis</strong></td>
<td><strong>Unit of analysis</strong></td>
<td><strong>Unit of analysis</strong></td>
<td><strong>Unit of analysis</strong></td>
<td><strong>Unit of analysis</strong></td>
</tr>
<tr>
<td>Process, action, or interaction (e.g., learning on rounds or role modeling as teaching strategy)</td>
<td>Perception of an event or experience (e.g., teaching or role modeling)</td>
<td>Group sharing a culture (e.g., senior residents)</td>
<td>Bounded event, activity, or program (e.g., rounds on specific ward)</td>
<td>One or more individuals (e.g., senior clinician teacher)</td>
</tr>
<tr>
<td><strong>Potential research question</strong></td>
<td><strong>Potential research question</strong></td>
<td><strong>Potential research question</strong></td>
<td><strong>Potential research question</strong></td>
<td><strong>Potential research question</strong></td>
</tr>
<tr>
<td>How do students learn on family-centered rounds?</td>
<td>How do team members define teaching on family-centered rounds?</td>
<td>Do senior residents incorporate “role model” as one of their roles?</td>
<td>How does teaching occur during family-centered rounds on a ward with high evaluation scores?</td>
<td>How has one clinical teacher’s experience with teaching on rounds evolved with changes in ward structure and over her career?</td>
</tr>
<tr>
<td>How does role modeling impact their learning?</td>
<td>Does it include role modeling?</td>
<td>Do they consciously model patient care for junior learners?</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Approach to data analysis</strong></td>
<td><strong>Approach to data analysis</strong></td>
<td><strong>Approach to data analysis</strong></td>
<td><strong>Approach to data analysis</strong></td>
<td><strong>Approach to data analysis</strong></td>
</tr>
<tr>
<td>Analyze by categorizing and relating data (coding) to generate a model of the process or action</td>
<td>Analyze for significant statements, units of meaning, and the what and/or how of participant experiences</td>
<td>Analyze the group’s behaviors for themes</td>
<td>Analyze for key themes important to understanding the case</td>
<td>Analyze story for key elements</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Interpret themes to develop cultural portrait of the group</td>
<td>Conduct thematic analysis across cases if using multiple cases</td>
<td>Develop themes with an emphasis on sequence, turning points, and context</td>
</tr>
<tr>
<td><strong>Potential study outcome</strong></td>
<td><strong>Potential study outcome</strong></td>
<td><strong>Potential study outcome</strong></td>
<td><strong>Potential study outcome</strong></td>
<td><strong>Potential study outcome</strong></td>
</tr>
<tr>
<td>Theoretical model explaining what and who students attend to on rounds and what factors allow an event to become a learning event</td>
<td>Description of the concept of teaching on rounds, and whether role modeling is experienced as teaching</td>
<td>Understanding of senior residents’ role modeling beliefs/behaviors and whether they see it as a responsibility</td>
<td>Recommended best practices for exemplar case for teaching on family-centered rounds</td>
<td>Understanding of how teaching on rounds can evolve as faculty gain experience and/or in response to changes in work environment</td>
</tr>
</tbody>
</table>

Reference:

Additional resources:

Author contact: hcarnechen@gmail.com
## Case Study Research in Health Professions Education

Sarah A. Bunton, PhD, research director, Medical School Operations, and Shana F. Sandberg, PhD, former senior research analyst, Center for Workforce Studies, Association of American Medical Colleges

In health professions education and the sciences, case-based teaching strategies—through which instruction and learning occur through discourse around specific, contextualized cases—are the norm. Case-based reports, in their detailed reporting of symptoms, diagnosis, treatment, and follow-up of patients, are also contextualized, and have facilitated new disease recognition and effects of treatments. Both case-based teaching and case-based medical reports provide a useful format for discussing complex symptoms or patients and ethical challenges in context.

Likewise, case study research—a qualitative research strategy that investigators within health professions education may apply—represents an effective methodology for examining a phenomenon within its real-life context. While case study research has sometimes been faulted for its lack of representativeness and rigor, it can, when approached with focused design, systematic data collection, careful analysis, and quality control procedures, facilitate evaluation of and insights into the relationships among innovations or interventions and health care and medical education. In this way, the research yields unique information that would not be achievable using other approaches.

### Considerations for Case Study Research in Health Professions Education

#### Types of Design
Comprise single or multiple cases, and may be (among others):
- Intrinsic cases, selected for their uniqueness to illustrate different approaches to the issue under study, or
- Instrumental cases, selected for being “typical” cases to help others better understand the issue under study

#### Types of Answerable Questions
A descriptive question: What is happening or has happened?
- What techniques are used to train medical residents in emergency medicine?
- What new roles are health professionals occupying in a teaching hospital?

An explanatory question: How or why did something happen?
- How are nurses involved in implementing quality improvement efforts in a specialty clinic?
- Why is the new medical education program not being received well by students?

#### Data Sources
Versatile; can employ both quantitative and qualitative data, including:
- Direct observations
- Interviews, focus groups
- Records, documents, and artifacts
- Quantitative data, such as survey or test results, to complement qualitative findings

#### Analysis
Dependent on the type of case study, but:
- Data from multiple sources should be integrated to understand overall case (i.e., carefully attend to all evidence)
- The question posed (i.e., research question) should (as in other research techniques) tightly frame the findings and discussion or explanation
- Consideration of alternative interpretations is an integral part of the process

#### Potential Disadvantages and Strategies to Mitigate
Limits to generalizability. To mitigate:
- Be clear about analytic generalizations (not statistical ones, which are not appropriate)
- Include multiple units within the case or multiple cases, where possible
- Describe findings as contextual; avoid claims of generalizability

Researcher integrity. To check quality:
- Train researcher and participants in data collection
- Triangulate information (check multiple sources for the same thing)

Researcher subjectivity (ethics of process). To mitigate:
- Use protocol, maintain chain of evidence, and ensure transparency of process
- Compare findings with other cases (build on existing knowledge)

#### Unique Information Gained Through Approach
- In-depth and multifaceted accounts of complex issues and processes (e.g., how a new model of care has been implemented in a health system)
- Insight into relationships (e.g., why a program initiative has not achieved desired results)
- Potential “new avenues for research or theory development” (e.g., factors revealed in focus groups can be included in a survey to measure their effect)
- The facilitation or improvement of effective practices across the academic medicine enterprise (e.g., identifying barriers to program success could lead to interventions targeting faculty or to the development of new curricula to address gaps)

### References

### Author Contact
sbunton@aaamc.org

First published online October 4, 2016
Secondary Data Sources for Health Professions Education Research: Where to Look and What You Will Find

Michael J. Dill, director, Workforce Studies, Association of American Medical Colleges; Emily D. Yunker, MPA, FMP, manager, Assessment Programs, Physician Assistant Education Association; Katherine Brandenburg, manager, Data Operations and Services, Association of American Medical Colleges; and Marie Caulfield, PhD, manager, Data Operations and Services, Association of American Medical Colleges

<table>
<thead>
<tr>
<th>What do I want to study?</th>
<th>Students and other trainees</th>
<th>Programs and pathways</th>
<th>Practitioners</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics</td>
<td>Educational experiences</td>
<td>Demographics</td>
<td></td>
</tr>
<tr>
<td>Preferences and attitudes</td>
<td>Curriculum content</td>
<td>Preferences and attitudes</td>
<td></td>
</tr>
<tr>
<td>Test scores</td>
<td>Decision points</td>
<td>Educational history</td>
<td></td>
</tr>
<tr>
<td>Graduation rates</td>
<td>(e.g., specialty, training program)</td>
<td>Practice characteristics</td>
<td></td>
</tr>
<tr>
<td>Employment outcomes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e.g., specialty, location)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Examples of the types of information a researcher might want

<table>
<thead>
<tr>
<th>What do I want to know about them?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training program organizations</td>
</tr>
<tr>
<td>(e.g., AAMC, ACOM, PAEA, AACN, ACGME, ADEA)</td>
</tr>
<tr>
<td>• Data from entrance exams, application processes, institutions, assessments, and student surveys</td>
</tr>
<tr>
<td>Professional organizations</td>
</tr>
<tr>
<td>(e.g., AMA, AOA, ADA)</td>
</tr>
<tr>
<td>• Data from institutions</td>
</tr>
<tr>
<td>Government agencies</td>
</tr>
<tr>
<td>(e.g., NCES)</td>
</tr>
<tr>
<td>• Data from institutions</td>
</tr>
</tbody>
</table>

Where do I look for the data?

<table>
<thead>
<tr>
<th>Examples of organizations that collect the information (specific content will vary)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training program organizations (e.g., AAMC, ACOM, PAEA, AACN, ADEA)</td>
</tr>
<tr>
<td>• Surveys of institutions, faculty, and trainees</td>
</tr>
<tr>
<td>• Curriculum inventory</td>
</tr>
<tr>
<td>Accrediting bodies (e.g., LCME, ACGME, ACEN, CCNE, ARC-PA, CODA)</td>
</tr>
<tr>
<td>• Data from institutions</td>
</tr>
<tr>
<td>Government agencies (e.g., NCES)</td>
</tr>
<tr>
<td>• Data from institutions</td>
</tr>
<tr>
<td>Professional associations (e.g., AMA, AAPA, ADA)</td>
</tr>
<tr>
<td>• Surveys of members</td>
</tr>
<tr>
<td>Government agencies (e.g., CMS, U.S. Census, BLS)</td>
</tr>
<tr>
<td>• Claims data</td>
</tr>
<tr>
<td>• Population surveys</td>
</tr>
<tr>
<td>• Employment data</td>
</tr>
<tr>
<td>State licensure boards (states, FSMB, USMLE)</td>
</tr>
<tr>
<td>• Data collection during licensure process</td>
</tr>
<tr>
<td>Certifying boards and agencies (e.g., ABMS, NCCPA, AADB)</td>
</tr>
<tr>
<td>• Surveys of members</td>
</tr>
</tbody>
</table>

Examples of Academic Medicine articles using some of the sources listed above

<table>
<thead>
<tr>
<th>What has been done with the data?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positioning Medical Assistants for a Greater Role in the Era of Health Reform</td>
</tr>
<tr>
<td>(Based on NCES data)</td>
</tr>
<tr>
<td>Learning About Medical Student Mistreatment From Responses to the Medical School Graduation Questionnaire</td>
</tr>
<tr>
<td>(Based on AAMC data)</td>
</tr>
<tr>
<td>The Road to an Academic Medicine Career: A National Cohort Study of Male and Female U.S. Medical Graduates</td>
</tr>
<tr>
<td>(Based on AAMC data)</td>
</tr>
<tr>
<td>Educational and Individual Factors Associated With Positive Change in and Reaffirmation of Medical Students’ Intention to Practice in Underserved Areas</td>
</tr>
<tr>
<td>(Based on AAMC data)</td>
</tr>
<tr>
<td>Thirty Years Training Rural Physicians: Outcomes From the Michigan State University College of Human Medicine Rural Physician Program</td>
</tr>
<tr>
<td>(Based on AAMC data)</td>
</tr>
</tbody>
</table>

A list of acronyms and full citations for the listed scholarly publications is available in Supplemental Digital Appendix 1 at http://links.lww.com/ACADMED/A902.

Author contact: mjdill@aamc.org
Primer for Submitting Health Professions Education Research to the IRB

Rebecca D. Blanchard, PhD, senior director, Educational Affairs, Baystate Health, and assistant professor, University of Massachusetts Medical School-Baystate; Stephen DeMeco, DO, MEd, attending neonatologist, WakeMed Health and Hospitals, and adjunct professor, Duke University Medical Center; and Alisa Nagler, JD, EdD, assistant director, Accreditation, Validation and Credentialing, American College of Surgeons, and adjunct professor, Duke University School of Medicine

Health professions education (HPE) research often represents a gray area for institutional review boards (IRBs) since HPE projects are variably categorized as research, quality improvement, or educational evaluation that is not research, depending on a number of considerations. Below, we address common questions asked by HPE researchers, define key IRB terms, and offer tips for successfully navigating the IRB review process.

**Common Questions**

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does my project need to be submitted to the IRB?</td>
<td>HPE scholarship is considerably different than clinical research,(^1) and can often be difficult to categorize as research. A detailed description of your project—including what data you will be collecting and how you will collect those data—should be provided to your institution’s IRB to help IRB members make that determination.</td>
</tr>
<tr>
<td>How do I know if my project is research or not research?</td>
<td>To be categorized as “research” with the IRB, the project must be both a systematic investigation (deliberate plan for studying an outcome) and generalizable (intending results to be representative of what would happen at another institution with the same intervention).(^2)</td>
</tr>
<tr>
<td>Who are the subjects in HPE research?</td>
<td>“Subjects” are likely to be learners or teachers in HPE research and not likely to be patients.</td>
</tr>
</tbody>
</table>

**Key Terms**

**Human Subjects Research**

- Human subjects research (HSR) is defined as any systematic investigation, either by intervention or interaction, designed to contribute to generalizable knowledge through the collection of data—from human subjects or living individuals, or from identifiable private information.\(^3\)
- In the case of HPE research, “interaction” could include any communication or interpersonal contact between investigator and subject (likely learners). Institutions vary greatly on their determination of which educational activities are considered HSR. For example, learner surveys, curriculum evaluation, and longitudinal analysis of standardized exam outcomes may be considered HSR at one institution and not at another.

**Risk**

Risk is the possibility of a negative consequence of participating in a study, including physical or psychological harm, or breach of privacy or confidentiality. "Minimal risk" is defined as risk that exceeds risk encountered in daily life or routine examination. The determination of minimal risk serves as the basis for the type of IRB review (exempt, expedited, or full).\(^2\)

**Exempt or Expedited Review**

- In an exempt review, a study usually involves no more than minimal risk, and an investigator may waive documented informed consent from participants, but (1) must have an appropriate process of informed participation, (2) must protect privacy and confidentiality, (3) must adhere to institution-specific policies and procedures, and finally, (4) is exempt from ongoing monitoring unless the protocol changes.\(^4\)
- In an expedited review, a study may involve more than minimal risk; a protocol may include direct contact with subjects (including recordings or transcripts), noninvasive procedures, or potential for breaches of privacy.\(^2\) These studies may be reviewed by the IRB chair only, but are subject to ongoing monitoring. These usually require subjects’ written informed consent for participation.

**Tips for Successfully Submitting HPE Research to the IRB**

1. Engage in a dialogue with the IRB early in the process (especially if you are unsure if the project meets criteria for HSR) and throughout with any updates, questions, or concerns regarding your project.
2. Find medical education mentors and collaborators to help you navigate the process.
3. Consider tools and templates to aid the process and to ensure in practice and documentation that learners are protected (at the institutional and national levels).\(^5\)
4. Remind the IRB that your subjects are learners, and state clearly if no patient data are being collected, especially if your intervention occurs in a clinical space.\(^6\)
5. Be sure to collaborate with your IRB. HPE research can evoke many questions, and IRB processes and decisions vary by institution!

**References**


**Author contact:** Rebecca.Blanchard@BaystateHealth.org; Twitter: @rdblanchard1
The Ethics and Etiquette of Research Collaboration

Holly Meyer, PhD, assistant professor of medicine, Uniformed Services University of the Health Sciences; Lara Varpio, PhD, associate professor of medicine, Uniformed Services University of the Health Sciences; Larry Gruppen, PhD, professor of learning health sciences and director, Master of Health Professions Education program, University of Michigan; and Gurjit Sandhu, PhD, assistant professor of surgery, University of Michigan

Collaboration among scholars benefits the scale of studies, broadens the scope of the health professions education (HPE) field, and further develops the skills of each scholar. Individuals working together for a common purpose increase the rigor and reach of scholarship. While there are numerous perks to collaboration, most researchers have also had to learn how to navigate the challenges of working in teams. The dilemmas HPE scholars face often stem from negotiating the ethics and etiquette of collaborative research. We explore ethical and etiquette considerations in two collaborative situations that HPE researchers commonly encounter.

**Research Dilemmas**

I am currently collaborating on a research team, and I have a spin-off idea. What are my obligations to include the original study team in my spin-off work?

**Ethical Considerations**

Recommendating and defending right and wrong conduct

- Research data are considered an asset of the principal investigator’s institution.  
- Withholding evidence and/or findings from the team that germinate during the research process is akin to interference and could be considered misconduct.

**Ethique Considerations**

Expectations for accepted behaviors

- Consider using a written collaboration agreement between authors to clarify access to original data and expectations around future use and publication.
- For each collaborator, ask: Would I have had this new idea without this team member? If not, consider inviting him/her to collaborate on the spin-off study.

When should I add or remove a collaborator from the research team?

**Authorship guidelines require participation:**

- In conception and design OR data collection and analysis, AND
- In drafting or revising the publication critically, AND
- Via approval of final publication, AND
- Via following up on all integrity and accuracy inquiries.

**Recommended Practices:** Early in the collaboration, explicitly discuss with the team:

1) Author requirements
2) How extensions of the study will be handled
3) A written collaboration agreement for authors

For a list of institutional Web sites or publications relating to ethical conduct, please see Supplemental Digital Appendix 1 at http://links.lww.com/ACADMED/A395.

**Acknowledgment:** The authors would like to thank Paul Trombley for his contribution to graphic design.

**Disclaimer:** The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Uniformed Services University of the Health Sciences, the Department of the Navy, the Department of Defense, or the U.S. Government.

**References:**


**Author contact:** holly.meyer.cs@usuhs.edu
Key Sampling Issues in Quantitative Research in Health Professions Education

Sonia Crandall, PhD, professor and director, Research and Scholarship, Department of Physician Assistant Studies, Reamer Bushardt, PharmD, professor and chair, Department of Physician Assistant Studies, and Edward Ip, PhD, MA, professor, Department of Biostatistics, Wake Forest School of Medicine

Background: The primary goal of sampling is to obtain a representative sample of the targeted population within the constraints set by the study. In health professions education, poor sampling design can lead to biased estimates, limited generalizability, and underpowered studies. Consider, for example, the results of a single-site study involving a convenience sample of the 10 residents from a radiology department. This sample design will likely produce results of little scientific value to other radiology residency programs.

Hypothetical case: Five faculty members from institutions in different states attend a regional Group on Educational Affairs meeting. They participate as a team in a workshop on career development for early-career faculty. As a follow-up, they decide to design a study that investigates the faculty development needs of early-career faculty members at U.S. medical schools. Specifically, they focus on faculty members at the rank of instructor or assistant professor. Using the 2015 Association of American Medical Colleges (AAMC) report on the distribution of U.S. medical school faculty, the research team (RT) knows that the potential population from which a sample can be drawn includes 88,277 faculty members at those ranks: 47,435 men (54%) and 40,842 women (46%). The RT creates a data collection instrument which includes objectives adapted from Successful Faculty in Academic Medicine. These items ask the faculty members to respond to their current and desired level of expertise for each objective. The RT considers the pros and cons of the following sampling designs.

<table>
<thead>
<tr>
<th>Design</th>
<th>Illustrative description</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
</table>
| Convenience sampling    | The RT selects the instructor and assistant professor attendees at a regional Group on Educational Affairs meeting. | • Easier to implement than other sampling designs  
• Least expensive  
• Appropriate for a small pilot study | • Participants may not be representative of the population of faculty  
• Prone to bias |
| Simple random sampling  | The RT randomly selects 1,000 faculty from the AAMC list of 88,277 using a random number generator. A power analysis will help researchers to determine the most appropriate sample size. A response rate of xx% (set a priori) is considered acceptable for this survey. | • Unbiased  
• Generalizable  
• A reasonably large sample (e.g., 1,000) usually will be representative and is the number of respondents often used in polls | • Involves a large number of institutions, so it may be expensive and difficult to interact with the participants |
| Stratified random sampling | Possible stratifications (i.e., subgroups) are rank, gender, region, and/or department type. The RT stratifies by gender and randomly selects 540 men and 460 women (representing the percentage of women and men in the population). | • Unbiased  
• Generalizable  
• Presents a more equal representation of the population | • Involves a large number of institutions, so it may be expensive and difficult to interact with the participants  
• Difficult to implement if too many strata are applied |
| Clustered random sampling | The RT randomly selects 10 schools from AAMC member schools (each school is a cluster). | • Easier to implement  
• Less expensive than a simple random sampling strategy  
• Appropriate for a multisite study | • Participants may not be representative of the population of faculty, especially when the number of clusters is small (if there is only one cluster, the cluster is a single-site sample, which resembles a convenience sample) |

Researchers must carefully think about sampling design before they start to collect their data because of the impact that the design has on the generalizability of the findings.

References:

Author contact: crandall@wakehealth.edu; Twitter @SoniaCrandall
AM Last Page: Survey Development Guidance for Medical Education Researchers

Hunter Gehlbach, PhD, assistant professor of Education, Harvard University; Anthony R. Artino, Jr, PhD, assistant professor of Preventive Medicine and Biometrics, Uniformed Services University of the Health Sciences; and Steven J. Durning, MD, professor of Medicine, Uniformed Services University of the Health Sciences.

Medical education researchers frequently rely on survey data. For example, of Academic Medicine’s 141 research articles from 2009, over half (56%) used surveys. Yet, the literature provides limited guidance on which processes best facilitate the development of surveys—particularly in the design of survey scales (i.e., several items that assess a single underlying construct such as physician empathy or teaching self-efficacy; see example below). This flowchart presents seven steps to facilitate the construction of valid and reliable survey scales.

**Step 1**
Conduct a literature review both to ensure that your construct definition aligns with relevant prior research and to identify extant survey scales or items that might be used or adapted for your research context.

**Step 2**
Conduct interviews and/or focus groups to learn how your population of interest conceptualizes and describes your construct of interest.

**Step 3**
Synthesize the literature review and interview/focus group data so that the conceptualization of the construct makes theoretical sense to scholars in the field and uses language that your population of interest understands. For example, a scale assessing teaching self-efficacy (i.e., confidence in one’s teaching ability) should use words like “confidence in trying out new teaching techniques,” not “efficaciousness in experimenting with novel pedagogies.”

**Step 4**
Develop items in accordance with current best practices in survey design. For example, the sample scale below uses response anchors that refer to the specific construct (rather than numbers or agree/disagree response anchors).

**Step 5**
Conduct an expert validation to assess the items’ clarity and relevance to the construct.

**Step 6**
Conduct cognitive pretesting through which participants restate each item aloud in their own words as they answer it. This step helps ensure that respondents interpret items in the manner that you intend.

**Step 7**
Pilot-test your items to check for adequate item variance, reliability, and convergent/discriminant validity with respect to other measures.

*Note:* After you complete each of these final steps, you may need to revise items and/or repeat steps from this part of the process.

**Sample Items From a Teaching Self-Efficacy Scale**

<table>
<thead>
<tr>
<th>Item</th>
<th>Not at all confident</th>
<th>Slightly confident</th>
<th>Moderately confident</th>
<th>Quite confident</th>
<th>Extremely confident</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How confident are you that you can help students remember what they learned in your class?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. When you need to teach less interesting topics, how confident are you that you can keep all students engaged?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. How confident are you that you can help students learn when they are unmotivated?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. How confident are you that you can get through to the most difficult students?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

References
Using Qualitative Research as a Means to an Effective Survey Instrument

Sarah A. Bunton, PhD, research director, Association of American Medical Colleges

The academic medicine and health care fields have seen a recent spate of studies using various qualitative research methods. Even more recently, a move to establish standards for reporting qualitative research has emerged. The contributions of these qualitative approaches, which facilitate rich insight into an issue by focusing on the how and why of phenomena, are unique and important. Another way qualitative methods contribute to research is through their systematic use in the survey instrument development process. That is, explicitly applying steps of qualitative methods during the development of a quantitative survey can strengthen the instrument and yield more meaningful results.

Instrument Development Process, Enhanced Using Qualitative Methods

**INITIAL INTERVIEWS with key stakeholders and representative members of the intended survey population**

Interviews can enhance understanding of an issue and help formulate a more comprehensive topic guide. The purpose of an initial interview is to explore responses from different people to identify all the critical aspects of, and gather in-depth insights into, the topic before designing the survey.

_Do you think about “climate” when you evaluate your job situation? Why or why not?_

**SURVEY DRAFT**

<table>
<thead>
<tr>
<th></th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Neither agree nor disagree</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I have meaningful interactions with departmental colleagues.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**FOCUS GROUPS with stakeholders and members of the target research population**

Semistructured focus groups help refine different content domains—developed in part though initial interviews—for the survey instrument. Focus group questions (as opposed to a formalized questionnaire) serve as guides, and emergent themes from the discussions can be used to generate specific survey items containing common and understandable language.

_Do you consider when judging whether your department is providing a work climate that you like?_

**COGNITIVE INTERVIEWS** with representatives of the survey population

These interviews can be used to test the adequacy of the instrument and gauge for response variations or difficulty comprehending particular items. Cognitive interviews, which should occur in the pre-testing phase of survey development, lead to the conceptual adequacy of the survey instrument items.

**FINAL SURVEY DRAFT**

<table>
<thead>
<tr>
<th></th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Neither agree nor disagree</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colleagues in my department generally support one another.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I feel appreciated by my departmental colleagues.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Designing high-quality survey instruments can be challenging. Systematic use of qualitative methods during the instrument development process can lead to enhanced survey outcomes, including:**

- a comprehensive understanding of the issues impacting the study,
- the discovery of issues or response options that may have been missed,
- insight into crafting focused items, and
- a more robust, accurate, and actionable survey.

**References:**


**Author contact:** sbunton@aamc.org
## AM Last Page: Avoiding Five Common Pitfalls of Survey Design

Anthony R. Artino, Jr, PhD, assistant professor of preventive medicine and biometrics, Uniformed Services University of the Health Sciences, Hunter Gehlbach, PhD, assistant professor of education, Harvard University, and Steven J. Durning, MD, professor of medicine and psychiatry, Uniformed Services University of the Health Sciences

Writing good survey items is both an art and a science. Over the last 30 years, scholars have amassed a great deal of scientific evidence on which questionnaire designers can rely. The guidelines below present some of the more frequently ignored, but more important, of these survey-design basics.

<table>
<thead>
<tr>
<th>Pitfall</th>
<th>Survey example(s)</th>
<th>Why it’s a problem</th>
<th>Solution(s)</th>
<th>Survey example(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creating a double-barreled item</td>
<td>How often do you talk to your nurses and administrative staff when you have a problem? Respondents have trouble answering survey items that contain more than one question (and thus could have more than one answer). In this example, respondents who talk to nurses often but talk to administrative staff infrequently will struggle to answer this question. Survey items should address one idea at a time.</td>
<td>When you have multiple questions/theses within a given item, either (1) create multiple items for each question that is important or (2) include only the more important question. Be especially wary of conjunctions in your items.</td>
<td>How often do you talk to your nurses when you have a problem? How often do you talk to your administrative staff when you have a problem?</td>
<td></td>
</tr>
<tr>
<td>Creating a negatively worded item</td>
<td>In an average week, how many times are you unable to start class on time? The chief resident should not be responsible for denying admission to patients. Negatively worded survey items are challenging for respondents to comprehend and answer accurately. Double-negatives are particularly problematic and increase measurement error. If a respondent has to say &quot;yes&quot; in order to mean &quot;no&quot; (or &quot;agree&quot; in order to &quot;disagree&quot;), the item is flawed.</td>
<td>Make sure “yes” means yes and “no” means no. This generally means wording items positively.</td>
<td>In an average week, how many times do you start class on time? Should the chief resident be responsible for admitting patients?</td>
<td></td>
</tr>
<tr>
<td>Using statements instead of questions</td>
<td>I am confident I can do well in this course. • not at all true • a little bit true • somewhat true • mostly true • completely true</td>
<td>A survey represents a conversation between the surveyor and the respondents. To make sense of survey items, respondents rely on “the tacit assumptions that govern the conduct of conversation in everyday life.” Only rarely do people engage in rating statements in their everyday conversations.</td>
<td>Formulate survey items as questions. Questions are more conversational, more straightforward, and easier to process mentally. People are more practiced at responding to them.</td>
<td></td>
</tr>
<tr>
<td>Using agreement response anchors</td>
<td>The high cost of health care is the most important issue in America today. • strongly disagree • disagree • neutral • agree • strongly agree</td>
<td>Agreement response anchors do not emphasize the construct being measured and are prone to acquiescence (i.e., the tendency to endorse any assertion made in an item, regardless of its content). In addition, agreement response anchors may encourage respondents to think through their responses less thoroughly while completing the survey.</td>
<td>Use construct-specific response anchors that emphasize the construct of interest. Doing so reduces acquiescence and keeps respondents focused on the construct in question. Doing so results in less measurement error.</td>
<td></td>
</tr>
<tr>
<td>Using too few or too many response anchors</td>
<td>How useful was your medical school training in clinical decision making? • not at all useful • somewhat useful • very useful</td>
<td>The number of response anchors influences the reliability of a set of survey items. Using too few response anchors generally reduces reliability. There is, however, a point of diminishing returns beyond which more response anchors do not enhance reliability.</td>
<td>Use five or more response anchors to achieve stable participant responses. In most cases, using more than seven to nine anchors is unlikely to be meaningful to most respondents and will not improve reliability.</td>
<td></td>
</tr>
</tbody>
</table>

**References:**

**Disclaimers:**
The views expressed in this article are those of the authors and do not necessarily reflect the official policy of the Department of Defense.
Dr. Steven Durning cautioned this Last Page prior to becoming assistant editor, AM Last Page.
# AM Last Page: Avoiding Four Visual-Design Pitfalls in Survey Development

Anthony R. Artino, Jr., PhD, associate professor, Preventive Medicine and Biometrics, Uniformed Services University of the Health Sciences, and Hunter Gehlbach, PhD, associate professor, Harvard Graduate School of Education

A previous AM Last Page presented five common pitfalls of survey design as well as several solutions. This AM Last Page presents four visual-design and layout pitfalls and offers solutions.

<table>
<thead>
<tr>
<th>Pitfall: Explanation and Example</th>
<th>Solution: Explanation and Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Labeling only the end points of your response options</strong></td>
<td><strong>Verbally label each response option</strong></td>
</tr>
<tr>
<td>Labeling only the end points leaves the meaning of the unlabeled options open to respondents’ interpretation. Different respondents can interpret the unlabeled options differently. This ambiguity increases measurement error.²</td>
<td>Labeling each response option increases consistency in the conceptual spacing between response options and increases the likelihood that all respondents will interpret the response options similarly. Additionally, the response options have comparable visual weight, so the respondents’ eyes are not drawn to certain options.</td>
</tr>
<tr>
<td>How interesting did you find this clinical reasoning course?</td>
<td>How interesting did you find this clinical reasoning course?</td>
</tr>
<tr>
<td>○ not at all interesting</td>
<td>○ not at all interesting</td>
</tr>
<tr>
<td>○ slightly interesting</td>
<td>○ moderately interesting</td>
</tr>
<tr>
<td>○ moderately interesting</td>
<td>○ quite interesting</td>
</tr>
<tr>
<td>○ extremely interesting</td>
<td>○ extremely interesting</td>
</tr>
</tbody>
</table>

| **Labeling response options with both numbers and verbal labels** | **Use only verbal labels** |
| Because of the additional information respondents must process, providing both numbers and verbal labels extends response time.³ The implied meaning of negative numbers can be particularly confusing and may introduce additional error. For example, in the item below, learning “a little bit” seems incongruous with learning the amount of “−1.” | In general, use only verbal labels for each response option. Doing so will reduce the cognitive effort required of your respondents and will likely reduce measurement error.² |
| How much did you learn in today’s workshop? | How much did you learn in today’s workshop? |
| ○ −2 almost nothing | ○ almost nothing |
| ○ −1 a little bit | ○ a little bit |
| ○ 0 some | ○ some |
| ○ 1 quite a bit | ○ quite a bit |
| ○ 2 a tremendous amount | ○ a tremendous amount |

| **Unequally spacing your response options** | **Maintain equal spacing between response options** |
| The visual spacing between options can attract respondents to certain options over others, which in turn might cause them to select these options more frequently. In addition, unbalanced spacing of the response options can shift the visual midpoint of the scale. | Maintaining equal spacing between response options will reinforce the notion that, conceptually, there is equal space or “distance” between each response option. As a result, the answers will be less biased, thereby reducing measurement error. |
| How much did you learn from your peers in this course? | How much did you learn from your peers in this course? |
| ○ almost nothing | ○ almost nothing |
| ○ a little bit | ○ a little bit |
| ○ some | ○ some |
| ○ quite a bit | ○ quite a bit |
| ○ a tremendous amount | ○ a tremendous amount |

| **Placing nonsubstantive response options together with substantive response options** | **Use additional space to visually separate nonsubstantive response options** |
| Placing nonsubstantive response options such as “don’t know,” “no opinion,” or “not applicable” together with the substantive options can shift the visual and conceptual midpoint of the response scales, thereby skewing the results.⁴ | Using additional space to visually separate nonsubstantive response options from the substantive options will align the visual midpoint with the conceptual midpoint, thereby reducing measurement error.⁴ This recommendation is an important exception to the guidance above about maintaining equal spacing between response options. |
| How satisfied are you with the quality of the library services? | How satisfied are you with the quality of the library services? |
| ○ not at all satisfied | ○ not at all satisfied |
| ○ slightly satisfied | ○ slightly satisfied |
| ○ moderately satisfied | ○ moderately satisfied |
| ○ quite satisfied | ○ quite satisfied |
| ○ extremely satisfied | ○ extremely satisfied |
| ○ not applicable | ○ not applicable |

---

**Disclaimer:**

The views expressed in this article are those of the authors and do not necessarily reflect the official policy of the U.S. Department of Defense.

**References:**


**Author contact:** artino@usuhs.edu
# The Tools of the Qualitative Research Trade

Elise Paradis, PhD, Canada Research Chair in Collaborative Healthcare Practice, assistant professor, Leslie Dan Faculty of Pharmacy and Department of Anesthesia, Faculty of Medicine, and scientist, The Wilson Centre, University of Toronto and University Health Network

Choosing the right qualitative research method is like choosing the right tool: When trying to cut a plank, a hammer is mostly useless. A key question to ask when attempting to choose the right qualitative research method is thus, *What are you trying to do?* Identifying your goal should help you choose the right method. Remember, however, that (1) your tools will be more useful if carried in the right toolbox (i.e., methodology or approach), (2) your research will be more impactful if you join a scholarly conversation by using theory and relating your work to that of others, and (3) each method has pros and cons.

<table>
<thead>
<tr>
<th>METHOD</th>
<th>GOAL</th>
<th>SAMPLE QUESTION</th>
<th>ISSUES TO CONSIDER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>INTERVIEWS</strong>[^2,^3]</td>
<td>Discover individual perceptions, experiences, or understandings of a specific topic, situation, or phenomenon.</td>
<td>How do residents working in the intensive care unit perceive the role of palliative care services?</td>
<td>Who should you be interviewing (sample)? Does your interviewer have the right training to elicit rich responses? Is his/her role likely to bias interviewees' responses? Is the topic likely to cause reputational or psychological harm to participants?</td>
</tr>
<tr>
<td><strong>FOCUS GROUPS</strong>[^4,^5]</td>
<td>Explore the range of perspectives on a topic within (and sometimes between) different stakeholder groups.</td>
<td>How do students and faculty members see the undergraduate professionalism curriculum? How is it meeting (or not meeting) their needs?</td>
<td>Is your moderator skilled enough to manage group discussions and different personalities? Is his/her professional role likely to alter participants' responses? Is the topic covered too private for a group discussion? Are hierarchies among group members problematic?</td>
</tr>
<tr>
<td><strong>OBSERVATIONS</strong>[^2,^5]</td>
<td>Develop an understanding of actual rather than narrated behavior. Situate behavior in its broader context. Evaluate the impact of a policy on practice.</td>
<td>How are faculty members modeling the CanMEDS Advocacy Role for students?</td>
<td>What kinds of observational data are you interested in—count data, workflow data, social interaction data? Do you have enough time to observe or to train an observer? Do you have access to sites? Do the ethics of your clinical role conflict with the nature of your role as a researcher?</td>
</tr>
<tr>
<td><strong>TEXTUAL ANALYSIS</strong>[^6]</td>
<td>Make discourse—the main sociohistorical influences on our world—visible. Explore relationships among people, organizations, and institutions, over time and in a specific place.</td>
<td>How has accreditation been used as an argument to implement new educational interventions since the 1960s?</td>
<td>Are there readily available and legitimate documents to answer your research question? Are you able to and interested in paying close attention to subtle changes in language? Do you have time to read hundreds if not thousands of pages, iteratively develop a coding scheme, and read and code again?</td>
</tr>
</tbody>
</table>

References:


*Author contact:* elise.paradis@utoronto.ca
Four Common Pitfalls of Quantitative Analysis in Experimental Research

Jimmie Leppink, PhD, postdoctoral researcher, Ellen M. Kok, MSc, PhD student, Esther M. Bergman, PhD, assistant professor, Mariëtte H. van Loon, PhD, assistant professor, and Anique B.H. de Bruin, PhD, associate professor, Maastricht University

A recently published AM Last Page presents five common methodological pitfalls of experimental research in medical education. In this Last Page, we present four statistical pitfalls and their more appropriate alternatives. Pitfalls are illustrated with a case of a fictitious researcher who conducts a study with elements that are common in many medical education experiments (Figure 1).

![Diagram of Pitfall 1: Treating a two-factor design as a one-factor design]

**Pitfall 1: Treating a two-factor design as a one-factor design**

To test for differences between the four conditions in the experiment, the researcher uses a one-way analysis of variance (ANOVA) with four groups.

**Problem with the one-way ANOVA approach**
- It does not test for an interaction effect between training and feedback, but compares each combination of two groups (see Figure 2, below).
- It is less likely to detect main effects of training and feedback.

**Alternative**
- A two-way (2x2) ANOVA that tests for a main effect of feedback, a main effect of training, and an interaction effect between training and feedback (see Figure 2, below).

![Graph showing One-way ANOVA (left) versus 2x2 ANOVA (right)]

**Pitfall 2: Treating the pretest and posttest as repeated measures instead of treating the pretest as a covariate**

In the experiment participants completed a pretest and posttest. The researcher treats the pre- and posttest as repeated measures in a within-subjects ANOVA.

**Problem with the repeated-measures approach**
- It is appropriate in quasi-experimental (nonrandomized group comparison) studies or when there are pretest differences between treatment conditions.
- When randomization of participants has resulted in no significant differences in pretest performance, there is no need to further test for these differences (Figure 3).

**Alternative**
- When successful randomization has resulted in no significant differences in pretest performance, treating pretest as a covariate in an analysis of covariance (ANCOVA) provides a more powerful test, because ANCOVA is more parsimonious than repeated-measures ANOVA.

![Graph showing Statistical analysis depends on whether there is a significant difference between groups on the pretest]

**Pitfall 3: Considering the time-on-task as a covariate versus a mediator**

The training took on average 1 hour and the feedback session on average ½ hour, resulting in differences in time-on-task between treatment conditions. In the analysis, the researcher includes time-on-task in an ANCOVA.

**Problem with the ANCOVA approach**
- The treatment influences time-on-task (or another variable of interest), and time-on-task may affect performance.
- Including time-on-task in an ANCOVA will either underestimate or overestimate the treatment effect.

**Alternative**
- Path analysis enables for treating time-on-task as a mediator and allows estimating both direct and indirect effects of treatment (see Figure 4).
- The total treatment effect is the sum of the treatment effect mediated by time-on-task (i.e., indirect effect) and the treatment effect not mediated by time-on-task (i.e., direct effect).

![Diagram showing Direct and indirect effect of treatment condition on performance]

**Pitfall 4: Ignoring a hierarchical structure of data by performing “ordinary” regression instead of multilevel regression**

Posttest performance and mental effort are measured per task. To analyze their relation, the researcher wants to average individual scores across tasks or treat every individual-by-task combination as an independent observation.

**Problem with these approaches**
- The independent-observations approach ignores the intrindividual correlation between the five tasks. Averaging over five tasks means loss of information.
- Both approaches can result in an incorrect interpretation of the correlation of interest.

**Alternative**
- Using multilevel models constitutes a best practice for dealing with intra-individual and/or intraclass correlations (see Figure 5).

![Diagram showing What can happen when using the averaging or independent observation approach (left), when the multilevel approach is more appropriate (right)]

References


Author contact: jimmie.leppink@maastrichtuniversity.nl
Avoiding Common Data Analysis Pitfalls in Health Professions Education Research

Jimmie Leppink, PhD, postdoctoral fellow, School of Health Professions Education, Maastricht University; and Kulamakan M. Kulasegaram, PhD, assistant professor, Department of Family Medicine, University of Toronto

Two recently published AM Last Pages present several study design\(^1\) and quantitative analysis\(^2\) pitfalls in experimental research. In this AM Last Page, we present four common data analysis pitfalls in health professions education research from our perspective as researchers, reviewers, and editors. We use a fictitious study, with three skill test moments for two groups of residents, that shares elements with a wide variety of studies (Figures 1 and 2).

### Mean skill test score (0–50) for two groups of residents

Bars around means indicate standard deviations

<table>
<thead>
<tr>
<th></th>
<th>Residents group A</th>
<th>Residents group B</th>
</tr>
</thead>
<tbody>
<tr>
<td>At start (day 1)</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>Halfway (after 3 weeks)</td>
<td>30</td>
<td>32</td>
</tr>
<tr>
<td>At the end (after 6 weeks)</td>
<td>35</td>
<td>30</td>
</tr>
</tbody>
</table>

**Figure 1:** No group-by-time interaction

<table>
<thead>
<tr>
<th></th>
<th>Residents group A</th>
<th>Residents group B</th>
</tr>
</thead>
<tbody>
<tr>
<td>At start (day 1)</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>Halfway (after 3 weeks)</td>
<td>30</td>
<td>32</td>
</tr>
<tr>
<td>At the end (after 6 weeks)</td>
<td>35</td>
<td>30</td>
</tr>
</tbody>
</table>

**Figure 2:** Group-by-time interaction

<table>
<thead>
<tr>
<th>Pitfall</th>
<th>Explanation and Advice</th>
</tr>
</thead>
</table>
| 1. **Needless hypothesis testing** | Hypothesis testing makes sense only when you wish to generalize the findings to a larger population. Hypothesis testing may be meaningless when samples are very small or very large, or when the entire population of interest is available.\(^3\) **Advice:**
  - Presenting frequencies, means, standard deviations, ranges, other key descriptive numbers, and when possible a graphical representation (e.g., Figures 1 and 2) is a prerequisite, should occur before hypothesis testing, and is always meaningful to audiences.\(^5\)
  - Be cautious in interpreting outcomes of statistical hypothesis tests, especially in small samples.\(^3\)

| 2: **Reporting means but no standard deviations** | When reporting means and mean differences to compare groups or conditions, standard deviations are sometimes omitted. **Advice:**
  - Standard deviations can be useful for evaluating means in terms of the size of the difference (effect size) and for conducting power analyses or required sample size calculations for subsequent studies that involve hypothesis testing.\(^3\)
  - Reporting confidence intervals is a supplement to help readers understand the uncertainty around sample estimates.\(^3\)

| 3: **Simple effects tests without testing for interaction** | Studies that compare groups over time test group differences for each time point. This practice increases the likelihood of type I errors (i.e., concluding a difference exists when there is none) in some cases, and the likelihood of type II errors (i.e., concluding no difference when there is one) in other cases. **Advice:**
  - Unless you are dealing with time-point-specific comparisons with a study endpoint as the primary outcome, or a purely descriptive study, the appropriate analysis would be an omnibus approach, including group, time, and group-by-time interaction altogether in one analysis prior to post hoc testing at each time point.\(^4\)
  - Start with an overall test on interaction and perform tests on group differences per time point only when there is a significant interaction (e.g., Figure 2); otherwise, focus on main effects of group and/or time (e.g., Figure 1)—whichever is of interest.\(^4\)

| 4: **Testing on individual items that form one scale** | Some studies report tests on group differences on individual items in a test or questionnaire which, based on appropriate psychometric analysis, can be considered indicators of the same skill or other underlying variable of interest (Figure 2), or of a limited number of constructs (Figure 4). In line with Pitfall 3, this increases the likelihood of type I and/or type II errors. **Advice:**
  - Perform appropriate psychometric analysis (e.g., factor analysis) to explore which items may be grouped together, and perform tests on group differences per summary for each group of items (Figures 1 and 2).\(^4\)

**References:**

**Author contact:** jimmie.leppink@maastrichtuniversity.nl; twitter: @jimmieleppink

---

First published online October 4, 2016
**P Value: Pitfalls and Solutions**
Timothy J. Wood, PhD, associate professor, Department of Innovation in Medical Education, Faculty of Medicine, University of Ottawa

Often, quantitative researchers strive for the magical P < .05 without considering the strength of the effect or the likelihood that a result could be replicated. This focus solely on the P value can unfortunately lead to a misinterpretation of a result. The purposes of this AM Last Page are (1) to illustrate the danger of overlying on P values to judge results and (2) to suggest methods that will help researchers interpret their findings.

**Case Study 1**

To study the benefits of a communication skills workshop on history-taking skills, 100 medical students are assigned to a workshop and 100 students are assigned to a control group. History-taking skills are subsequently tested during a 10-minute OSCE station.

The mean OSCE score for the intervention group is 6.6/10 (SD = 1.9), and the mean score for the control group is 6.1/10 (SD = 1.3). The difference is statistically significant, P = .05.

These case studies with simulated data reflect the dangers of relying solely on a P value for evaluating the impact of an intervention.

1. A P value is the probability of obtaining the observed result if the null hypothesis is true (i.e., no difference between groups). It does not tell you how strong the effect is, how important it is, or whether the result could be replicated. For example, just because the result in Case Study 1 is significant, it does not mean that if the study were repeated, the replicated result would also be significant. Figure 1 illustrates the weakness of relying on a P value to determine the success of a study. If Case Study 1 were replicated 26 times with the same means and SDs, the resulting P values would vary substantially (Figure 1).

2. A P value is directly influenced by the number of participants; the more people, the more likely a small effect will be significant (i.e., Case Study 1), and vice versa. Case Study 2 has a larger difference in means compared with Case Study 1, but with fewer people the difference is less likely to be significant.

**Solution**

In addition to reporting P values, researchers should report effect size measures and/or confidence intervals in order to evaluate their results. This additional information will allow researchers and readers to determine the magnitude and importance of an effect and not just whether a result was statistically significant.

**Effect Sizes**

Effect sizes are used to measure the strength of a comparison; the bigger the effect size, the larger the difference in the distribution of scores associated with the groups in question, and therefore the greater the likelihood that the null hypothesis can be rejected. That is, the larger the effect size, the more likely that the groups differ, that the intervention worked, and that the results can be replicated. The following graphs display distributions of scores for two groups with three different effect sizes.

**Confidence Intervals**

Scores obtained in any study always represent a sample of the population of all possible scores, so the population mean is not known. A 95% confidence interval quantifies the precision of the sample; that is, it is a range of values that we are confident contains the population mean. Confidence intervals are useful for comparing groups of scores. The width of the interval reveals the degree that the sample mean contains the population mean, and the overlap between intervals reflects the degree that groups of scores differ (e.g., P value). The less overlap, the stronger the effect.

**How to Use Effect Size**

For Case Study 1, there was a statistical difference, but the standardized effect size (Cohen’s d = 0.30) indicated a small effect and a high overlap in the distribution of scores. It is likely that the two sets of scores came from the same population, and therefore the intervention did not actually work, and the result would not be replicated. For Case Study 2, the result is not significant, but the effect size of 0.61 indicates a moderate to large effect. It is possible that the two sets of scores came from two different populations, and therefore the intervention worked, but did not have enough people to detect a significant difference.

**References**


Author contact: twood@uottawa.ca

OSCE indicates objective structured clinical exam; SD, standard deviation.

First published online October 4, 2016
How to Calculate a Survey Response Rate: Best Practices
Andrew W. Phillips, MD, MEd, clinical fellow of critical care medicine, Stanford University; Benjamin T. Friedman, MD, NREMT-P, resident physician of emergency medicine, University of Washington; and Steven J. Durning, MD, PhD, professor of medicine and pathology, Uniformed Services University of the Health Sciences

- Approximately 50% of original medical education research utilizes surveys, yet many survey studies do not report a response rate, and most do not use a standard response rate definition.1
- In broad terms, the definition of response rate is the number of people who responded divided by the total number of potential respondents, expressed simply as:

\[
\text{Respondents} = \frac{\text{Nonrespondents} + \text{Respondents}}{\text{Nonrespondents} + \text{Respondents}}
\]

- The equation becomes more complicated when we consider which nonrespondents should be included in the calculation based on whether or not they were eligible to participate in the survey.

Recommendations
- Use one of the six American Association of Public Opinion Research (AAPOR) definitions2 outlined in the table below.
- Two decision points determine the definition to use. One decision point has three options, and the other has two, for a total of six possible definitions.
  1. Determine the eligibility of nonrespondents for the survey. Determining eligibility of those who did not respond depends on the characteristics of the potential respondents.
  2. Determine whether or not to include partially completed surveys. Some researchers include surveys with skipped questions in the numerator of the response rate equation. Other researchers count only surveys returned with every question answered.
- Report the definition used when you report the response rate, such as “47 of 100 (47%) potential respondents returned surveys, AAPOR RR6.” See example articles for each definition in Phillips et al.3

Vignette: A researcher is studying expectations of life as a physician among medical students who have at least one parent who is a physician. Nonrespondent eligibility depends on the sampling frame, which corresponds with different response rate definitions, described below.

<table>
<thead>
<tr>
<th>Response rate definition</th>
<th>Nonrespondent eligibility</th>
<th>Which surveys included</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Assume all nonrespondents are eligible</td>
<td>Fully completed</td>
<td>Surveys are distributed to a “physician son/daughter” group. All nonrespondents are included in the calculation based on an assumption that everyone in the group is eligible.</td>
</tr>
<tr>
<td>2</td>
<td>Fully and partially completed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Estimate the probability of nonrespondent eligibility</td>
<td>Fully completed</td>
<td>Surveys are distributed to all medical students. Based on student application data, 40% of students have a physician parent, so only 40% of nonrespondents are included in the calculation based on an estimate that 40% would be eligible.</td>
</tr>
<tr>
<td>4</td>
<td>Fully and partially completed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Use other records to know that all nonrespondents are eligible</td>
<td>Fully completed</td>
<td>Surveys are distributed only to students who indicated on their application that they have a physician parent. All nonrespondents are included in the calculation based on known information that all survey recipients are eligible.</td>
</tr>
<tr>
<td>6</td>
<td>Fully and partially completed</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Decision #1

Decision #2

Disclaimer: The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Uniformed Services University of the Health Sciences, the Department of the Navy, the Department of Defense, or the U.S. Government.

References:

Author contact: warell@stanford.edu; Twitter: @warellb
Presenting Data From Health Professions Education Research in Tables and Figures

Jennifer Campi, senior staff editor, and Elizabeth S. Karlin, MA, staff editor, Academic Medicine, Association of American Medical Colleges

Tables, figures, and other exhibits provide ways of presenting data succinctly and accessibly. In this AM Last Page, we offer six tips for creating exemplary tables and figures.

**Six Tips for Creating Successful Exhibits**

- Follow the target journal’s instructions for tables and figures, usually available on the journal’s Web site.
- Look for and emulate published examples of tables and figures in the target journal.
- Include enough information in each table or figure for readers to understand it independent of the full report. Some readers will skip the text and go first to the tables and figures. Some will read only the exhibits.
- Create online-only appendices for additional data or for data or exhibits that would not fit well within the print pages of the journal.
- Keep it simple! Clear is better than intricate.
- Seek feedback from a colleague (not a coauthor) on your exhibits before submitting the manuscript.

The table and figure below, exemplars published in Academic Medicine, illustrate the effective use of the tips.

<table>
<thead>
<tr>
<th>Table 1 Workforce Production Outcomes: Professional Practice Characteristics of 1999–2001 Graduates From the 124 Existing U.S. MD-Granting Medical Schools by School Expansion Group, as of 2009–2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice characteristic</td>
</tr>
<tr>
<td>Direct patient care</td>
</tr>
<tr>
<td>Primary care</td>
</tr>
<tr>
<td>Rural practice</td>
</tr>
<tr>
<td>Underserved area practice</td>
</tr>
<tr>
<td>Service as full-time faculty*</td>
</tr>
</tbody>
</table>

*Expanded levels by medical school expansion group were as follows: no growth, < 1.0%; low growth, 1.3% to 9.8%; moderate growth, 10.0% to 17.7%; high growth, 18.5% to 138.3%. Less than 3.7% of the data are missing data for all practice outcomes. For individual schools listed by expansion group, see Supplemental Digital Appendix 1 at http://dx.doi.org/10.1097/ACM.0000000000000167.

*Cochran-Armitage test for trend across the four expansion groups.

*Source: 2010 American Medical Association (AMA) Physician Masterfile.

*Source: Geocoded 2009 AMA Physician Masterfile.

*Served as full-time faculty member at any time between graduation and 2010. Source: Association of American Medical Colleges Faculty Roster, accessed September 2012.

Reference:

Author contact: okarlin@acm.org; Twitter: @AcadMedJournal
It’s a Story, Not a Study: Writing an Effective Research Paper
Lorelei Lingard, PhD, professor, Department of Medicine, and Chris Watling, MD, PhD, associate professor, Department of Clinical Neurological Sciences, Schulich School of Medicine & Dentistry, Western University

Advice abounds for education researchers hoping to publish their work. Authors are commonly told to include a clear question and purpose statement, at least one theoretical framework for the work, sufficiently detailed methods, balanced reporting of results, thoughtful limitations, and conclusions appropriate to the research design. Helpful though such advice is, we think it misses the fundamental point. Because what separates a mediocre research paper from a great research paper is not such bits and pieces. It is something much more essential.

A decent research paper reports a study. But a great research paper tells a story.

What’s the difference between study and story?
First, the difference is structural:
- A study lives in the methods and results of a report.
- A story unfolds in the introduction and discussion/conclusion.

Second, the difference is rhetorical:
- The study must be reported accurately.
- The story must be told persuasively.

A good story is understandable, compelling, and memorable. It needs a good study at its core, but it uses that study as a launching point to contribute to a conversation in the world about a shared problem.

Below, we illustrate the standard manuscript format according to this story concept, detailing for each section the key questions writers should ask themselves in order to achieve a good story. While we distinguish between study and story for the sake of clarity, study and story likely interweave throughout a report’s sections.

**Introduction**
What problem did you explore? What’s the hook—why does the problem matter?

**Literature review**
What conversation are you joining? What’s the gap in knowledge?

**Methods**
What did you do? What was the rationale for the research design? Is the explanation accessible?

**Conclusions**
What’s the key lesson from your story? What is the inevitable story-in-waiting?

**Discussion**
How does your story add to the conversation? How have you filled the gap? How does the design limit your contribution?

**Results**
Who are the main characters in your results? Have you illustrated them convincingly?

We do not intend for researchers to see their reports as creative nonfiction. Published condemnations of selective and biased reporting in the clinical trials setting could equally apply to medical education research. Authors must root their stories in science. They should narrate honestly and thoroughly, and they must grapple with results that surprise, deviate, or even disappoint. This scientific storytelling approach will elevate published research, expanding its audience and raising its potential to influence.

References:
3. @WBoledaResearch, Twitter.

Author contact: lorelei.lingard@schulich.wlu.ca; Twitter: @LingardLorelei
The Ethics and Etiquette of Research Dissemination

Holly Meyer, PhD, assistant professor of medicine, Uniformed Services University of the Health Sciences; Lara Varpio, PhD, associate professor of medicine, Uniformed Services University of the Health Sciences; Larry Grupen, PhD, professor of learning health sciences and director, Master of Health Professions Education program, University of Michigan; and Gurjit Sandhu, PhD, assistant professor of surgery, University of Michigan

Unethical behavior (e.g., fictionalizing data) is relatively uncommon in health professions education (HPE); nevertheless, recent highly publicized stories1 should prompt HPE researchers to reflect on our community’s ethical standards. Disseminating research broadly and promptly has led some researchers to make unethical choices. The ethical dilemmas HPE researchers face often stem from the subtle nuances inherent in interpreting ethical codes of conduct related to dissemination (i.e., how scholarly work is communicated to the broader community) of research. We explore these ethical and etiquette considerations by addressing two dissemination situations that HPE researchers commonly encounter.

<table>
<thead>
<tr>
<th>Research Dilemmas</th>
<th>Ethical Considerations: Recommending and defending right and wrong conduct</th>
<th>Etiquette Considerations: Expectations for accepted behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td>How can I justify multiple publications and/or presentations of the same body of work? What is the difference between “building on previous work” vs. “recycling”?</td>
<td>• Creating publications with a family resemblance involves recasting data presentations to address different outcomes of the same study or alternative research questions underlyng the study.1</td>
<td>• Contact the journal’s editorial staff or the conference’s program staff to discuss your study’s particular circumstances since journals/conferences may differ in how they interpret the rules. Their staff members’ insights can help you determine if your findings are inappropriately repetitious.</td>
</tr>
<tr>
<td>How do I know when I’ve pushed too much data into a paper (i.e., it is too dense) vs. when I have divided the data across too many papers (i.e., salami slicing)?</td>
<td>• Aim to present significant findings and rich data and to offer novel and noteworthy insights.3</td>
<td>• “Publish or perish” has led to discussions about the least publishable unit.5 Researchers have to determine how many unique and insightful significant findings their data can offer to the research literature. That is the number of papers to write.</td>
</tr>
</tbody>
</table>

Recommended Guidelines

1) Duplicate publications are unethical.
2) Duplicate presentations fall in a gray zone.
3) When in doubt, ask the editor, or just don’t do it.
4) Prioritize quality of publications over quantity.

For a list of institutional Web sites or publications relating to ethical conduct, please see Supplemental Digital Appendix 1 at http://links.lww.com/ACADMED/A396.

Acknowledgement: The authors would like to thank Paul Trembliey for his contribution to graphic design.

Disclaimer: The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Uniformed Services University of the Health Sciences, the Department of the Navy, the Department of Defense, or the U.S. Government.

References:

Author contact: holly.meyer.ctr@usuhs.edu
Ten Tips to Move From “Revisions Needed” to Resubmission

Holly S. Meyer, PhD, assistant professor, Department of Medicine, Uniformed Services University of the Health Sciences; Jan Carlile, PhD, professor, Biomedical Informatics and Medical Education, Department of Family Medicine and Pharmacy, University of Washington; and Steven J. Durning, MD, PhD, professor, Medicine and Pathology, Department of Medicine, Uniformed Services University of the Health Sciences

Authors may struggle when they receive a decision letter that requires revisions to their manuscript. Reframing the submission and review process to align with climbing a mountain helps authors visualize the process as challenging, but rewarding. There are three phases of the climb:

- Write: writing your paper for the initial submission,
- Prepare: preparing yourself for feedback, and
- Revise: revising your paper based on feedback received.

The 10 tips below relate to the “prepare” and “revise” portion of the climb. These tips frame the revision process as an opportunity to transform and improve one’s manuscript.

**PREPARE**

1. Expect revisions! Even excellent manuscripts require (and benefit from) revisions.
2. Engage the peer-review process by having colleagues review the revised manuscript prior to submission.

**REVISE**

3. Ask for clarification if you don’t understand a suggestion.
4. Address every comment in the revision unless the editor instructs otherwise.
5. Do your best to respond; if appropriate, consider stating the concern as a limitation.
6. Throughout, follow the advice of resources in the literature.1
7. Be professional and tactful in your reply.
8. Be timely in responding; early is better than late.
9. Make changes easy to follow:
   - Edit using the track-changes function.
   - Provide a table organized by requested changes, revisions (location), and explanations.
   - Reread the manuscript to ensure the whole text holds together post revisions.
10. Be persistent! Almost every paper has a home. A rejection is an opportunity for resubmission to another journal.

Disclaimer: The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Uniformed Services University of the Health Sciences, the Department of the Navy, the Department of Defense, or the U.S. Government.

Reference:

Author contact: holly.meyer.cbi@usuhs.edu
Congratulations! Your Article Has Been Accepted. Now What?
Media, Social Media, and Other Outlets for Promoting Your Work

Toni Gallo, MA, senior staff editor, Academic Medicine, Association of American Medical Colleges

First, know the journal’s embargo policy.
- All journals have one. It dictates when you can share information publicly about your article.
- Most begin when you initially submit your article and lift once your article is published.
- While your article is under embargo, you can still share it privately—with colleagues and your institution’s communications office (CO). You can also discuss it with trusted journalists, as long as no news stories are released until the embargo lifts.

Next, consider what you want to say and who you want to reach.
- This will dictate which outlets are most appropriate for sharing your work.
- It will also help you think about what you can do, what your institution can do, and what the journal can do. Each has unique expertise, experience, and reach.

<table>
<thead>
<tr>
<th>Promoter</th>
<th>Role</th>
<th>Reach</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>You</strong></td>
<td>- Provide a first-person account of your work</td>
<td>- Personal and professional networks</td>
</tr>
<tr>
<td></td>
<td>- Explain the applications and implications of your work</td>
<td></td>
</tr>
<tr>
<td><strong>Your institution</strong></td>
<td>- Provide guidance on which outlets (see below) are best for sharing your work</td>
<td>- Institutional community</td>
</tr>
<tr>
<td></td>
<td>- Offer expertise on shaping the message for each audience</td>
<td>- Local and national media</td>
</tr>
<tr>
<td><strong>The journal</strong></td>
<td>- Give weight to your work because of its reputation</td>
<td>- Journal readers and society members</td>
</tr>
<tr>
<td></td>
<td>- Provide perspective on how your work fits into current scholarship</td>
<td></td>
</tr>
</tbody>
</table>

Finally, consider these communication strategies for crafting your message and reaching the right audiences.

<table>
<thead>
<tr>
<th>Outlet</th>
<th>Communication strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Media</strong></td>
<td>- Contact the CO at your institution as soon as your article is accepted. Explain in clear, non-scientific language your work, its importance, and any critical next steps.</td>
</tr>
<tr>
<td></td>
<td>- The CO personnel can help you determine the appropriate outlets for your work (e.g., local/national media, social media, news services, health Web sites, trade publications) and develop a plan to reach those outlets.</td>
</tr>
<tr>
<td></td>
<td>- They also can provide you with media training to ensure you are comfortable discussing your work and to help you shape your message for specific audiences.</td>
</tr>
<tr>
<td></td>
<td>- Be available before and after your article is published for interviews.</td>
</tr>
<tr>
<td></td>
<td>- Consider visuals to accompany news stories (e.g., videos, graphics, B-roll footage).</td>
</tr>
<tr>
<td></td>
<td>- Share any specialty-specific or foreign outlets or contacts you have with the CO personnel—they do not know your field as well as you do and might not know about these channels.</td>
</tr>
<tr>
<td></td>
<td>- Connect the CO with the journal’s editorial staff so they can coordinate efforts.</td>
</tr>
<tr>
<td><strong>Social media</strong></td>
<td>- Ask the journal and your institution to tag your handle in tweets they send.</td>
</tr>
<tr>
<td></td>
<td>- Tag the journal’s handle and those of your coauthors, institution, relevant funders, or other organizations involved in your work in tweets you send, as the character count allows.</td>
</tr>
<tr>
<td></td>
<td>- Use relevant hashtags to reach a broader audience (see Symplur Healthcare Hashtags).</td>
</tr>
<tr>
<td></td>
<td>- Include a link to your article on the journal’s Web site. Use a shortened URL (from bit.ly or owl.ly) to reduce the character count.</td>
</tr>
<tr>
<td></td>
<td>- Include a visual when appropriate (e.g., photo you took or graphic from your article—it does not have to be professional).</td>
</tr>
<tr>
<td><strong>Facebook</strong></td>
<td>- Post a one- to two-sentence description of your article (Facebook will truncate your post after 400 characters).</td>
</tr>
<tr>
<td></td>
<td>- Include a link to your article and a visual when appropriate.</td>
</tr>
<tr>
<td><strong>LinkedIn</strong></td>
<td>- Post a one- to two-sentence description of your article both to your personal page so your network will see your work and to any groups whose work is relevant.</td>
</tr>
<tr>
<td></td>
<td>- Include a link to your article and a visual when appropriate.</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td>- Share your work in medical school/hospital/university/alumni newsletters and Web sites and in appropriate trade publications. The CO can help you facilitate this process.</td>
</tr>
<tr>
<td></td>
<td>- Share your work on appropriate blogs. Ask the journal to share your work on its blog, if available.</td>
</tr>
<tr>
<td></td>
<td>- Post a description of your work on your personal blog or Web site.</td>
</tr>
<tr>
<td></td>
<td>- Share your article on scholarly networking Web sites like Medscape or ResearchGate.</td>
</tr>
<tr>
<td></td>
<td>- Whenever possible, include a link to your article.</td>
</tr>
</tbody>
</table>

Remember, publishing your article is not the end. Getting the word out can help ensure that your work has an impact.

Acknowledgments: The author thanks: Lisa Worley, University of Miami Miller School of Medicine; Maria Ober, Boston University School of Medicine; Jacqueline Kozlowski, Thomas Jefferson University and Jefferson Health; Jennifer Aron, Barnes-Jewish Hospital; and Stephanie Weiner, Association of American Medical Colleges, for sharing their expertise on this topic.

Further reading:

Author contact: tgallo@aacu.org; Twitter @AcadMedJournal

First published online September 27, 2016
A Medical Educator’s Guide to #MedEd

Andrew Micieli, MM, medical student, Jason R. Frank, MD, professor, Department of Emergency Medicine, and Alineza Jalalí, MD, teaching chair, Faculty of Medicine, University of Ottawa

• Twitter is a growing social media platform in medical education. This forum is facilitated by the use of the hashtag (#).
• The hashtag/word combo #MedEd is used for tweets related to medical education and is the gold standard for medical education news on Twitter.

Why consider Twitter for medical education?
The purpose of this Last Page is to educate novice, but frequent “tweeters” (i.e., Twitter users) on the possible uses of Twitter in medical education by ...
• Highlighting the usefulness of #MedEd as an educational tool in independent learning and teaching
• Explaining how to search and customize your Twitter feed for high-quality medical education news
• Considering the future of #MedEd among Twitter users

#MedEd is an educational tool for medical educators and medical students. It’s a forum for...

- Links to news or journal articles
- Medical educators discussing teaching strategies
- Trending news about postgraduate training programs
- Real-time content from health care conferences
- Study tools (students tweet questions to each other)
- Targeted campaigns from special medical interest groups

Permission to reprint MedEdChat tweet granted by Ryan Madanick, MD. Permission to use AMA logo and tweet granted by American Medical Association.

Using, Personalizing, and Optimizing #MedEd

To the right is a flow chart highlighting the different options medical educators have for searching and customizing their Twitter page for #MedEd. The pathway individual educators choose more often will depend on how frequently they use Twitter, their career and preferences, their reasons for using Twitter, and their social media influence (i.e., their number of followers).

Future of #MedEd

There has been a dramatic increase over time in both #MedEd tweets (see figure to the left) and in Twitter users using this hashtag. This increase may be leveraged for personalized teaching and learning in the classroom setting. For example, professors have successfully displayed live stream Twitter chats so students can post questions for concurrent viewing and discussion during lectures.

References

Author contacts (on Twitter):
Micieli: @medstudent_blog
Frank: @jrfrank
Jalalí: @ARJalalí

Author contact (via e-mail):
Micieli: andrew.micieli@ottawa.ca