Writing for publication in medical education

Mitra Amini

Professor of Shiraz University of Medical Sciences, Shiraz, Iran

FAIMER fellow, Assessor of WFME, Associate Editor of BMC Medical Education Journal

Medical education publication

- Sharing experience, innovation and lessons learnt with other professionals and educators.
- Stimulating scholarly debate and suggestions for future development that could prompt new research and further studies.
- Improving the standards of students' education and learning and hence the standards of career development and the quality of patients' healthcare.

Medical education publication

- Contributing to the advancement of the profession and enhancing the body of theory and understanding.
- Contributing to the improvement of research quality in the university where researchers work, and improving academic standing and their opportunities for research funding and promotion.

Medical education publication

- Gaining recognition and prestige for their research/publication contributions.
- Receiving feedback from the peer reviewers and editors that can help in improving the work as well as the publication.
- Establishing a network of academics working on similar projects and sharing the same area of research interests.

Challenges

- Inadequate knowledge and skills in writing for publication
- Lack of time and commitment to write, and a lack of confidence and motivation to start writing
- Rejection of manuscript

Some resources for writing

- A series of articles by the Journal of Clinical Epidemiology on successful academic writing
- Textbooks on writing research papers
- online resources; including the Purdue University Online Writing Lab, OWL (https://owl.english.purdue.edu), and the Online Research Skills Module by University Graduate College, Cardiff University (http://cardiff.ac.uk /ugc/training/online-research-skills-modules) and BMJ learning (http://learning.bmj.com)

Tips for writing

- Writing for success is a systematic, disciplined process.
- The research question should be focused, imbedded in the available literature, and achievable given the available resources.
- The research design is determined by the question, should conform to ethical educational standards, and should be comprehensively described.

- Strategies for writing include starting where it is easiest to do so, spontaneously and uncritically writing the first paragraphs, and identifying and reducing specific barriers to writing.
- Getting the final submission ready requires very careful attention to detail and accuracy.

Choosing a topic and getting started

- Choose a topic with a question that is doable.
- Choose a topic area for which you have enthusiasm.
- Identify the importance or significance of the topic.
- Imbed the topic and question in the related literature.
- Look for mentorship and constructive criticism on the research idea.
- Choose capable, enthusiastic, and compatible team members.

Importance of the question

- Is the question to be answered relevant to many people or, perhaps, relevant to fewer people but very influential or problematic?
- Reviewers and editors will view this consideration as crucial in the judgment of the suitability of a manuscript for publication

- Significance pertains to the prevalence and/or seriousness of an issue and the likelihood that the results will benefit educators and their learners.
- Significance is also understood by how the study's results might add to the available literature.

Sharpen your good idea

- Examine the related literature and what was published about your idea – what do we know and what do we need to know?
- Assess your idea again and examine if it adds anything new to our knowledge in this area.
- Sharpen your research question, evaluate what you are trying to prove.

Think about the methodology

- Quantitative
- Qualitative
- Mixed method

Authorship

- Discuss authorship and who you think could be a potential co-author and bring something useful to the project

Ethical approval

- Work out ethical approval required for your project.

Think about journal

- Which journal?

Preparing the manuscript

Mitra Amini

Professor of Shiraz University of Medical Sciences, Shiraz, Iran

FAIMER fellow, Assessor of WFME, Associate Editor of BMC Medical Education Journal

- Get started, even with the first paragraph alone
- Start where it is easiest to do so
- Follow a relatively set script or structure applicable to the anatomy of the
- section being written
- Spontaneously and uncritically write the first draft
- Find time to write, relatively free of distractions
- Create a reward system based on the amount written
- Use the team to help in overcoming specific barriers to writing
- Be patient, persevere, and have fun

Original Article Structure

- **Introduction**
  - What question was asked?
- **Materials and Methods**
  - How was it studied?
- **Results**
  - What was found?

and

- **Discussion**
  - What do the findings mean?

Introduction

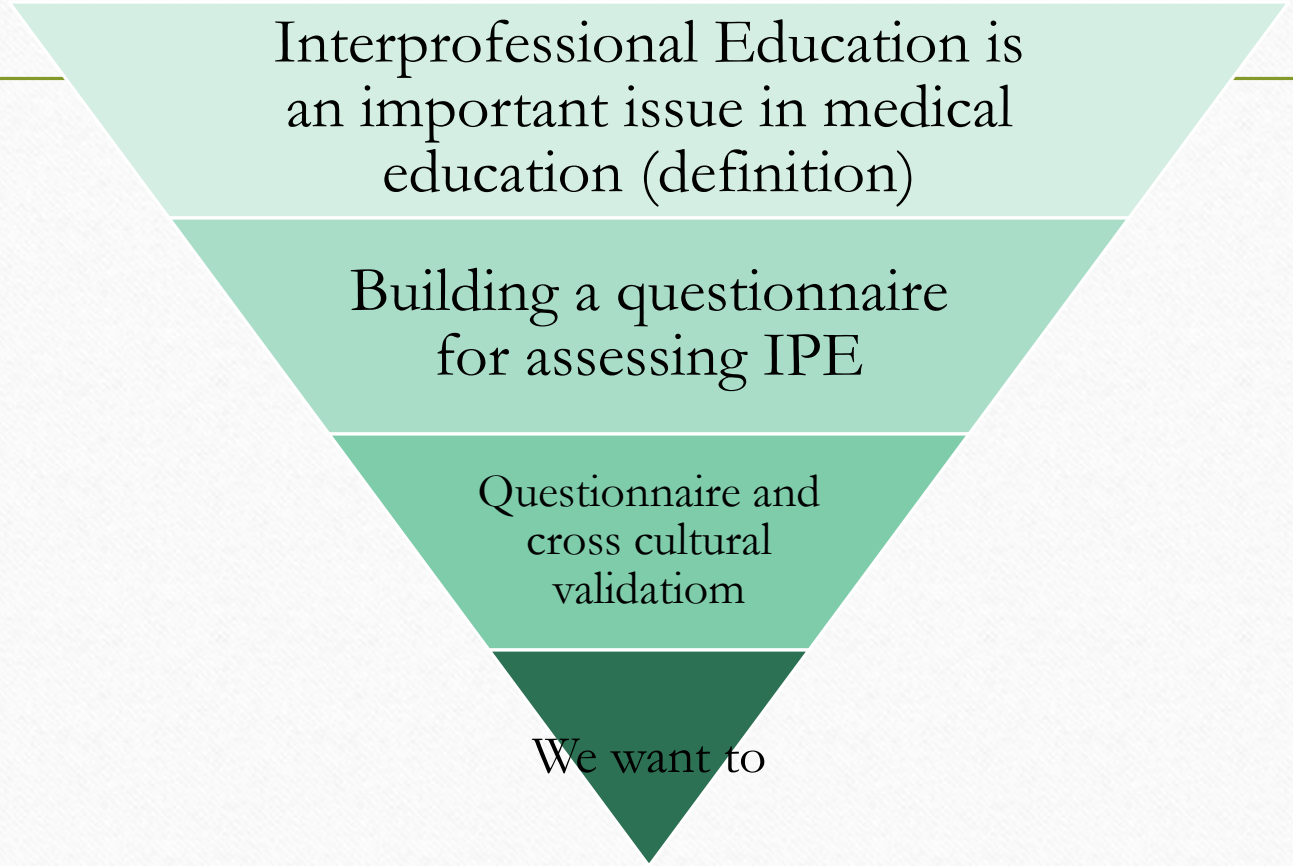
Three parts

Part one : The definition

Part two : Previous studies

Part three: The importance of the study and the research question

Inverted Pyramid



The last sentence

- Along with this growth in IPE, however, there is an increasing concern over the lack of a valid and reliable Persian scale to measure the readiness of students for IPE. Therefore, we felt that it is essential to construct a Persian scale, which fits into a Persian culture. The purpose of this research is to determine whether or not the RIPLS can be adapted for Persian healthcare professions.

Educational paper introduction

- It should also be appreciated that some educational research papers will require a theoretical or conceptual framework in the Introduction. In this case the Introduction might be longer than was indicated above. In this way, papers for educational journals differ from those for biomedical journals, when the latter tend to leave theoretical issues to the “Discussion” section.

Material and Methods

Methods: The most important part

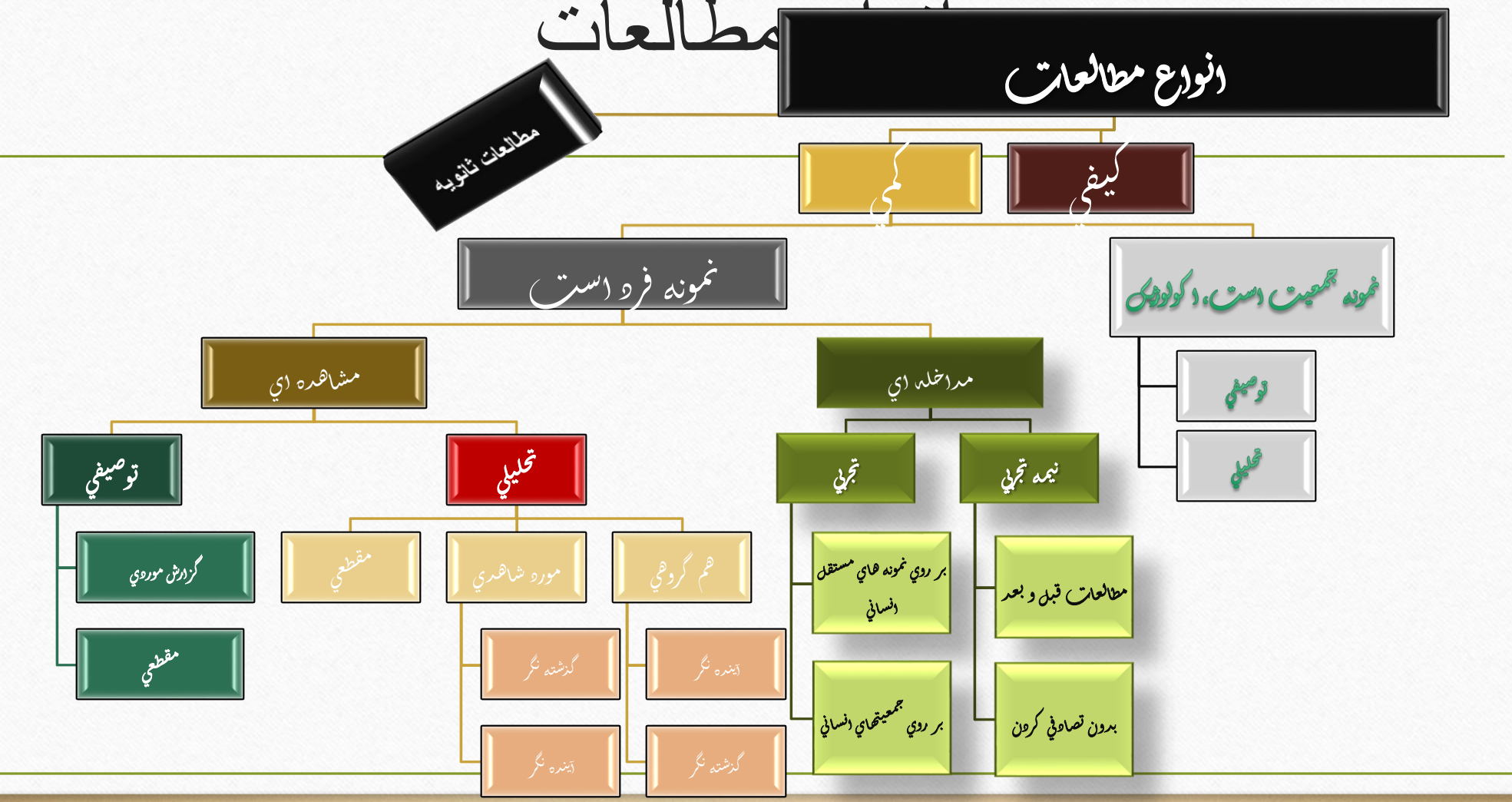
- The nature of the educational research question determines the choice of methods to be employed in the planned and disciplined approach to securing its answer and to delineating the parameters of the study

Qualitative and quantitative research

Qualitative: hypothesis generation

Quantitative: hypothesis testing

مطالعات


انواع مطالعات
کیفی
کمی
نمونه جمعیت است، اکولوژیک
توصیفی
تحلیلی
نمونه فرد است
مداخله ای
نیمه تجربی
تجربی
بدون تصادفی کردن
مطالعات قبل و بعد
بر روی جمعیتهای انسانی
بر روی نمونه های مستقل انسانی
مشاهده ای
تحلیلی
توصیفی
هم گروهی
مورد شاهدی
مقطعی
آینده نگر
گذشته نگر
آینده نگر
گذشته نگر
گزارش موردی
مقطعی
مطالعات ثانویه

Methods

- Because medical education practice is so variable across countries and schools, it might be helpful to include a specific subsection of the Methods describing the context of the study.

- Study design and participants
- Data collection method
- Statistical analysis
- Ethical consideration

Materials and Methods

- The most important section

- Materials
  - Who, when, where, what (instrument, …)
- Methods
  - How
  - (Chrono)logical order
- Enough detail to be reproducible
- In past tense

Materials and Methods (cont.)

- Study design
- Inclusion/exclusion criteria
- Sample size
- Study population
- Sampling technique
- Data collection
  - Questionnaire
  - Data collection sheet
- Ethics
  - Informed consent
  - Institutional Review Board approval (IRB approval
  - Conflicts of Interest (COI)
- Statistical analysis

ICMJE INTERNATIONAL COMMITTEE of MEDICAL JOURNAL EDITORS

Enter search terms SEARCH

Recommendations | Conflicts of Interest | Journals Following the ICMJE Recommendations | About ICMJE | News & Editorials

Recommendations

Read the Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly work in Medical Journals.

BROWSE DOWNLOAD

Conflicts of Interest

Use the ICMJE Form for Disclosure of Potential Conflicts of Interest to generate a disclosure statement for your manuscript.

ACCESS THE FORM

ANNOUNCEMENTS

The application period for new member journals ended July 1. The ICMJE does on occasion solicit new applications. If you would like to receive an alert regarding future application periods and other announcements from ICMJE, sign up here.

The deadline for submitting feedback regarding the proposed requirements for Sharing Clinical Trial Data has ended. You may view the feedback submitted here. - *April 2016*

Member Publications & Organizations

equator network

Enhancing the QUAlity and Transparency Of health Research

EQUATOR resources in Portuguese | Spanish

Home | Library | Toolkits | Courses & events | News | Blog | Librarian Network | About us | Contact

Your one-stop-shop for writing and publishing high-impact health research

find reporting guidelines | improve your writing | join our courses | run your own training course | enhance your peer review | implement guidelines

Library for health research reporting

The Library contains a comprehensive searchable database of reporting guidelines and also links to other resources relevant to research reporting.

- Search for reporting guidelines
- Not sure which reporting guideline to use?
- Reporting guidelines under development
- Visit the library for more resources

Reporting guidelines for main study types

Randomised trials	CONSORT	Extensions	Other
Observational studies	STROBE	Extensions	Other
Systematic reviews	PRISMA	Extensions	Other
Case reports	CARE	Extensions	Other
Qualitative research	SRQR	COREQ	Other
Diagnostic / prognostic studies	STARD	TRIPOD	Other
Quality improvement studies	SQUIRE		Other
Economic evaluations	CHEERS		Other
Animal pre-clinical studies	ARRIVE		Other
Study protocols	SPIRIT	PRISMA-P	Other

See all 343 reporting guidelines

Funders: reporting guidelines key for research reproducibility and reliability

Reporting Guidel

- CONSORT
  - Clinical trials
- STROBE
  - Observational studies
- PRISMA
  - Systematic reviews and meta-analyses

www.equator-network.org/resource-centre/library-of-health-research-reporting

workgroup

- Please write introduction and methods for your topic
- Please check the checklists for writing methods

Preparing the manuscript
Part 2

Mitra Amini

Professor of Shiraz University of Medical Sciences, Shiraz, Iran

FAIMER fellow, Assessor of WFME, Associate Editor of BMC Medical Education Journal

Results

- The Results should not be a mere "laundry list" of data and various statistical comparisons. In approaching the development of the Results, one helpful method is to order the findings in parallel with how the goals were identified in the Introduction and the findings discussed in the "Discussion" section

Tips for effective tables

- Consider whether the table is really necessary
- The “information density” (information per square inch) should be greater than just putting this information in the main text; data that would require fewer than 2 columns and rows should be presented in the text rather than a table.

Tips for effective tables

- The information should complement rather than duplicate information available elsewhere in the manuscript.
- Create a brief but explanatory title.
- Carefully select the data (not all data are equally important) and thoughtfully organize the data to communicate a clear message.

Tips for effective tables

- Ensure that the table can be easily interpreted without reference to the main text.
- Keep the table simple, clean, and free of extraneous detail.
- Explain all abbreviations; special use of italics, parentheses, and dashes; special symbols; and empty cells * Example: a cell entry “46/50 (92%)” should have a column heading or footnote explaining that this means “No./N (%)"

Tips for effective tables

- Keep abbreviations consistent with the main text; define all abbreviations using footnotes (so that the table can stand alone).
- Apply similar formatting for all tables in the manuscript.
- Follow all journal-specific instructions on table creation.
- Look at recent back issues of the journal for examples.

Tips for effective tables

- If the table or its data are from another source, cite the original source.
- Refer to the table in the text.
- Place the table in the manuscript according to journal instructions (i.e., appended at the end, embedded in the main text, or submitted in a separate document).
- For additional information on table preparation, see the Purdue Online Writing Lab (owl.english.purdue.edu) and Wainer (1984), Morgan (1985), and Schriger et al. (2006).

Tips for effective figures

- Consider whether the figure is really necessary.
- The “information density” (information per square inch) should be greater than just putting this information in the main text or in a table.
- Carefully select the data (not all data are equally important) and thoughtfully organize the data to communicate a clear message.

Tips for effective figures

- Follow established guidelines and norms for specific figure types (e.g. participant flow diagram for experimental studies (Schulz et al., 2010), or study flow diagram for systematic reviews (Moher et al., 2009))
- Ensure that the figure can be easily interpreted without reference to the main text.
- Ensure that the visual metaphor of the figure accurately reflects both the data and the intended message.

- Ensure that that all scales (e.g., x and y axes) are consistently used and correctly proportioned; disproportionate scaling (e.g., scales that vary irregularly along the axis, scales that do not start at 0, and nonlinear scales) can be misleading.
- Keep the figure simple, clean, and free of extraneous detail; avoid using special effects (e.g. 3-D effects, shading, and layered text).

- Verify that all data are accurate and are plotted accurately.
- Explain all line, symbol, and color styles; text emphasis (bold, italics); and abbreviations in the legend or caption.
- Keep abbreviations consistent with the main text; define all abbreviations.

- Follow all journal-specific instructions on figure creation, including figure resolution and file format (e.g., JPEG, TIFF, PNG).
- Place the figure in the manuscript according to journal instructions (i.e., appended at the end, embedded in the main text, or submitted as a separate file).
- For additional information on figure preparation and the visual display of data, see the Purdue Online Writing Lab (owl.english.purdue.edu) and Tufte (2001), Wainer (1984), and Schriger and Cooper (2001).

Discussion

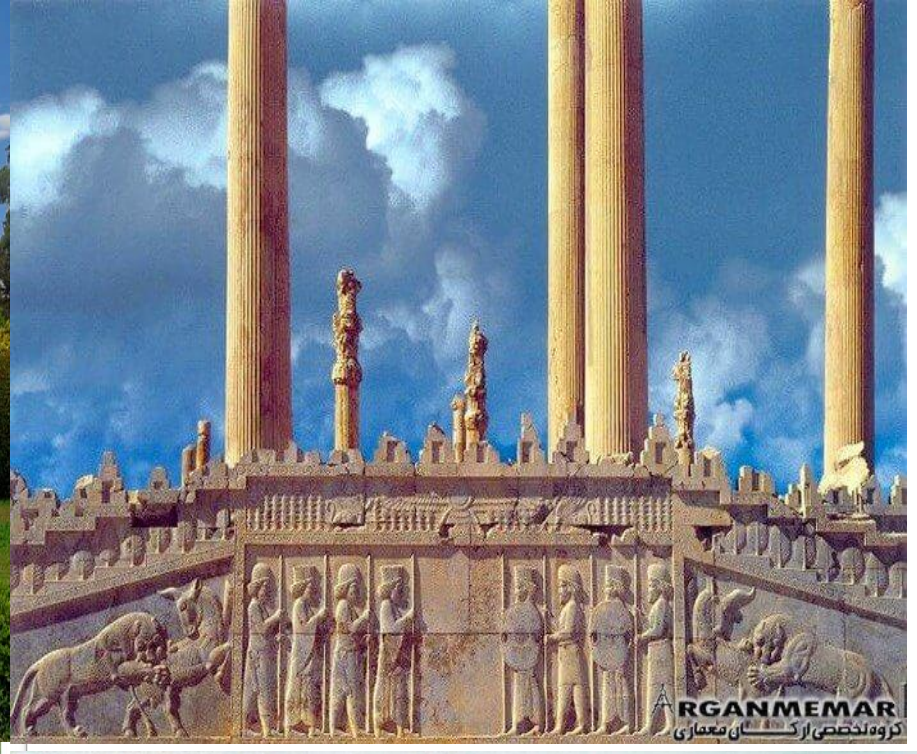
- One of the key goals of the discussion is to link the aims and findings with relevant prior research. In this way, the discussion links back to the introduction to inform the reader about how these new findings are placed into an appropriate context

- In quantitative research, findings should be clearly stated and understood in relation to the rationale for the study and previously published findings of interest, possible alternative explanations, and implications for readers in their roles as educators, educational researchers, or administrators.

- The findings also should be discussed in relation to the strengths and limitations of the data (e.g. a one-site study, small number of subjects, low response rate, and other contextual factors can limit the generalizability of the findings), which is usually the work of the penultimate paragraph of the "Discussion" section.

- It might also be noted that statistically significant differences are not necessarily educationally meaningful. The final paragraph of the Discussion briefly reiterates the main findings and their implications for readers

- Now write results and discussion for your topic


RGANMEMAR
گروه تخصصی ارگان معماری

Developing A questionnaire

Mitra Amini

MD MPH Ms

Surveys are used throughout medical education.

- Student evaluation of medical school courses and clerkships
- Patient satisfaction and student self-assessment surveys.

Poor surveys!

- When surveys are poorly designed, they may fail to capture the essence of what the survey developer is attempting to measure due to different types of measurement error.
- Poor questions
- Wording
- Confusing question layout
- Inadequate response options

Survey

- Questions used in a phone interview,
- The set of items employed in a focus group
- The questions on a self administered patient survey

Self administered survey named as questionnaire

- Questionnaires are good for gathering data about abstract ideas or concepts that are otherwise difficult to quantify, such as opinions, attitudes and beliefs.

Construct

- A construct is the model, idea or theory that the researcher is attempting to assess.
- Student satisfaction with a new curriculum
- Patients' ratings of their physical discomfort
- Love
- People perspectives
- Motivation

Seven steps survey

- Conduct a literature review
- Conduct interviews and/or focus groups
- Synthesize the literature review and interviews/focus groups
- Develop items
- Conduct expert validation
- Conduct cognitive interviews
- Conduct pilot testing

Conduct a literature review

- (1) To clearly define the construct
- (2) To determine if measures of the construct (or related constructs) already exist. Cross-Cultural validation

Conduct interviews and/or focus groups

- Semi structured: Conduct interviews and/or focus groups interview
- Focus group

Synthesize the literature review and interviews/focus groups

- One suitable way to conduct Step 3 is to develop a comprehensive list of indicators for the construct by merging the results of the literature review and interviews/focus groups

Develop items

- The goal of this step is to write survey items that adequately represent the construct of interest in a language that respondents can easily understand.
- One important design consideration is the number of items needed to adequately assess the construct.
- There is no easy answer to this question.

- In general, it is good practice to develop more items than will ultimately be needed in the final scale (e.g. developing 15 potential items in the hopes of ultimately creating an eight-item scale), because some items will likely be deleted or revised later. In general, it is good practice to develop more items than will ultimately be needed in the final scale (e.g. developing 15 potential items in the hopes of ultimately creating an eight-item scale), because some items will likely be deleted or revised later in the design process

- Ultimately, deciding on the number of items is a matter of professional judgment, but for most narrowly defined constructs, scales containing from 6 to 10 items will usually suffice in reliably capturing the essence of the phenomenon in question.

- The next challenge is to write a set of clear, unambiguous items using the vocabulary of the target population.
- Another important part of the questionnaire design process is selecting the response options that will be used for each item.
- Closed-ended survey items can have (nominal) response options or (ordinal) response options.

- Moreover, survey items can ask respondents to complete a ranking task (e.g. “rank the following items, where 1 worst and 6 best”) or a rating task that asks them to select an answer on a Likert-type response scale.

Pitfall	Solution(s)	References
1. Labeling only the end points of your response options 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. Problematic item: How interesting did you find this clinical reasoning course? not at all interesting ○ ○ ○ ○ extremely interesting	**Verbally label each response option.** 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. Improved item: How interesting did you find this clinical reasoning course? ○ not at all interesting ○ slightly interesting ○ moderately interesting ○ quite interesting ○ extremely interesting	Krosnick, 1999
2. Labeling response options with both numbers and verbal labels Because of the additional information respondents must process, including numbers and verbal labels extends response time. 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." Problematic item: How much did you learn in today's workshop? ○ -2 almost nothing ○ -1 a little bit ○ 0 some ○ 1 quite a bit ○ 2 a great amount	**Use only verbal labels** 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. Improved item: How much did you learn in today's workshop? ○ almost nothing ○ a little bit ○ some ○ quite a bit ○ a great amount	Christian et al., 2009; Krosnick, 1999
3. Unequally spacing your 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. Problematic item: How much did you learn from your peers in this course? ○ almost nothing ○ a little bit ○ some ○ quite a bit ○ a great amount	**Maintain equal spacing between response options.** 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. Improved item: How much did you learn from your peers in this course? ○ almost nothing ○ a little bit ○ some ○ quite a bit ○ a great amount	Dillman et al., 2009
4. Placing non-substantive response options together with substantive response options Placing non-substantive 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. Problematic item: How satisfied are you with the quality of the library services? ○ not at all satisfied ○ slightly satisfied ○ moderately satisfied ○ quite satisfied ○ extremely satisfied ○ not applicable	**Use additional space to visually separate non-substantive response options from the substantive options.** Using additional space to visually separate non-substantive response options from substantive options will align the visual midpoint with the conceptual midpoint thereby reducing measurement error. This recommendation is a beneficial exception to the guidance above about maintaining equal spacing between response options. Improved item: How satisfied are you with the quality of the library services? ○ not at all satisfied ○ slightly satisfied ○ moderately satisfied ○ quite satisfied ○ extremely satisfied ○ not applicable	Dillman et al., 2009

Adapted with permission from Lippincott Williams and Wilkins/Wolters Kluwer Health: Artino AR & Gehlbach H (2012). AM last page: Avoiding four visual-design pitfalls in survey development. *Academic Medicine*, 87: 1452.

Figure 1 Visual-design "best practices" based on scientific evidence from questionnaire design research.

Pitfalls

- Creating a double barreled item:
- 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, the respondent may talk to his nurses often but talk to administrative staff much less frequently. If this were the case, the respondent would have a difficult time answering the question. Survey items should address one idea at a time

Creating a negatively worded item

- 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 “yes” in order to mean “no” (or “agree” in order to “disagree”), the item is flawed

- In an average week, how many times do you start class on time?
- Should the chief resident be responsible for admitting patients?

Table 3. Examples of various Likert-type response options.

Construct being assessed	Five-point, unipolar response scales	Seven-point, bipolar response scales
Confidence	• Not at all confident • Slightly confident • Moderately confident • Quite confident • Extremely confident	• Completely unconfident • Moderately unconfident • Slightly unconfident • Neither confident nor unconfident (or neutral) • Slightly confident • Moderately confident • Completely confident
Interest	• Not at all interested • Slightly interested • Moderately interested • Quite interested • Extremely interested	• Very uninterested • Moderately uninterested • Slightly uninterested • Neither interested nor uninterested (or neutral) • Slightly interested • Moderately interested • Very interested
Effort	• Almost no effort • A little bit of effort • Some effort • Quite a bit of effort • A great deal of effort	
Importance	• Not important • Slightly important • Moderately important • Quite important • Essential	
Satisfaction	• Not at all satisfied • Slightly satisfied • Moderately satisfied • Quite satisfied • Extremely satisfied	• Completely dissatisfied • Moderately dissatisfied • Slightly dissatisfied • Neither satisfied nor dissatisfied (or neutral) • Slightly satisfied • Moderately satisfied • Completely satisfied
Frequency	• Almost never • Once in a while • Sometimes • Often • Almost always	

Conduct expert validation

- This step involves collecting data from content experts to establish that individual survey items are relevant to the construct being measured and that key items or indicators have not been omitted

- Using experts to systematically review the survey's content can substantially improve the overall quality and representativeness of the scale items

- One useful approach to finding experts is to identify authors from the reference lists of the articles reviewed during the literature search.

- There is no consensus in the literature regarding the number of experts that should be used for content validation; however, many of the quantitative techniques used to analyze expert input will be impacted by the number of experts employed. Rubio et al. (2003) recommends using 6–10 experts, while acknowledging that more experts (up to 20) may generate a clearer consensus about the construct being assessed, as well as the quality and relevance of the proposed scale items

key domains to assess through an expert validation process are

- Representativeness
- Clarity
- Relevance
- Distribution.

- Representativeness is defined as how completely the items (as a whole) encompass the construct.
- Clarity is how clearly the items are worded and relevance refers to the extent each item actually relates to specific aspects of the construct.

- The distribution of an item is not always measured during expert validation as it refers to the more subtle aspect of how “difficult” it would be for a respondent to select a high score on a particular item.

- In other words, an average medical student may find it very difficult to endorse the self-confidence item, “How confident are you that you can get a 100% on your anatomy exam”, but that same student may find it easier to strongly endorse the item, “How confident are you that you can pass the anatomy exam”. In general, survey developers should attempt to have a range of items of varying difficulty

روایی(validity)

- روایی، به هدفی که ابزار برای تحقق بخشیدن به آن درست شده است اشاره می کند. به عبارتی دیگر، ابزاری دارای روایی است که برای اندازه گیری آنچه مورد نظر است مناسب باشد. روایی ابزار، عبارت است از میزان کارایی آن برای اندازه گیری خصیصه ای که به منظور اندازه گیری آن خصیصه ساخته شده است.

انواع روایی

- روایی صوری
- روایی محتوا
- روایی ملاک
- روایی سازه

که در اینجا فقط به دو مورد اول می پردازیم.

- برای ارزیابی روایی محتوایی از نظر متخصصان در مورد میزان محتوای ابزار اندازه گیری و هدف پژوهش، استفاده میشود. برای این منظور دو روش **کیفی و کمی** در نظر گرفته میشود.
- در بررسی **کیفی محتوا** پژوهشگر از متخصصان درخواست می کند تا بازخورد لازم را در ارتباط با ابزار ارائه دهند که بر اساس آن موارد اصلاح خواهند شد.
- در بررسی **کمی محتوا** از دو ضریب نسبی روایی محتوا **(CVR)**و شاخص روایی محتوا **(CVI)** استفاده میشود.

روایی صوری

- روایی صوری یا ظاهری یک شاخص ابتدائی و حداقل برای روایی محتوا به شمار می آید.
- این نوع روایی نشان می دهد که عناصر مورد سنجش به طور ظاهری توانایی اندازه گیری مفهوم پژوهش را دارند.
- به عبارتی روایی صوری بیانگر این است که سوال ها یا گویه های آزمون تا چه حد شبیه به موضوعی است که برای اندازه گیری آن منطقی که در واقع با قضاوت متخصصان و یا خبرگان مربوط به موضوع است. یعنی ارزیابی ذهنی پژوهشگر از وسیله و ابزار اندازه گیری. برای روایی صوری پرسشنامه، باید شکل سوال ها، منطقی و متناسب با ویژگی های پاسخ دهندگان باشد.

روایی محتوی

- روایی محتوا، یعنی ترسیم و تعیین خوب ابعاد و مؤلفه های مفهوم یا متغیر است.
- روایی محتوا، اشاره می کند که نمونه سؤال های مورد استفاده در یک آزمون تا چه حد معرف کل جامعه سؤال ها است. به عبارتی، روایی محتوا به این سوال پاسخ میدهد که آیا سوال ها یا گویه های ابزار مورد نظر، به طور منطقی و مناسب، مفهوم و متغیر مورد نظر را اندازه گیری می کند؟
- یا این که اطمینان می دهد که ابزار مورد نظر به تعداد کافی پرسشهای مناسب برای اندازه گیری مفهوم (متغیر) مورد سنجش در بردارد؟
- برای مثال اگر استاد یک درس خاص بخواهد در پایان ترم یک آزمون «روا» از دانشجویان به عمل آورد، آزمون وی باید دارای سوالاتی باشد که معرف تمام مطالب تدریس شده در کلاس باشد و دانشجویان را به خوبی در آن درس ارزیابی کند.
- روایی محتوا، به پژوهشگر اطمینان میدهد که ابزار مورد نظر به حد کافی توان مناسب برای اندازه گیری مفهوم مورد سنجش را در بر دارد.

شاخص نسبت روایی محتوایCVR

- این شاخص توسط لاوشه طراحی شده است. جهت محاسبه این شاخص از نظرات کارشناسان متخصص در زمینه محتوای آزمون مورد نظر استفاده می شود و با توضیح اهداف آزمون برای آن ها و ارائه تعاریف عملیاتی مربوط به محتوای سؤالات به آن ها، از آن ها خواسته می شود تا هریک از سؤالات را بر اساس طیف سه بخشی لیکرت «گویه ضروری است»، «گویه مفید است ولی ضروری نیست» و «گویه ضرورتی ندارد» طبقه بندی کنند. سپس بر اساس فرمول زیر، نسبت روایی محتوایی محاسبه می شود:

$$CVR = \frac{\text{تعداد متخصصینی که گزینه ضروری را انتخاب کرده اند} - \frac{\text{تعداد کل متخصصین}}{۲}}{\frac{\text{تعداد کل متخصصین}}{۲}}$$

حداقل مقدار □□□ قابل قبول بر اساس تعداد متخصصین نمره گذار□					
تعداد متخصصین□	مقدار □□□□	تعداد متخصصین□	مقدار □□□□	تعداد متخصصین□	مقدار □□□□
□۵	□ـ□	□۱۱	□ـ□	□۲۵	□ـ□
□۶	□ـ□	□۱۲	□ـ□	□۳۰	□ـ□
□۷	□ـ□	□۱۳	□ـ□	□۳۵	□ـ□
□۸	□ـ□	□۱۴	□ـ□	□۴۰	□ـ□
□۹	□ـ□	□۱۵	□ـ□	□	□
□۱۰	□ـ□	□۲۰	□ـ□	□	□

شاخص روایی محتوایی CVI

- جهت بررسی شاخص روایی محتوا از روش والتز و باسل (Waltz & Bausell) استفاده می شود . بدین صورت که متخصصان «مربوط بودن»، «واضح بودن» و «ساده بودن» هر گویه را بر اساس یک طیف لیکرتی ۴ قسمتی مشخص می کنند. متخصصان مربوط بودن هر گویه را از نظر خودشان از ۱ «مربوط نیست»، ۲ «نسبتاً مربوط است»، ۳ «مربوط است»، ۴ «کاملاً مربوط است» مشخص می کنند. ساده بودن گویه نیز به ترتیب از ۱ «ساده نیست»، ۲ «نسبتاً ساده است»، ۳ «ساده است»، ۴ «ساده و مربوط است» و واضح بودن گویه نیز به ترتیب از ۱ «واضح نیست»، ۲ «نسبتاً واضح است»، ۳ «واضح است»، تا ۴ «واضح و مربوط است» مشخص می شود

$$CVI = \frac{\text{تعداد متخصصینی که به گویه نمره ۳ و ۴ داده اند}}{\text{تعداد کل متخصصین}}$$

- حداقل مقدار قابل قبول برای شاخص CVI برابر با 0.79 است و اگر شاخص CVIگویه ای کمتر از 0.79 باشد آن گویه بایستی حذف شود.

نمونه پرسشنامه

به نام خدا

همکار گرامی،

لطفا نظرات خود در مورد هر آیتم پرسشنامه را با تیک زدن در ستون مورد نظر مشخص کنید. همچنین خواهشمند است، نظرات اصلاحی خود در مورد هر آیتم در زمینه شیوه نگارش (نگارش سلیس و روان و استفاده از واژه‌های ساده و قابل فهم)، و قرارگیری سوالات در جای مناسب را در فضای خالی تعبیه شده در زیر هر آیتم مرقوم فرمائید.

حوزه ۱:

ردیف	عبارات	نسبت روایی محتوا (CVR) یکی از سه مورد زیر انتخاب شود			شاخص روایی محتوا (CVI)											
		ضروری	مفید، ولی غیر ضروری	غیر ضروری	مرتبط بودن یکی از چهار مورد زیر انتخاب شود				واضح بودن یکی از چهار مورد زیر انتخاب شود				سلیس و روان بودن یکی از چهار مورد زیر انتخاب شود			
					کاملاً مرتبط	مرتبط	نسبتاً مرتبط	غیر مرتبط	کاملاً واضح	واضح	نسبتاً واضح	غیر واضح	کاملاً سلیس	سلیس	نسبتاً سلیس	غیر سلیس
۱																

پیشنهادات اصلاحی:

ردیف	عبارات	نسبت روایی محتوا (CVR) یکی از سه مورد زیر انتخاب شود			شاخص روایی محتوا (CVI)											
		ضروری	مفید، ولی غیر ضروری	غیر ضروری	مرتبط بودن یکی از چهار مورد زیر انتخاب شود				واضح بودن یکی از چهار مورد زیر انتخاب شود				سلیس و روان بودن یکی از چهار مورد زیر انتخاب شود			
					کاملاً مرتبط	مرتبط	نسبتاً مرتبط	غیرمرتبط	کاملاً	واضح	نسبتاً واضح	غیر واضح	کاملاً سلیس	سلیس	نسبتاً سلیس	غیر سلیس
۱	مدرس به موقع در کلاس حضور پیدا میکند	*			*				*				*			
پیشنهادات اصلاحی:																
۲	مدرس طول زمان کلاس را رعایت میکند و مبحث را در زمان تعیین شده به اتمام میرساند		*		*					*				*		
پیشنهادات اصلاحی:																

Conduct cognitive interviews

- Similar to how experts are utilized to determine the content validity of a new survey, it is equally important to determine how potential respondents interpret the items and if their interpretation matches what the survey designer has in mind

- Results from cognitive interviews can be helpful in identifying mistakes respondents make in their interpretation of the item or response options.
- The sample sizes used for cognitive interviewing are normally small and may involve just 10–30 participants

Table 4. Examples of commonly used verbal probes.

Type of verbal probe	Example
Comprehension/interpretation	''What does the term 'continuing medical education' mean to you?''
Paraphrasing	''Can you restate the question in your own words?''
Confidence judgment	''How sure are you that you have participated in 3 formal educational programs?''
Recall	''How do you remember that you have participated in 3 formal educational programs?''
	''How did you come up with your answer?''
Specific	''Why do you say that you think it is very important that physicians participant in continuing medical education?''
General	''How did you arrive at that answer?''
	''Was that easy or hard to answer?''
	''I noticed that you hesitated. Tell me what you were thinking.''
	''Tell me more about that.''

Conduct pilot testing

- Reliability analysis
- Cronbach's alpha coefficient

Factor Analysis

- Construct validity

Glossary

Closed-ended question – A survey question with a finite number of response categories from which the respondent can choose.

Cognitive interviewing (or cognitive pre-testing) – An evidence-based qualitative method specifically designed to investigate whether a survey question satisfies its intended purpose.

Concurrent probing – A verbal probing technique wherein the interviewer administers the probe question immediately after the respondent has read aloud and answered each survey item.

Construct – A hypothesized concept or characteristic (something "constructed") that a survey or test is designed to measure. Historically, the term "construct" has been reserved for characteristics that are not directly observable. Recently, however, the term has been more broadly defined.

Content validity – Evidence obtained from an analysis of the relationship between a survey instrument's content and the construct it is intended to measure.

Factor analysis – A set of statistical procedures designed to evaluate the number of distinct constructs needed to account for the pattern of correlations among a set of measures.

Open-ended question – A survey question that asks respondents to provide an answer in an open space (e.g. a number, a list or a longer, in-depth answer).

Reliability – The extent to which the scores produced by a particular measurement procedure or instrument (e.g. a survey) are consistent and reproducible. Reliability is a necessary but insufficient condition for validity.

Journals and databases

Mitra Amini

Professor of Shiraz University of Medical Sciences, Shiraz, Iran

FAIMER fellow, Assessor of WFME, Associate Editor of BMC Medical Education Journal

Select the right journal

- Think about the type of papers published in medical education journals and which type is suitable for your idea.
- Examine papers published in the journal and learn whether the journal publishes the type of work you plan to conduct.

- Check the average time between submission of a manuscript, finalization of the review process and the final acceptance and publication.
- Study the author's guidelines, the journal's requirements, and whether you need to discuss your idea with the editor first.

- Journal Impact Factor
- The journal's rejection rate
- Are there processing fees?
- Is it an open access journal?

Academic Medicine

- Impact factor: 5.083
- Journal Abbreviation: Acad Med
- Association of American Medical College.
- Articles, Perspectives, Commentaries, Reviews, Research reports, Innovation reports, Letter to the Editor, Teaching & learning moments.

Medical Education

- Impact factor: 4.619
- Journal Abbreviation: Med Educ
- Association For the Study of Medical Education (ASME)
- Original articles, Reviews, Commentaries, The cross-cutting edge articles, Really Good Stuff, Letter to the editor

Medical Teacher

- Impact factor: 2.706
- Journal Abbreviation: Med Teach
- Association for Medical Education in Europe(AMEE).
- Original articles, Reviews, Commentaries, AMEE Guide, Twelve Tips, Letter to the editor, How we, Short communication

Teaching and Learning in Medicine

- Impact factor: 2.216
- Journal Abbreviation: Teach Learn Med
- Taylor & Francis Group.
- Groundwork, Validation , Investigations ,Educational Case Reports ,Observations

Advances in Health Sciences Education

- Impact factor: 2.761
- Journal Abbreviation: Adv Health Sci Educ
- Springer Netherlands
- Research article, review, reflection

The Clinical Teacher

- Journal Abbreviation: Clin Teach
- Willey online Library
- Original article, insights, letter to editor

BMC Medical Education

- Impact factor: 1.870
- Journal Abbreviation: BMC Med Educ
- Biomed central group
- Research article, Technical advance article, Database article, Software article, Study protocol, Review

BMC Research notes

- Journal Abbreviation: BMC Res Notes
- BMC group
- Research notes, Data notes

Iran

- Journal of Education and Health Promotion(Pubmed Scopus).
- Journal of Advances in Medical Education and Professionalism(Pubmed).
- Medical Journal of Islamic Republic of Iran(Pubmed Scopus)
- Archives of Iranian Medicine(Impact factor :1.2)

Search

AMEE

http://www.amee.org/

سایت انجمن آموزش پزشکی اروپا

ارائه کنندۀ اطلاعات کاربردی در زمینه آموزش پزشکی می باشد.

امکانات:

- فراهم نمودن امکان دسترسی به ژورنال Medical Teacher برای اعضا
- submit مقالات در این مجله
- ارائه دهنده آخرین اطلاعات در زمینه کنفرانس‌های برگزار شده و آتی در آموزش پزشکی
- امکان جستجو در سایت

بخش Links and Resources:

با هدف گسترده نمودن دسترسی به منابع اطلاعاتی در آموزش پزشکی امکان پیوند با:

- سازمان ها
- پایگاه ها
- مجلات تخصصی در آموزش پزشکی

فراهم گردیده

- Medical Education Associations
- Medical Education Journals
- Useful databases and sites

AMEE لیست مجلات آموزش پزشکی به شرح ذیل را به شما ارئه می دهد

- Academic Medicine (The journal of AAMC)
- Education for Health (the journal of The Network: TUFH)
- Teaching and Learning in Medicine
- JAMA
- Medical Education (the journal of ASME)
- Medical Teacher (the journal of AMEE)
- BMC Medical Education

جهت جستجو اطلاعات در آموزش **AMEE** از پایگاه های معرفی شده درسایت پزشکی می توان به موارد زیر اشاره نمود.

- British Education Index
- BUBL Information Service
- Combined Health Information Database
- Education Online
- ERIC
- Medcast.com
- MedEdCentral
- Medline via PubMed
- METRO - Medical Education Taxonomy Research Organisation
- School of Health Professions Education (SHE)
- TIMELIT
- TRIP

Best Evidence (BEME)

دراین بخش با توجه به اهمیت مستندات در نتیجه گیری های آموزشی مقالات مروری در آموزش پزشکی جهت اطلاع رسانی فراهم می باشد.

Pubmed سایت

- ارائه کننده پایگاه مدلاین، فراهم کننده اطلاعات ارزشمند در رشته های علوم پزشکی می باشد.

- این سایت در جستجو اطلاعات مورد نیاز در زمینه **آموزش پزشکی** نیز بسیار توانا می باشد

- علاوه بر ارائه مقالات مجلات موجود در آموزش پزشکی مانند (Medical teacher, Medical education, Academic…) اطلاعات از سایر مجلات در آموزش پزشکی را نیز برای شما فراهم می نماید.

جهت آشنایی با مجلات ارائه کننده مقالات در آموزش در این سایت با رفتن به بخش **sign citation**, کلمات Education، learn، teach به شما لیستی از مجلات در زمینه آموزش در رشته پزشکی و سایر رشته های مرتبط ارائه خواهد نمود.

ERIC database

پایگاه ERIC

پایگاه **ERIC** فراهم کننده دسترسی رایگان به بیش از ۱۰۲ میلیون اطلاعات کتابشناسی از مقالات مجلات و دیگر منابع مرتبط به آموزش می باشد.

- در مواردی که Full Text مقالات موجود باشند این پایگاه دسترسی به مقاله را به صورت رایگان در اختیار بازدید کننده قرار می دهد.

پایگاه **ERIC** توسط بخش آموزش مؤسسه علوم آموزش آمریکا حمایت می گردد.

پایگاه ERIC

منابع نمایه شده در پایگاه به سال ۱۹۹۶ بر می گردد و در بر گیرنده:

- ✓مقالات مجلات
- ✓ کتاب ها
- ✓ تحقیقات
- ✓ مقالات کنفرانس‌ها
- ✓ گزارشات فنی، سیاسی
- ✓ دیگر اطلاعات مرتبط با آموزش

پایگاه ERIC

- بیش از ۶۰۰ عنوان مجله در این پایگاه نمایه می گردد. در اغلب موارد کلیه مقالات در پایگاه نمایه شده و در مورد بعضی از مجلات فقط مقالات مرتبط با آموزش در آن نمایه می گردد.

-کلیهٔ ژورنالهای مهم در آموزش پزشکی مانند **Medical Education** و**Medical Teacher** در این پایگاه نمایه می گردد.

پایگاه ERIC

جستجو از **ERIC** از دو دیدگاه امکان پذیر می باشد.

- جستجو پایه
- جستجو پیشرفته

- در جستجو پایه میتوان در عنوان مقاله، نام نویسنده، جستجو در توصیف گرها (جستجو موضوع) استفاده شود.
- در جستجوی پیشرفته امکان جستجو با استفاده ازمحدودیتهای بیشتر و بهتر فراهم میباشد.

در بخش اول جستجو پیشرفته می‌توانید آن واحد جستجو خود را به عنوان، نویسنده و نام مجله محدود نمایید و در صورت نیاز به استفاده از محدودیت در فیلدهای بیشتر Add Another Row این امکان را برای شما فراهم می نماید.

پایگاه ERIC

محدودیت به بازیابی مقالات تمام متنی (Full Text) محدودیت به سال، نوع منبع (مجله، کتاب، گزارشات و ...) و سطح آموزشی(آموزش بزرگسالان، آموزش ابتدایی، آموزش ...) در بخش دوم جستجو فراهم می باشد.

ویژگی جستجو در ERIC

فراهم بودن مقولهٔ جستجو در نتایج، در این پایگاه باعث افزایش دقت و سرعت در رسیدن به منابع مورد نظر می‌گردد.

آشنایی با پایگاه BMC

http://www.biomedcentral.com/browse/bysubject

آشنایی با پایگاه BMC

- سایت یا بهتر است بگوییم ژورنال الکترونیکی BMC Medical education یک ژورنال با دسترسی آزاد و رایگان می باشد.

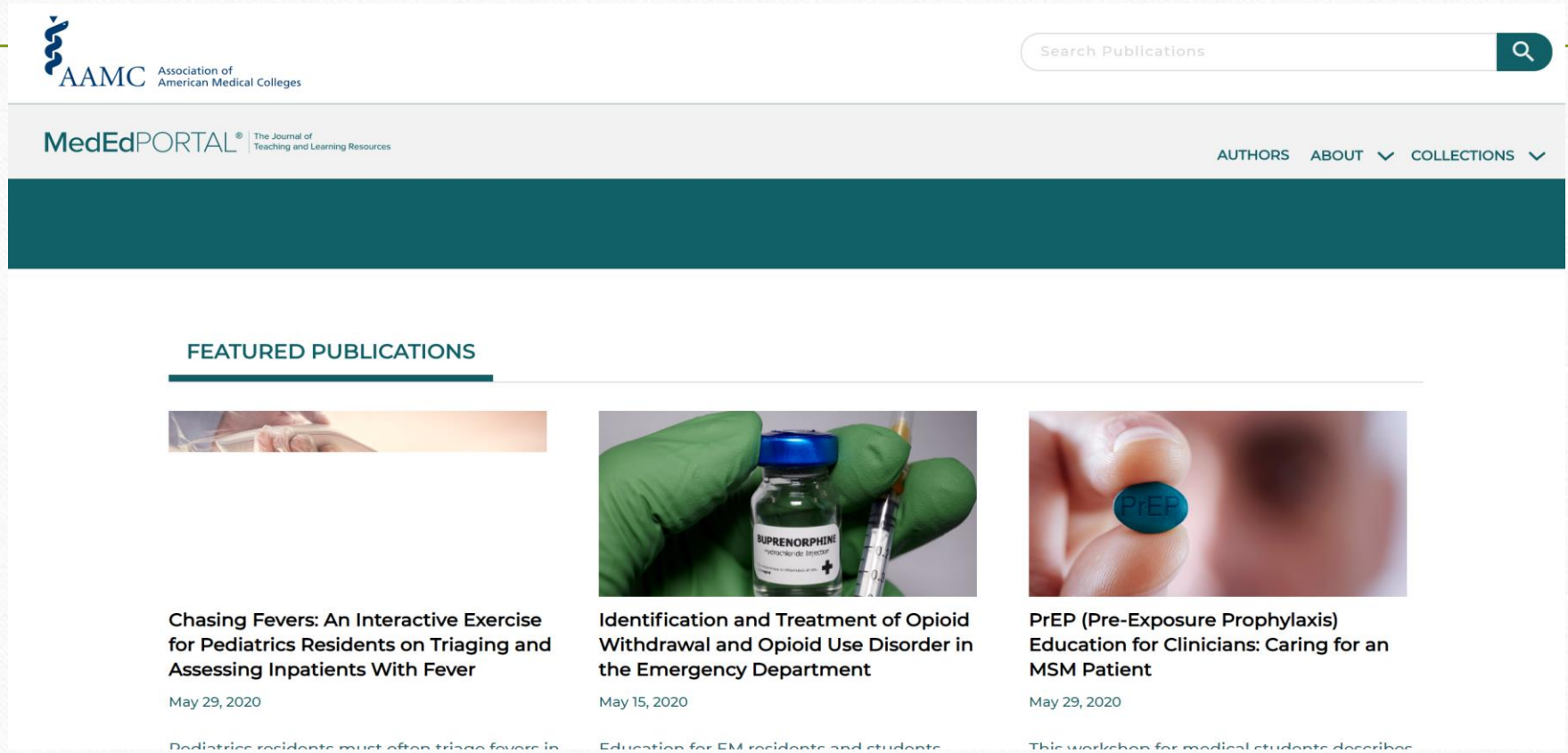
- این ژورنال الکترونیکی ارائه کننده مقالات پژوهشی در زمینه آموزش پزشکی در تحصیلات علوم پایه، تکمیلی و آموزش مداوم می باشد.

- این ژورنال الکترونیکی در پایگاه Pubmed و Medline، CAS، Scopus و EMBAE نمایه می گردد. برای دسترسی آسان به این ژورنال الکترونیکی عضویت در سایت الزامی می باشد. بدین منظور می بایست در بخش My Biomed Central که در منوی بالای صفحه سایت موجود است اقدام به عضویت نمایید.

آشنایی با پایگاه BMC

- سایت مجهز به جستجوی پیشرفته، و همچنین سیستم Alert جهت ارسال مقالات روزآمد در زمینه موضوعی مورد در حواست شما را می نماید که این امکان در صورت عضویت و دریافت Email آدرس ممکن می گردد.

MedEd Portal



Workgroup

- Please choose a topic
- Please choose a journal
- Search about your topic

Final step
Sending the manuscript to a Journal

Mitra Amini

Professor of Shiraz University of Medical Sciences, Shiraz, Iran

FAIMER fellow, Assessor of WFME, Associate Editor of BMC Medical Education Journal

Polish carefully before you submit

- Editors will help you tailor your message to their audience, but you should not expect them to correct poor grammar or typographical errors.
- Use Grammarly premium software if possible.
- Use a consistent font style for each level of subheading.
- Review the title of each table and figure to ensure that it accurately and completely describes the information.

Polish carefully before you submit

- Confirm that data and other details in the abstract match those in the main text.
- Verify the format and accuracy of each reference, including adherence to the journal's formatting style.
- Remove all comments and resolve all tracked changes.

- For those writing in a nonnative language, a skilled native speaker should always proofread the manuscript (note that just because someone speaks English does not mean he or she is a good writer or good proofreader).
- McGaghie (2009) enumerated several additional suggestions for those writing in a non-native language.(in references of 12 tips articles).

Chosing the right journal

- Impact factor.
- Time for acceptance.
- Rejection rate.

Follow journal instructions precisely

- Topics of interest
- Article type
- Manuscript length and required elements
- Abstract structure and length
- Reference citation format
- And more….

Cover letter important structure

- A brief statement about authorship.
- Conflicts of interest.
- Prior publication.

Useful for cover letter

- Focus on the importance of this topic,
- The anticipated impact of these findings on the field
- Why this manuscript is a good match for the chosen journal.

When you are rejected (because you will be)

- Unfortunately, since most journals accept fewer than 20% of the manuscripts they receive, your manuscript will most likely be rejected.
- You will naturally feel discouraged when you get that rejection letter.
- However, rejection is simply part of the publishing game.

- You should never give up after the first rejection!
- Get it back out the door quickly!
- Change it, correct it and submit to other journal.
- Take seriously all reviewer and editor suggestions.(fter a short cooling-down period (rejection is always hard!).

When you are invited to revise (because you will be, eventually)

- Respond carefully to every suggestion, even if you disagree.
- The reviewers are always right!

Classify the comments

- Recognition of poor writing
- Identification of an error;
- Suggestion to elaborate on a theme
- Opinion without suggestions
- Compliment.

Large group work

Some samples of reviewers' comments

Recognition of poor writing

- “The sentence on page 4, line 3, is difficult to understand.”
- Your manuscript needs English editing.

Response

- Correct it

Identification of an error

- “There appears to be an inconsistency between the data reported in the main text and in Table 2.”
- “The investigators used the t-test, but the Wilcoxon rank sum test would have been more appropriate.”
- “There are several other studies addressing this question, including work by [author].
- These should be cited in the Introduction.”
- “The claim that the results apply to practicing physicians is not justified because it extrapolates beyond the data.

Response

- First double check your work, then seriously consider: Is the reviewer correct? If yes correct it.
- If you believe the reviewer is wrong, did his or her error arise because of ambiguous writing? Clarify it.
- Tactfully explain your rationale for change or no change in the response letter. "We neglected to report that we verified the assumptions for the use of parametric tests such as the t-test. We have clarified this point in the Methods, and continue to use the same statistical test."

Suggestion to elaborate on a theme

- “It would be good for the authors to elaborate on the finding that _____.”
- “In discussing this point, the authors may wish to draw in the work by [author].” [note this suggestion is less forceful than the “error” quoted above]
- "The authors spend too much time talking about _____, which is only tangentially related to this topic.
- This should be deleted.”

Response

- For suggested elaborations, seriously consider: Is the message strengthened if you follow this advice?
- If you opt to incorporate a suggested elaboration, it is often appropriate to keep it short.
- If you choose not to make a change, defend your decision by stating something like, “This is an excellent suggestion, and we agree with the reviewer. However, due to space constraints we are not able to address this point fully."

Opinion without suggestion

- "It is interesting to note that this issue arises in the work on cognitive load theory as well."
- "I disagree with the interpretation of finding ____." [a suggestion to change is implied]

Response

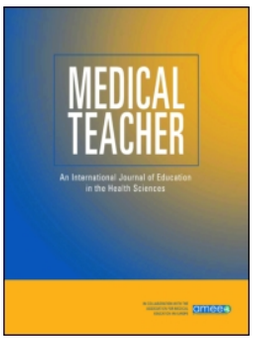
- First carefully consider: Is a specific suggestion hidden in this opinion?
- If yes, then respond accordingly.
- If no, then treat this as a suggested elaboration (above) or compliment (below).

Compliment

- “This is an important question and a very timely study.”
- “The randomized design adds rigor.”

Response

- Acknowledge this graciously and without elaboration in the response letter.



Medical Teacher

ISSN: 0142-159X (Print) 1466-187X (Online) Journal homepage: https://www.tandfonline.com/loi/imte20

Twelve tips on writing abstracts and titles: How to get people to use and cite your work

David A. Cook & Georges Bordage

To cite this article: David A. Cook & Georges Bordage (2016) Twelve tips on writing abstracts and titles: How to get people to use and cite your work, Medical Teacher, 38:11, 1100-1104, DOI: 10.1080/0142159X.2016.1181732

To link to this article: https://doi.org/10.1080/0142159X.2016.1181732

MEDICAL TEACHER, 2016
VOL. 38, NO. 11, 1100–1104
http://dx.doi.org/10.1080/0142159X.2016.1181732

MEDICAL TEACHER
Taylor & Francis
Taylor & Francis Group

TWELVE TIPS

Twelve tips on writing abstracts and titles: How to get people to use and cite your work

David A. Cook[a] and Georges Bordage[b]

[a]Office of Applied Scholarship and Education Science, Mayo Clinic Online Learning, Knowledge Delivery Center, and Division of General Internal Medicine, Mayo Clinic College of Medicine, Rochester, MN, USA; [b]Department of Medical Education, University of Illinois at Chicago, Chicago, IL, USA

ABSTRACT

The authors share 12 practical tips on creating effective titles and abstracts for a journal publication or conference presentation. When crafting a title authors should: (1) start thinking of the title from the start; (2) brainstorm many key words, create permutations, and ask others for input; (3) strive for an informative and indicative title; (4) start the title with the most important words; and (5) wait to finalize the title until the very end. When writing the abstract, authors should: (6) wait until the end to write the abstract; (7) copy and paste from main text as the starting point; (8) start with a detailed structured format; (9) describe what they did; (10) describe what they found; (11) highlight what readers can do with this information; and (12) ensure that the abstract aligns with the full text and conforms to submission guidelines.

Introduction

An engaging title and informative abstract, whether for a journal publication or a conference presentation, will help capture the attention of readers long enough for them to stop and learn more about your study and its implications. For a journal publication, you want readers (e.g. researchers, clinicians, educators, or policymakers) to discover your work, recognize its relevance and merit, read it, use it, and cite it in their own publications. For a conference abstract, you want participants to attend your session or stop at your poster, listen to your brief presentation, recognize the relevance and merit of your work, join your professional network, and anxiously await your full-text publication. In either case, potential readers will be quickly skimming many other titles and abstracts (e.g. PubMed search results, journal table of contents, conference proceedings, or other poster boards) all competing for their attention. Your title and abstract must stand out from the rest and communicate in very few words a captivating message.

Evidence suggests that there is substantial room for improvement in both titles and abstracts (Narine et al. 1991; Taddio et al. 1994; Pitkin & Branagan 1998; Pitkin et al. 1999, 2000; Dryver & Hux 2002; Siegel et al. 2005; Cook et al. 2007b). Guidelines developed by national workgroups recommend specific abstract structure and content (Haynes et al. 1990; Hopewell et al. 2008; Moher et al. 2009), and others have offered suggestions on the content and structure of titles and abstracts (Bordage 1989; Huth 1999; Day & Gastel 2012; Bordage et al. 2015; Cook 2016). The purpose of these Twelve Tips is to provide a practical guide for creating effective titles and abstracts. We do not dwell on specific content, but rather focus on the process. Our primary audience are authors writing full-text journal manuscripts, but most tips are also relevant to conference proposals.

Practice points

- The most important factor in getting your work found, read, used, and cited is an informative, indicative title in which the key words come first.
- Reporting detailed results in the abstract is the second most important factor in getting your work used and cited. We encourage use of the "more informative abstract" structure or similar subheadings, and the reporting of actual numeric data and qualitative themes.
- Reporting specific, actionable conclusions in the abstract is the third most important factor. Don't make readers guess at how these results will change what they do.

The title

Tip 1: Start thinking of the title from the start

Authors often neglect the title—putting it off until the end, and perhaps investing little time or thought in its creation. This is a significant oversight! The title is the first—and, if you're not careful, the *last*—thing a potential reader will read about your work. An engaging, descriptive title will entice the reader to read more, whereas titles that fail to accurately and concisely convey the message of the study will allow readers to skip ahead to the next abstract or article.

An informative title is the *single most important thing* that will get your article read, used, and cited, or your conference presentation attended and applied.

As such, the title merits more attention, at least proportionate to the number of words, than any other section of your manuscript. Start thinking about potential titles with the very first manuscript draft. Occasionally you will begin a

CONTACT David A. Cook, MD, MHPE cook.david33@mayo.edu Division of General Internal Medicine, Mayo Clinic College of Medicine, Mayo 17-27W, 200 First Street SW, Rochester, MN 55905, USA

manuscript with a fairly good idea for the title, but it should still be refined as suggested in the following tips.

Tip 2: Brainstorm lots of key words, create permutations, and ask coauthors and non-authors for input

The title is your "shortest possible abstract" (Bordage et al. 2015). Not only that, but long titles don't get read (to the point that some journals limit the length of titles). You'll need to make the most of those very few words!

Start by making a list of key words about your study. These might include words related to the topic ("communication skills"), intervention ("virtual patient"), theory ("self-regulated learning"), participants ("surgery residents"), outcomes ("skills retention"), the message ("improved feedback"), or study design ("randomized trial"). For each term in your initial list, try to come up with several alternatives, synonyms, or related expressions. For example, if your study is about resident "communication skills," your list of alternatives might include "counseling skills," "motivational interviewing," or "shared decision-making." Depending on your intended audience and your study's central message, each expression might attract more or less attention. Likewise, if your study enrolled first-year surgery residents, your alternatives might include "junior residents," "postgraduate trainees," "physicians in training," or "surgical interns."

Once you have generated a fairly extensive list of key words, use varying combinations and sequences to generate as many titles as possible (we usually produce at least a dozen candidate titles, often more). Reflect on these permutations, adding to and deleting from the list as you write and refine the manuscript main text.

Get feedback on your candidate titles from as many people as possible. Ask your coauthors to vote on their favorite three titles. Seek input from non-author friends, asking them which one(s) would most entice them to read the abstract.

Tip 3: Strive for an informative and indicative title

Huth (1999) noted that titles can describe what you found (informative titles, e.g. "Video-supported feedback is superior to audio-only feedback"), describe what you did (indicative titles, e.g. "randomized trial"), or both ("Video-supported feedback is superior to audio-only feedback: a randomized trial"). The best titles are usually both informative and indicative. A colon (":") can help to append indicative information to the title, using expressions such as "a systematic review," "a cohort study," "a national survey," or "a grounded theory study." We encourage authors to create titles that contain both informative and indicative elements.

We tend to avoid catchy, dramatic, fad, or gimmicky titles for original research studies or rigorous review articles, because they are easily misinterpreted and typically waste precious words with little specific information (i.e. are less informative and indicative). We also avoid questions ("What is the effect of personalized feedback?") because the answers—the informative element—are more useful. By contrast, for an editorial, commentary, perspective, or less formal review, a catchy, casual, fun, or provocative title is appropriate and often highly effective in attracting desired attention.

As a rule, avoid using abbreviations in the title.

Tip 4: Start the title with the most important words (but don't start with the method)

Don't begin the title with "randomized trial" or "systematic review." These words, while important, do nothing to engage the reader regarding your main message. Numerous randomized trials and systematic reviews are published every week. You must communicate—in the first few words—the key features that distinguish your study from all the rest! Imagine a potential reader sifting through a list of 350 articles from a PubMed search, looking for studies relevant to a study-in-planning or a systematic review. She is scanning articles quickly—looking for certain key words that she has identified as relevant to her work—and her brain is getting tired. You will make her life much easier, and increase the likelihood that she will discover your article, if appropriate key words come *first* in the title. An important indicative phrase (e.g. "randomized trial") can come at the end, once interest is already aroused, to confirm that this article is indeed worth reading.

Bottom line: It is essential not only to have the right key words in the title, but also to position these words where they will most readily catch readers' attention.

Tip 5: Don't finalize the title until the very end

Although you should begin writing the title with the very first manuscript draft, you should not finalize the title until the manuscript is otherwise complete. You need time for the title to percolate and evolve as you receive feedback and generate additional alternatives (see Tip #2, and examples in Table 1).

Moreover, the title should reflect, as accurately, completely, and concisely as possible, the central message of your study; yet that message usually becomes more focused as the manuscript matures. Only when the manuscript is complete can you select the title that most eloquently captures your bottom line. Be especially mindful that your title does not mislead, either by overstating your results or overstating the limits of your study design.

Some journals place limits on the number of words or characters in the title, and some have additional requirements regarding subtitles and phrasing. Read the Instructions to Authors carefully, and adhere to any requirements.

The abstract

Tip 6: Wait until near the end to write the abstract

In contrast with the title, wait until the manuscript is nearly complete before starting on the abstract. Sometimes creating a draft abstract early on can help to organize your thinking and provide structure for the manuscript as a whole. However, for practical reasons it usually makes sense to defer writing the final abstract until the very end. First, creating the abstract is easier once you have the full manuscript to draw from. You will work from what you actually said, rather than what you anticipate saying. Second, as with the title, you may not know until the end what data and conclusions are most central to your message. Third, waiting prevents problems with version control—that is, the

Table 1. Examples of title transformations.

Poor title	Better title	Rationale
Systematic review of assessments of medical student self-regulated learning	Assessing self-regulated learning in medical students: a systematic review	Don't start off with the study design; begin with words most likely to attract attention of potential readers
Does CBL work for medical student lectures?	Case-based vs non-case-based lectures for second-year medical students: a nonrandomized controlled study *Or* Improved retention with case-based vs non-case-based lectures for medical students: a nonrandomized controlled study	The acronym "CBL" could mean many other things including computer-based learning The poor title does not mention the comparison or the study design The poor title does not specify the stage of medical student training or the main findings; each better title resolves one of these deficits
Evaluating the role of Schmidt's intermediate effect in facilitating the cognitive development of third year medical students in a nephrology clinical rotation in Kenya	Reproducing the intermediate effect in third-year medical students: a randomized trial *Or* Clinical reasoning in third-year medical students: reproducing the intermediate effect	The poor title is very long and contains details that distract from the main message The better titles are shorter, more focused, and put key words up front. The choice between these two would depend on which key words would best attract your target audience
Give me credit for what I've done: improving maintenance of certification	Facilitating maintenance of certification for internal medicine physicians: a focus group study	The poor title might be appropriate as an editorial or perspective, but it is a bit informal for most original research articles. Note that the better title puts the key words near the front, to better attract attention

inconsistencies that inevitably creep in when you revise the main text after writing the abstract.

This rule is slightly different for short abstracts submitted for conference presentation. In those cases, preparing the abstract based on preliminary data and tentative analyses is common.

In creating an abstract you will invest meticulous effort in selecting key content, weighing the importance of each word, and iteratively polishing the prose. These steps are best done once—near the end of the writing process.

Tip 7: Copy and paste from the main text as the starting point

Copying entire sentences, including data, from the main text into the abstract is a highly efficient way to begin the abstract. As a rough start, pull in two to three key sentences each from the Introduction (including the statement of study intent [research question, hypothesis, or purpose] (Cook et al. 2007a)), Methods, and Discussion (often from the first or concluding paragraph), and several sentences from the Results (at least one sentence from each main analysis).

The resulting text will invariably be too long, unfocused, and disjointed. You will need to edit extensively to eliminate unnecessary details, extra words, and tangential thoughts to provide coherence and a natural flow, but at least you will be refining text rather than starting from scratch. Remove all references to other publications.

Tip 8: Start with a detailed structured format—Even if the journal doesn't require or allow it

Many journals favor unstructured abstracts or the familiar "Purpose–Methods–Results–Conclusions" format. Such simple abstracts commonly lack important details regarding methods and results (Taddio et al. 1994; Cook et al. 2007b). In 1987 the Ad Hoc Working Group for Critical Appraisal of the Medical Literature proposed standards for the "more informative abstract" (Huth 1987). These standards have since been revised (Haynes et al. 1990), and now include headings for background, objective, design, setting, participants, interventions, outcomes, results, and conclusions. In addition, many reporting guidelines now recommend specific abstract formats depending on the study design, such as systematic reviews (Moher et al. 2009) and randomized trials (Hopewell et al. 2008).

We strongly encourage the use of detailed structured abstracts (i.e. using the "more informative abstract" headings or headings appropriate to the specific study design). The detailed structure itself promotes readability, but more importantly the detailed structure encourages the inclusion of more information. Some authors and journals view the abstract like a movie trailer—a teaser to entice the potential reader to read the entire article. While it certainly does need to attract readers, the abstract facilitates additional tasks including peer review, database indexing, literature searches, and critical appraisal. The more information an abstract contains, the better it serves these additional functions. Also, when choosing which of several "competing" articles to read, readers will welcome the information contained in a detailed abstract. The more information in your abstract, the more likely it is that readers will find, recognize as relevant, read, and cite your work.

We encourage authors when writing their abstract to use the detailed structure most closely aligned with their study, even if the journal prefers a different format. The detailed headings encourage inclusion and sequencing of essential information and will also help you visualize the relative amount of text in each section (e.g. is the Conclusion longer than the Results?). If the journal does not permit a detailed structured format, you can always omit the structured headings and make minor adjustments in wording during later stages. For example, when writing the abstract to accompany a systematic review, use the structure recommended in the PRISMA guidelines (Moher et al. 2009) during the initial drafting, and then modify the headings if needed (e.g. to a Purpose–Methods–Results–Conclusions format) prior to submission.

Include in the abstract the key words you could not fit into the title. Literature searches in databases like MEDLINE or PubMed usually search the abstract along with the title

and formal indexing terms. Strategically ensuring that alternative expressions for key concepts are present in the abstract will increase the likelihood that a search will identify your article.

Tip 9: Describe what you did

Using the most appropriate structure, succinctly summarize the key aspects of your study. If you tested a theory, mention that theory by name. If you evaluated an intervention, briefly describe that intervention. If you compared that intervention against another intervention or a control group, describe what happened in the comparison arm. Briefly describe the eligible participants, study procedures, main outcome measures, and methods for qualitative data collection and analysis. We usually do not spend words naming specific quantitative statistical tests (e.g. "chi-squared test" or "t-test") unless they are particularly important in correctly interpreting the study results (e.g. a novel, unusual, or controversial statistical technique).

Tip 10: Describe what you found

Dedicate adequate space and attention to reporting your results. The Results section might easily comprise more than one-third of the total abstract length. Again, the more detail, the better. In quantitative studies, avoid using vague terms such as "increased" or "statistically significant difference." Instead, report the actual numbers and *p*-value or confidence interval; for example, "Posttest knowledge scores were similar in the case-based (mean [standard deviation]: 75.0 [12.3]) and non-case-based groups (74.7 [12.6]); 95% confidence interval for the difference, −4.4 to 5.0 ($p = 0.90$)." For qualitative research studies, instead of stating that "four themes were identified," report the actual themes. In a review or perspective, summarize the key points in as much detail as possible.

Reporting detailed results is the *second most important factor* in getting people to read and cite your work, second only to an informative, indicative title.

Tip 11: Highlight what the reader can do with this information

End the abstract with a concise conclusion that highlights defensible bottom line messages. Don't force readers to make inferences about your study findings; tell them outright what the results mean.

The conclusions should be brief—two or at most three sentences. They should not restate or summarize the results; the Results section is itself already a brief summary. Rather, use the conclusions to unambiguously but realistically and justifiably tell readers why these results are important and how this information will advance the field and change what they do.

Conclusions should be supported by the Results presented in the abstract (i.e. readers should not have to read the main text in order to justify the abstract's conclusions). If abstract results do not support the conclusions, then either add information to the abstract or adjust the conclusions accordingly. If you've properly selected the most important results and highlighted the most salient messages, such lack of alignment should not occur.

Reporting specific, actionable conclusions is the *third most important factor* in getting people to read and use your work.

Tip 12: Make sure the abstract aligns with the full text and conforms to submission guidelines

As a final step before submission, *print the abstract and manually verify each element* against the main text, with special attention to the results. Nothing should be reported in the abstract that is not reported in the main text, including methods, results, and conclusions (Bordage et al. 2015); yet inconsistencies arise when, for example, a preliminary statistical analysis is re-run with a slightly different technique or deleted altogether. Thus, every count, response rate, mean, standard deviation, *p*-value, qualitative theme, etc. must be checked for consistency. Carefully compare each datum in the abstract (numeric or narrative) one by one against the main text to verify that all data reported in the abstract were also reported in the main text, and that all data in the abstract match the data in the main text. Waiting to write the abstract until the main text is complete, and copy-and-pasting as suggested above, can prevent most problems in this regard.

Although journal instructions or conference submission guidelines should have been consulted from the outset, now is the time to verify one last time that the abstract adheres to requirements regarding word length, structure, and style (e.g. the journal may require that their abstracts be written in third person).

Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

Notes on contributors

David Cook, MD, MHPE, is an Associate Director of Mayo Clinic Online Learning, Director of the Knowledge Delivery Center Research and Measurement Group, and Professor of Medicine and Medical Education, Mayo Clinic College of Medicine; and a Consultant in the Division of General Internal Medicine, Mayo Clinic, Rochester, Minnesota, USA.

Georges Bordage, MD, PhD, is a Professor of Medical Education in the Department of Medical Education in the College of Medicine at the University of Illinois at Chicago, Chicago, USA.

References

Bordage G. 1989. Considerations on preparing a paper for publication. Teach Learn Med. 1:47–52.

Bordage G, McGaghie WC, Cook DA. 2015. Title, authors, and abstract. In: Durning SJ, Carline JD, editors. Review criteria for research manuscripts. Washington (DC): Association of American Medical Colleges. p. 148–156.

Cook DA. 2016. Twelve tips for getting your manuscript published. Med Teach. 38:41–50.

Cook DA, Beckman TJ, Bordage G. 2007a. Quality of reporting of experimental studies in medical education: a systematic review. Med Educ. 41:737–745.

Cook DA, Beckman TJ, Bordage G. 2007b. A systematic review of titles and abstracts of experimental studies in medical education: many informative elements missing. Med Educ. 41:1074–1081.

Day R, Gastel B. 2012. How to write and publish a scientific paper. Cambridge (UK): Cambridge University Press.

Dryver E, Hux JE. 2002. Reporting of numerical and statistical differences in abstracts. Improving but not optimal. J Gen Intern Med. 17:203–206.

Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. 1990. More informative abstracts revisited. Ann Intern Med. 113:69–76.

Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, Schulz KF. 2008. CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. PLoS Med. 5:e20.

Huth EJ. 1987. Structured abstracts for papers reporting clinical trials. Ann Intern Med. 106:626–627.

Huth EJ. 1999. Writing and publishing in medicine. Baltimore: Williams and Wilkins.

Moher D, Liberati A, Tetzlaff J, Altman DG. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med. 151:264–269.

Narine L, Yee DS, Einarson TR, Ilersich AL. 1991. Quality of abstracts of original research articles in CMAJ in 1989. Can Med Assoc J. 144:449–453.

Pitkin RM, Branagan MA. 1998. Can the accuracy of abstracts be improved by providing specific instructions?: a randomized controlled trial. JAMA. 280:267–269.

Pitkin RM, Branagan MA, Burmeister LF. 1999. Accuracy of data in abstracts of published research articles. JAMA. 281:1110–1111.

Pitkin RM, Branagan MA, Burmeister LF. 2000. Effectiveness of a journal intervention to improve abstract quality. JAMA. 283:481.

Siegel PZ, Thacker SB, Goodman RA, Gillespie C. 2005. Titles of articles in peer-reviewed journals lack essential information: a structured review of contributions to 4 leading medical journals, 1995 and 2001. In: Rennie D, editor. Fifth international congress on peer review. Chicago, Illinois: Biomedical Publication.

Taddio A, Pain T, Fassos FF, Boon H, Ilersich AL, Einarson TR. 1994. Quality of nonstructured and structured abstracts of original research articles in the British Medical Journal, the Canadian Medical Association Journal, and the Journal of the American Medical Association. Can Med Assoc J. 150:1611–1615.

Medical Teacher

ISSN: 0142-159X (Print) 1466-187X (Online) Journal homepage: https://www.tandfonline.com/loi/imte20

Writing for academia: Getting your research into print: AMEE Guide No. 74

John H. Coverdale, Laura Weiss Roberts, Richard Balon & Eugene V. Beresin

To cite this article: John H. Coverdale, Laura Weiss Roberts, Richard Balon & Eugene V. Beresin (2013) Writing for academia: Getting your research into print: AMEE Guide No. 74, Medical Teacher, 35:2, e926-e934, DOI: 10.3109/0142159X.2012.742494

To link to this article: https://doi.org/10.3109/0142159X.2012.742494

2013; 35: e926–e934 MEDICAL TEACHER

WEB PAPER

Writing for academia: Getting your research into print: AMEE Guide No. 74

JOHN H. COVERDALE[1], LAURA WEISS ROBERTS[2], RICHARD BALON[3] & EUGENE V. BERESIN[4,5]

[1]Baylor College of Medicine, USA, [2]Stanford University School of Medicine, USA, [3]Wayne State University School of Medicine, USA, [4]Harvard University, USA, [5]Massachusetts General Hospital and McLean Hospital, USA

Abstract

The authors identify and describe strategies for success in writing for publication, including how to choose an educational research topic, define the question and choose the correct design, know the anatomy of a research paper, write each of the sections, optimize the writing before publication, choose a journal, and respond to editors and reviewers. The research question should be focused, modest, and achievable given the constraints of the setting, significant, and appropriately imbedded in the available literature. The choice of methods is determined by the nature of the educational research question and should conform to ethical standards. Specific strategies for writing include starting where it is easiest to do so, spontaneously and uncritically writing the first paragraphs through, minimizing external impediments to the work, and knowing how each section of a manuscript is routinely structured. All papers require a number of revisions with careful attention to accuracy and detail as well as to specific requirements of the target journal before submission. Authors should respond positively, not defensively, and in detail to all of the editor's and reviewers' suggestions for revision. Writing for success is therefore a disciplined and systematic process following prescribed steps, which, although hard work, is rewarding.

Introduction

Medical educational research is optimally a systematic inquiry intended to extend knowledge or to solve a research question of interest in the educational preparation of medical students, residents, specialty and subspecialty fellows, biomedical scientists, allied health trainees, and practicing clinicians. This field of scholarship serves the critically important functions of enriching educational theory and practice by substantiating "best evidence medical education" (Harden et al. 2000; Hart & Harden 2000) and, ultimately, it is hoped that this work will benefit patients.

Medical education research is undergoing enormous expansion, and wide-ranging opportunities are available for contributing to this growth. This AMEE Guide is intended to promote medical education research by helping researchers at all levels of experience to be successful in preparing and publishing an educational research project. This Guide is the third in a medical education series on general research topics and follows papers on a general introduction to research (Ringsted et al. 2011) and on writing an educational research and grant proposal (McGaghie 2009).

Our goals for this Guide are to identify and describe strategies for success in writing for publication. These strategies include choosing an educational research topic, defining the question and choosing the research design, knowing the "anatomy" of a research paper and how to write each of the sections, and optimizing the writing before submission to a journal. We will also talk about the factors to consider in the choice of a journal in which to publish and how to respond to any comments by editors and reviewers. Our focus is on educational research, not other forms of writing such as reviews, annotated bibliographies, and commentaries. We want readers to get started and to succeed in their quest to become productive educational researchers and writers.

Practice points

- Writing for success is a systematic, disciplined process.
- The research question should be focused, imbedded in the available literature, and achievable given the available resources.
- The research design is determined by the question, should conform to ethical educational standards, and should be comprehensively described.
- Strategies for writing include starting where it is easiest to do so, spontaneously and uncritically writing the first paragraphs, and identifying and reducing specific barriers to writing.
- Getting the final submission ready requires very careful attention to detail and accuracy.

Correspondence: L.W. Roberts, Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, 401 Quarry Road, Room 3215, Stanford, CA 94305-5717, USA. Tel: 650 723 8290; fax: 650 723 8216; email: RobertsL@Stanford.edu

ISSN 0142–159X print/ISSN 1466–187X online/13/020926–9 © 2013 Informa UK Ltd.
DOI: 10.3109/0142159X.2012.742494

Choosing a topic and getting started

Prospective researchers will appreciate that medical education research grows by accretion or by small gains. Researchers should not think that they have to do a landmark study or a major study, because such an ambition is daunting and can detract from the pleasure of the work. In order to get started, the research goal should be feasible and achievable given resources, including the availability of time, potential participants, finances, and administrative assistance. It should also be appreciated that any question worth answering is worth examining for program evaluation purposes and may be worth publishing. This means that researchers should not let down themselves, their topics of inquiry, their study volunteers, or their potential prospective readership by a lack of rigor or attention to the methodology. Thus, the research question should be focused to allow for an achievable and well-supported answer given the constraints of the setting where the research is to take place (Table 1).

Educational questions can arise from everyday experiences and ideas, whether from clinical rounds or clinical team discussions, *ad hoc* conversations with colleagues, reading educational texts or articles, educational conferences, or any educational teaching or learning experience. As Ringsted et al. (2011) noted, the challenge is to place a concrete idea, interest, or problem within a general context of learning, teaching, and education. Educational research thus is not just about answering local questions but general questions about learning, teaching, and education that are studied in local contexts (Ringsted et al. 2011).

Importance of the question

In choosing a question, the researcher should ask what its significance is. Is the question to be answered relevant to many people or, perhaps, relevant to fewer people but very influential or problematic? Reviewers and editors will view this consideration as crucial in the judgment of the suitability of a manuscript for publication (Roberts et al. 2004). Significance pertains to the prevalence and/or seriousness of an issue and the likelihood that the results will benefit educators and their learners. Significance is also understood by how the study's results might add to the available literature, whether there are few studies on a topic that is timely, whether the study allows for a reassessment of the confidence attributable to the findings of more well-studied topics, or whether the study improves on earlier methodologies (Coverdale et al. 2005). The potential generalizability of study results is also a prime consideration in assigning importance to a question for potential study (Coverdale et al. 2005; Ringsted et al. 2011).

Table 1. Choosing a topic.

Choose a topic with a question that is doable
Choose a topic area for which you have enthusiasm
Identify the importance or significance of the topic
Imbed the topic and question in the related literature
Look for mentorship and constructive criticism on the research idea
Choose capable, enthusiastic, and compatible team members

Embeddedness in the literature

The researcher therefore should carefully appraise the literature on the topic area in order to establish what, if anything, has been written on the topic before. Previous studies on the topic should be appraised for their methodological strengths and weaknesses so that the methodology and context of the currently proposed study is understood. This appraisal will enable a preliminary assessment of how rigorous the study needs to be in order to contribute to that literature. Earlier methodological deficiencies should be identified and addressed, when possible, in the proposed study design. These are critical steps in the decision whether or not to proceed with a research proposal as well as in writing a grant proposal (McGaghie 2009).

Literature searching begins with a well-defined question, including the population of interest, the intervention (and comparison group, when relevant), and relevant outcomes. The search should be relatively comprehensive for the previously mentioned reasons of understanding the potential merits of the proposed study. Thus, search strategies should emphasize sensitivity over specificity. Comprehensive descriptions of how to search the educational literature are available (Haig & Dozier 2003a, b). Greater sensitivity is achieved by using the Boolean operator OR as opposed to AND when combining search terms and by using synonyms of keywords or search terms. Searching should also use more than one database, especially because some educational research articles might be difficult to retrieve from MEDLINE due to inadequate subject headings (Haig & Dozier 2003b). MEDLINE does not include all journals that publish articles on medical education research (Maggio et al. 2011). Education Resources Information Center, for example, is the largest educational research database. Checking the citations in relevant publications on a topic can enhance the process of looking for valuable articles, although such checking is less systematic than searching the literature. Additionally, educational researchers should be sure to pay due attention to studies originating from other countries, because ignoring international studies constitutes a bias and results in a lessened understanding of the field.

At the same time, in the early phases of planning the research, it is not necessary to be exhaustive and thoroughly comprehensive in the search, as would be expected for a systematically conducted review on a topic (Haig & Dozier 2003a). Instead, the goal is to be confident enough that the proposed research will be contributory to the field. To this end, it is also well worth reading the "Introduction" and "Conclusion" sections of similar studies, in order to see how those studies and findings were justified as important.

Choosing a team

Educational research is rarely conducted alone, although this way of proceeding is certainly an option. A team can contribute by providing constructive criticism and mentorship,

providing an independent check of the literature that serves as background for the research, supporting research processes administratively, and participating in particular aspects of the study such as getting the requisite approvals from the Ethics Committee, also known as the Institutional Review Board, in order to proceed, contributing to study design, collecting and interpreting the data, and writing components of the paper. Thus, it is important to think about the skills and expertise that a potential team member might bring, that person's compatibility with other team members, and whether that person has both the requisite enthusiasm and ability to meaningfully contribute. Being certain that there is sufficient methodological expertise, for example, in study design or qualitative or quantitative analysis, is vital to the success of the team. Choosing team members who are enthusiastic, hard working, and capable can also add substantially to the pleasure inherent in the work. The team should also take its time on discussing the value of the research and the prospective paper rather than be in a rush to get started in order to optimize the processes and the final product.

One consideration in creating a team is an expectation that the members will contribute sufficiently to warrant authorship. In general, contributions should be substantial for this purpose. Criteria for authorship include a substantial contribution to the conception and design, acquisition of data or analysis and interpretation of data, drafting the article or revising it critically for important intellectual content, and final approval of the version to be published. The International Committee of Medical Journal Editors requires all three components to be present (International Committee of Medical Journal Editors 2009), although there is debate about the reasonableness of such stringent criteria (Shaw 2011). In our view, authors are publicly accountable for the rigor and professional integrity of the work, and they should have participated in a sufficiently rich manner that the scholarship is strengthened by their work and influence throughout the process. Discussion about the order of authorship may also arise at this early stage of choosing a team, with a general principle being that the person who conceived of the study, and/or the person who does the most work, has the strongest claim to the first authorship. The order of subsequent authorships is determined by the amount of work completed. One convention is that the most senior academic author goes last, although it is not clear how widely this convention is accepted or applied. A very helpful approach is to establish the ground rules and expectations early on for the work ahead.

Under many circumstances, it is especially helpful to identify a statistician in advance of formally beginning a quantitative research project because the design of the study will be shaped by the hypotheses and outcome measures envisioned. Understanding the statistical tests can also be challenging for many researchers. The statistician can help by reviewing the study design and the instruments used to define outcome measures. It is important to rectify identifiable problems in study design before starting and to use valid and reliable outcome measures when these are available. In the absence of valid outcome measures, care should be taken to develop and pilot test a new instrument in accordance with acceptable standards (Sullivan 2011). A statistician's advice can also be sought regarding the practicability of the anticipated analyses for answering the research question(s), which is advice that should at least qualify for an acknowledgment and perhaps co-authorship should the statistician's work be sufficient to fulfill other criteria for authorship. Choosing the right statistical tests and getting the statistics done correctly is an important consideration in the decision by an editor whether or not to publish (Bordage 2001).

Choice of methods

The nature of the educational research question determines the choice of methods to be employed in the planned and disciplined approach to securing its answer and to delineating the parameters of the study (Sackett & Wennberg 1997). As previously indicated, questions should be carefully crafted and focused in order to facilitate the choice of educationally relevant outcome measures. In qualitative research, however, the focus is typically on hypothesis generation as opposed to hypothesis testing. Other than enabling an answer to the research question, the choice of methods should be plausible, address potential confounding variables or biases, validly address subject selection and settings, and allow for unexpected outcomes or events to occur (McGaghie et al. 2001).

Selection of the research design, moreover, should conform to ethical standards that seek to ensure that the overall aim of the work is valuable and that the methods of research are appropriate. These ethical standards seem less salient in education research in which the potential, for example, of true physical risks to volunteers are minimal. Nevertheless, the appropriateness of the question and the adequacy of efforts to limit harm to participants may be important considerations. For these reasons, educational research in the United States is included under the umbrella of federal regulations for human subjects research (Table 2) and, in both American and European settings, must be prospectively approved and overseen by an Institutional Review Board or formally deemed exempt from institutional review (Roberts et al. 2001; Roberts et al. 2005; Hoschl et al. 2012). If a study seeks to clarify whether learners who are women or who are under-represented ethnic/racial minority students perform similarly to male or majority learners, for example, and they do not, the anticipated consequence of negative labeling should be considered by the research team, as well as by the institutional reviewers, and the potential negative impact lessened. To illustrate, in a multicenter study on health care policies and practices of students performed by one of us, there was the possibility that certain medical schools would appear less sophisticated or less compassionate in their policy approaches. The intent of the study was clearly not to expose individual schools but to help raise understanding of how institutional milieu may influence student self-care practices, so the analyses were performed and presented in publications in a manner that allowed for the pattern of compassionate policies and increased appropriate care-seeking to be apparent. Similarly, data regarding women, under-represented minorities, and particularly women who are also under-represented minority students were aggregated across schools to lessen the likelihood that individual students would

Table 2. Definitions and guidelines relevant to educational research involving human subjects (Adapted from www.hhs.gov/ohrp/index.html [Accessed 02 December 2012]).

Research is defined as ''a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge.''

A human subject is defined as ''a living individual about whom an investigator (whether professional or student) conducting research obtains (1) Data through intervention or interaction with the individual, or (2) Identifiable private information.''

Educational research may be formally deemed exempt, but is not required to do so, by an Institutional Review Board if several conditions are met:

(1) Research conducted in established or commonly accepted educational settings, involving normal educational practices, such as (i) research on regular and special education instructional strategies, or (ii) research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods

(2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, and achievement), survey procedures, interview procedures or observation of public behavior, unless:

(i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation

(3) Research involving the use of educational tests (cognitive, diagnostic, aptitude, and achievement), survey procedures, interview procedures, or observation of public behavior that is not exempt under paragraph (b)(2) of this section, if:

(i) the human subjects are elected or appointed public officials or candidates for public office; or (ii) federal statute(s) require(s) without exception that the confidentiality of the personally identifiable information will be maintained throughout the research and thereafter

(4) Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects

(5) Research and demonstration projects which are conducted by or subject to the approval of department or agency heads, and which are designed to study, evaluate, or otherwise examine:

(i) public benefit or service programs; (ii) procedures for obtaining benefits or services under those programs; (iii) possible changes in or alternatives to those programs or procedures; or (iv) possible changes in methods or levels of payment for benefits or services under those programs

be identifiable. Although the study had extensive confidentiality safeguards in place, the small number of underrepresented minority women students placed them at risk for stigma and potential identification. The design of educational protocols, as well as all planned analyses and publication steps, must take considerations such as these into account in order to fulfill the ethical standards of the field.

Moreover, consideration should be given to how the dependent status of students can constitute a constraint on students' autonomy (Roberts et al. 2005). Subtle coercion is a very important issue in medical education research because the research is mostly conducted by teachers with their subjects as students or trainees. Offering credits for class in exchange for participation in a study may constitute such a form of subtle coercion. The same faculty who evaluate students often conduct research, creating a potential conflict of interest (Roberts et al. 2005), and students' opportunities to complain or to appeal may be limited if they think that negative consequences will accrue if they do so. Safeguards include having an independent person or researcher distributing and collecting questionnaires, foregoing rewards in the form of credits for participation, and protecting anonymity of responses (Table 2).

There are a variety of ways for classifying educational research design (Gall et al. 2003; Fraenkel & Wallen 2006; Horn et al. 2009; Ringsted et al. 2011). The AMEE Guide identified four main categories of design: exploratory, experimental, observational, and translational (Ringsted et al. 2011). Separate publications of the Guide series are devoted to describing more fully each of these methodologies. In brief, exploratory studies include descriptive qualitative studies that are used to identify and explain elements of phenomena and their relationships. Observational studies include cross-sectional or correlational studies such as surveys, cohort studies following volunteers forward in time, and case–control studies looking backward in time from a particular outcome. Experimental studies include randomized and non-randomized controlled trials. Translational studies focus on implementing the findings of educational research to real-life settings. Systematic reviews, which are methods for combining and synthesizing the information from studies on the same or similar educational questions of interest (Reed et al. 2005; Hammick et al. 2010), should also be added to this list of categories of design.

All of these methods can provide valuable information. An important strength of controlled trials is the allowance of an assessment of possible causal outcomes. The most rigorous of experimental methods is the randomized controlled trial, although a randomized controlled trial can be difficult or impractical to achieve in some educational settings, especially when there are ethical barriers to randomizing learners. Given important limitations of randomized controlled trials (Prideaux 2002), it is recognized that the quality of research is as much defined by the integrity and transparency of the research philosophy and methods as by the superiority of one research design over another (Bunniss & Kelly 2010).

Once the method for study is selected and a decision has been made to pursue the work, as noted earlier in this text, the ethical safeguard of institutional review is necessary because educational research is human research. According to federal regulations governing human studies in the United States (Department of Health and Human Services), human research is defined as obtaining "data through intervention or interaction with the individual," or obtaining "identifiable private information" (Hoschl et al. 2012). Even if the project merely involves the review and analysis of existing data, the intent to contribute to scholarship and generalized knowledge creates the obligation to obtain approval, or formal exemption, from

the jurisdictional institutional or educational research review board (Hoschl et al. 2012).

Strategies for writing

Elegant writing is always difficult to attain, and for many authors even terrible writing can be hard to birth. Indeed, there are a variety of types of writing problems, which include distaste for writing, lack of time, lack of confidence, anxiety in writing, perfectionism, and difficulty in starting and finishing (Boice 1990). For each author it is important to assess and reflect on specific barriers to writing and sometimes to seek help in developing strategies to overcome them. Writing is not easy for anyone all of the time, and developing the requisite skills requires effort. Keywords in this process are patience, perseverance, and fun (Coverdale et al. 2005). A great deal of practice and perseverance is required to complete the research and the writing, and having fun in this work promotes patience and perseverance.

Getting started, even with the first paragraph alone, will bring focus to the project and builds momentum to follow through. It is as though writing the first paragraph commits the writer to the task at hand, from which point it becomes very difficult to put it aside. It is a very good idea therefore to start where it is easiest to do so, which is often the "Methods" section or the "Introduction." The Methods follow a relatively set script, to be described in the subsequent section, in simply outlining the research design and what was specifically done in meeting the goals set for the study.

Starting with what one is ready to do and spontaneously and uncritically writing the first paragraphs help writers to become unstuck when a lack of confidence, anxiety, a desire for perfectionism, or an inability to get going impedes writing. Spontaneous writing bypasses internal censors, generates rhythm and voice, and builds confidence and abilities to be spontaneous, playful, and creative (Boice 1990). External impediments include lack of time and potential distracters such as noise in the environment, e-mail to attend to, television in the background, or child-care responsibilities and require setting aside even brief periods of personal time relatively free from those distracters.

Many dedicated authors have rituals to help support their effectiveness in writing. Opening the curtains, making a cup of tea, sitting in a particular spot, having necessary books or resources nearby, turning off one's phone, and other such activities may create a comfort in the routine of entering the writing task. In order to protect against daydreaming and a general lack of productivity, it helps to develop a reward system based on the amount written as opposed to time spent. For example, a break might be taken as a form of reward after a self-prescribed number of words or paragraphs are completed. Daily maintenance of such a fixed-ratio schedule of reinforcement fosters considerable productivity over time. Moreover, the fun of writing arises in part from social engagement with other members of the team, working together and not alone, and using the available mentorship of the team is a strategy that fosters productivity (Table 3).

Table 3. Some strategies for writing.

Get started, even with the first paragraph alone
Start where it is easiest to do so
Follow a relatively set script or structure applicable to the anatomy of the section being written
Spontaneously and uncritically write the first draft
Find time to write, relatively free of distractions
Create a reward system based on the amount written
Use the team to help in overcoming specific barriers to writing
Be patient, persevere, and have fun

Anatomy of the paper

All educational research papers, including each of the sections, follow a certain logic and possess a standardized structure. Knowing the anatomy of an educational research paper is an important strategy for success in writing. This section is oriented toward quantitative research while qualitative research will be more thoroughly addressed in another Guide.

Introduction

The "Introduction" section has three important components. The first is to demonstrate the importance or seriousness of the topic area, as well as the relevance or significance (Pangaro & McGaghie 2001; Coverdale et al. 2005; Ringsted et al. 2011) of that topic to the community of readers that the target journal serves. This is the "hook" or the rationale for the paper: why does the question — and therefore, the empirical report addressing the question — *matter*? The second component is to describe what research has been conducted on the topic area previously, including the strengths and weaknesses of the earlier research. The third is to indicate why the current study was undertaken and how it plans to rectify any weaknesses and contribute to the field.

These components together set the stage for a statement of the specific research goals or hypotheses for the current project. In this last paragraph of the Introduction, it is also sometimes helpful to add an additional summary comment about what the reader might gain from the study. In these ways therefore the Introduction serves to reel the reader into reading further.

It should also be appreciated that some educational research papers will require a theoretical or conceptual framework in the Introduction. In this case the Introduction might be longer than was indicated above. In this way, papers for educational journals differ from those for biomedical journals, when the latter tend to leave theoretical issues to the "Discussion" section.

Methods

The "Methods" is the most important section because it provides a sufficiently detailed description to enable exact replication, facilitate critical appraisal of the study and decision making about whether to incorporate the findings into educational practice, and permit an understanding of the modifications needed in order to improve the validity of

subsequent designs and methods (McGaghie et al. 2001; Coverdale et al. 2006). Authors should justify the appropriateness of the Methods in relation to the specific research question. The Methods should describe the population from which the sample was drawn and the means for selecting the study participants and reasoning supporting their selection, the particulars of the setting and possible contextual effects on the procedures, the specific outcome measure and methods used to generate and collect data, and procedures for analyzing the data. Because medical education practice is so variable across jurisdictions, countries and schools, it might be helpful to include a specific subsection of the Methods describing the context of the study. The Methods should also note that subjects provided informed consent and that Institutional Review Board approval was obtained or that the study was deemed exempt from approval (Table 2).

In quantitative research, randomized trials should describe the methods of randomization and concealment of allocation. In randomized and non-randomized controlled trials, the following information should be provided: the sequence of procedures; group differences at baseline; presence or absence of blinding and methods for blinding; similarities or differences in the treatment of groups; adequacy of follow-up or intention to treat; and the justification, validity, and reliability of the outcome measures, when that information is available. Describing the methods of selection to groups and group differences at baseline will assist readers in their evaluation of the potential for confounding. Any discrepancies or deviances from the researchers' intended methods of implementation of the study that might influence the outcomes should be identified (Gall et al. 2003). The CONSORT statement, for example, serves to improve the quality of reporting of trials, by providing a comprehensive method for organizing and communicating the Methods (Moher et al. 2001). In quantitative scholarship the reader should have sufficient understanding to evaluate the likely generalizability of the results garnered by the study.

In qualitative research, such as focus group interviews or ethnographic research, as in quantitative research, authors should identify the steps that were taken to reduce possible biases in the collection and interpretation of data (Inui & Frankel 1991; Giacomini & Cook 2000a, b). In particular, reasoning should be justified regarding how the participants were selected and how those participants might enable an understanding of a range of perspectives or social phenomena. Similarly, the instructions given to participants and precise methods for collecting and analyzing the data and the reliability of those methods should be provided. In qualitative scholarship, the reader should have adequate appreciation for the approach of the work in that the data gathering and analyses conform to the expectations of the field. It is understood in qualitative work (Giacomini & Cook 2000a, b) that the process of inquiry and the approach of the investigators may influence in discernible ways the results that are obtained. For these reasons, a rigorous qualitative study may be conducted, yet it may not be possible to assess the generalizability of the results, particularly in small studies.

In quantitative studies, the data analysis procedures should be identified and discussed in the light of the study question and the methods and measures used to answer the question. Because small sample sizes are common in educational research, a calculation of the power (Gall et al. 2003) of a study helps to determine the probability of finding an effect of a certain size, if such an effect truly exists. It should also be appreciated that when multiple outcome measures are used, the possibility of finding a significant difference when none truly exists increases. In this case, the level of significance might be adjusted to reduce this possibility (McGaghie & Crandall 2001). Qualitative studies require more description and that will be covered in other AMEE Guides.

Results

The "Results" section of a research paper should concisely portray the key findings. To be effective, the study findings should be clearly presented and ordered in relation to the research questions (Regehr 2001). The order of the narrative presentation should be clear and coherent; in other words, the Results should not be a mere "laundry list" of data and various statistical comparisons. In approaching the development of the Results, one helpful method is to order the findings in parallel with how the goals were identified in the Introduction and the findings discussed in the "Discussion" section (Regehr 2001). It is only necessary to publish the results that are of high quality and that relate most directly to the specific goals; it is not necessary to publish extraneous data (Louie et al. 2006). Tables or Figures can help provide the requisite detail and complex data or highlight key findings. The headings should be concise and summarize the contents of the Tables or Figures precisely, and the legends should inform the readers about any abbreviations that were used. At the same time, journals' printed space requirements often limit the use of Tables, and data from Tables should not be repeated in their entirety in the text. When Tables are used, the general strategy is to provide the requisite details of the data within them so that the text of the Results can emphasize the key findings without replicating all of the details.

Discussion

The "Discussion" section focuses on the main outcomes of the study first, establishing their context. In quantitative research, which this Guide is primarily about, these findings should be clearly stated and understood in relation to the rationale for the study and previously published findings of interest, possible alternative explanations (Crandall & McGaghie 2001), and implications for readers in their roles as educators, educational researchers, or administrators. One of the key goals of the Discussion is to link the aims and findings with relevant prior research. In this way, the Discussion links back to the Introduction to inform the reader about how these new findings are placed into an appropriate context, including the practical implications of the new findings in relation to prior work as well as any implications for future research. Conclusions must be clearly supported by the data. The findings also should be discussed in relation to the strengths and limitations of the data (e.g. a one-site study, small number of subjects, low response rate, and other contextual factors can

limit the generalizability of the findings), which is usually the work of the penultimate paragraph of the "Discussion" section. It might also be noted that statistically significant differences are not necessarily educationally meaningful. The final paragraph of the Discussion briefly reiterates the main findings and their implications for readers.

Fitting the sections together

The research question, methods, results, and discussion should all include the same elements. This is to say that the Methods should not include something that is not formulated as a research question and the Results should not include new information that is not described in the research question or Methods. The Discussion in turn should not include more or new information that is not part of the Methods, such as an additional description of the context of the study or of the intervention or of the circumstances of the control. Moreover, the Conclusions should follow precisely from the findings and not serve as an extension of the discussion or of the authors' own thinking.

Abstract and title

The writing of the Abstract is usually left until last because the Abstract summarizes the final version of the main body of the paper. It should provide information that is sufficiently complete, within required word limits, in order to accurately convey the main elements of each of the sections of the paper. Abstracts may be structured or narrative, dependent on the requirements of the target journal. The Title should be representative of the study, incite interest, and include keywords that are readily identifiable by search strategies. Because the Title and Abstract set a first impression for editors, reviewers, and readers, it is especially important to write these well. Researchers should therefore not scrimp on the time they dedicate to writing these sections, especially when tired at this last stage of manuscript preparation. After all, readers might only read the Abstract, and the Abstract can also be the basis for a decision as to whether to include a study in a systematic review.

Optimizing the writing

Most papers require a number of revisions and very careful attention to the editing before they are ready for submission (Table 4). For example, the Abstract should be checked to see that the requisite detail in it precisely matches what is contained in the paper. Similarly, information in the Tables should exactly match what was written in the text. The references should be individually checked for their accuracy and concordance with the target journal's requirements for citations. Definitions or terms should be strictly chosen and authors should stick to these rather than change the phrasing at different points in the text. The writing should be concise in using as few words as possible. In addition, citations in the text should be individually checked for the validity of comments ascribed to them. That is also to say that review article texts or abstracts should not be taken at face, and original sources should always be checked.

Table 4. The close-to-submitted version.

Revise and correct until the writing is optimized
Be sure that the findings are discussed in relation to the strengths and weaknesses of the methods for answering the specific research question
Check that the information in the tables and abstract exactly match what was said in the text
Check the validity of comments related to each of the cited references
Check to see that the references are accurate
Keep the style and requirements of the intended journal in mind
Have an experienced reviewer critically read it through
Re-read again with fresh eyes

Consideration ought to be given to the prior published reasons for potential acceptance or rejection of a manuscript during this process of revision and review. Top reasons for success include a clearly and succinctly written manuscript, practical and useful implications, and a discussion that adequately takes account of methodological limitations (Bordage 2001). Top reasons for rejection include incomplete or insufficiently described statistics, over-interpretation or under-interpretation of results, inaccurate or inconsistent data, and defective Tables or Figures (Bordage 2001). It is surprising how often one final read can reveal additional, even minor, issues for attention. Many capable authors suggest that one should permit a manuscript to sit for a week after it is "done" – a careful read-through with fresh eyes allows one to pick up on phrasing and subtleties that help produce the best possible empirical report. Authors who are writing in a different language for an international readership should have someone with expertise in that language read the paper through or seek help earlier to ensure that the editing and language are acceptable. There is also a difference between the English of the United States and the United Kingdom, and most text programs provide the opportunity for tailoring the writing accordingly. It might also help then to have an experienced reviewer critically read a close-to-final version before submission to catch any problems. Such assiduousness in preparation of the final manuscript, coupled with patience and perseverance in the revision processes, promotes the integrity of writing and editorial acceptance of the manuscript. It also protects against a negative bias by journal reviewers.

Choosing the journal

In effect, the choice of journal is a decision considered throughout all stages of writing and preparation of the manuscript. In turn, the choice of journal will have an important impact on the structure of the article and so the authors should read some articles from the preferred journal in order to see what the paper should look like. There are several, sometimes competing, factors in the choice of journal. These include the "goodness of fit" of a paper for the journal and the relevance of findings for the journal's readership, the prestige of the journal (usually judged by its Impact Factor, as discussed in the subsequent text), word limits of educational

research articles, and if known, acceptance or rejection rates, anticipated time to an editorial decision, and time between acceptance and publication. Although there are relatively few education research journals (e.g. Medical Teacher, Academic Medicine, Academic Psychiatry, Medical Education, Teaching and Learning in Education, Advances in Health Sciences Education, BMC Medical Education, International Journal of Medical Education, Journal of Graduate Medical Education, Journal of Continuing Education in Health Professions, and Journal of the International Association of Medical Science Educators), some specialty and general medical journals seek to publish education research. It is important to choose a journal that is interested in the context of the research. For example, some United Kingdom and United States journals may be less interested in research conducted outside of their jurisdictions. Knowledge of the range of options and the proclivity of specific journals for publishing on a topic of interest is helpful.

The Impact Factor, which is published annually by the Institute of Scientific Information in its Journal Citation Reports, is defined as the number of cites to articles in a particular (current) year divided by the number of substantive articles published over the two preceding years (Garfield 2006). Thus, an impact factor of 1 suggests that an "average article" published in two preceding years is cited on average once in a more recent year. Aiming high leaves open the possibility of acceptance in a relatively prestigious journal, but more likely invites rejection. It is often difficult to predict how reviewers and journals will respond, and at the cost of rejection and loss of time and hurt feelings, the reviews received at a relatively prestigious journal should enable the writing of an improved paper and enhanced success at the next journal. Authors should especially take care to reference all relevant articles from the journal to which they are submitting because the editors will likely know of relevant articles omitted from their own journal, and such omissions may lead to concerns about the adequacy of the authors' methods of searching. In addition, these citations might contribute to the journal's impact factor.

One strategy, underutilized in our experience as editors, is to contact the journal in advance of submission to ascertain its interest in a particular idea. Calling or e-mailing the editorial office for advice creates interest and perhaps generates a sense of responsibility and commitment by the editors to have the author become successful. Editors usually appreciate being consulted and given an opportunity to help authors.

Responding to editors and reviewers

Few papers become accepted without being revised. An invitation to revise and resubmit is a very good result because it is uncommon for such manuscripts to subsequently become rejected. When editors signal that they will be willing to entertain a revision – without specifically inviting the revision – the possibility of future rejection is higher, but this opportunity is still positive for the author and should be pursued.

Comprehensive and constructive reviews are a gift (Roberts et al. 2004) and warrant the utmost respect in turn. Reviewers who take time to develop a comprehensive set of suggestions enhance the quality of the final written product, as well as assist the editors in forming a decision concerning publication, ensure scientific rigor, and foster advancement of the field (Roberts et al. 2004). Moreover, reviewers truly try to help and some are experts in the topic of study. To this end, authors should respond positively, non-defensively, and in detail to every reviewer's comment in turn (Table 5). The authors should make the job easy for reviewers and editors by saying what precisely was changed in the text as opposed to just indicating that the text was revised while also avoiding long explanations.

Table 5. Responding to reviewers as consultants and colleagues.

Anticipate that reviewers will provide many suggestions for improvement
Respond positively, with thanks, and non-defensively to every comment in turn
Provide thoughtful, well-argued, and reasoned responses to important or major recommendations
Balance conflicting recommendations
Make changes in line with the reviewers' suggestions at every opportunity

On occasion, an author may not agree with a comment by a reviewer. Reviewers can also make mistakes, and some of their recommendations (such as to obtain a larger sample size) may not be achievable. Frequently, reviewers also will proffer contradictory advice to an author. A thoughtful, well-argued, and reasoned response should facilitate a favorable decision by the editors in this context of expert disagreement. Moreover, being courteous and thankful can count as to whether a journal will accept a manuscript (Guyatt & Brian Haynes 2006) and is a professional responsibility. The guiding principle here is to approach the reviewer as a consultant (Provenzale 2010) or colleague (Roberts et al. 2004) rather than as an adversary. It is important to remember that reviewers do not make publication decisions; editors do. Editors will weigh the insights of the reviewers alongside their own views, plus issues that extend beyond the specific manuscript. For example, the editor may know – although the author and reviewers may not – that an entire set of already-accepted papers on a similar topic are "in the queue" for publication in the very near future. The editors, thus, may attribute more or less weight to the newly submitted manuscript in accordance with how it fits into this set of papers. In sum, the author's primary relationship should be with the editor or editors who are making the difficult decision about whether the piece should be published and how it may be improved. Moreover, the author should understand the nuances of correspondence with editors and the kinds of factors that editors must consider, both intrinsic and extrinsic to the submitted manuscript.

Conclusions

Writing for success is a disciplined and systematic process following prescribed steps. We have emphasized how, though hard work, writing should be wonderfully rewarding and fun.

It is a pinnacle of academic success to see one's research in print and available for others to read and appreciate. Our own starting point was that we wanted readers to get started and to succeed in their quest to become productive educational researchers. The strategies that we have presented here should facilitate success in the academic processes of writing for publication and promote educational research.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

Notes on contributors

JOHN H. COVERDALE, MB ChB, MD, M. Ed, FRANZCP, is a Professor of Psychiatry and of Medical Ethics at the Menninger Department of Psychiatry, Baylor College of Medicine, Houston, TX.

LAURA WEISS ROBERTS, MD, MA, is Chairman and Katharine Dexter McCormick and Stanley McCormick Memorial Professor, Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA.

RICHARD BALON, MD, is a Professor, Department of Psychiatry and Behavioral Neurosciences and Department of Anesthesiology, Wayne State University School of Medicine, Detroit, MI.

EUGENE V. BERESIN, MD, MA, is a Professor, Department of Psychiatry, Harvard University, and Director of Child and Adolescent Psychiatry Residency Training, Massachusetts General Hospital and McLean Hospital, Boston, MA.

References

Boice R. 1990. Professors as writers. A self-help guide to productive writing. Stillwater, OK: New Forums Press Inc.

Bordage G. 2001. Reasons reviewers reject and accept manuscripts: The strengths and weaknesses in medical education reports. Acad Med 76(9):889–896.

Bunniss S, Kelly DR. 2010. Research paradigms in medical education research. Med Educ 44(4):358–366.

Coverdale J, Louie A, Roberts LW. 2005. Getting started in educational research. Acad Psychiatry 29:14–18.

Coverdale J, Roberts L, Louie A, Beresin E. 2006. Writing the methods. Acad Psychiatry 30(5):361–364.

Crandall SJ, McGaghie WC. 2001. Discussion and conclusion: Interpretation. Acad Med 76:942–944.

Fraenkel JR, Wallen NE. 2006. How to design and evaluate research in education. Boston, MA: McGraw-Hill.

Gall MD, Gall JP, Borg WR. 2003. Educational research: An introduction. White Plains, NY: Longman.

Garfield E. 2006. The history and meaning of the journal impact factor. JAMA 295(1):90–93.

Giacomini MK, Cook DJ. 2000a. Users' guides to the medical literature: XXIII. Qualitative research in health care A. Are the results of the study valid? Evidence-Based Medicine Working Group. JAMA 284(3):357–362.

Giacomini MK, Cook DJ. 2000b. Users' guides to the medical literature: XXIII. Qualitative research in health care B. What are the results and how do they help me care for my patients? Evidence-Based Medicine Working Group. JAMA 284(4):478–482.

Guyatt GH, Brian Haynes R. 2006. Preparing reports for publication and responding to reviewers' comments. J Clin Epidemiol 59(9):900–906.

Haig A, Dozier M. 2003a. BEME Guide No. 3: Systematic searching for evidence in medical education, part 2: Constructing searches. Med Teach 25:463–484.

Haig A, Dozier M. 2003b. BEME Guide No. 3: Systematic searching for evidence in medical education, Part I: Sources of information. Med Teach 25:352–363.

Hammick M, Dornan T, Steinert Y. 2010. Conducting a best evidence systematic review. Part 1: From idea to data coding. BEME Guide No. 13. Med Teach 32(1):3–15.

Harden RM, Grant J, Buckley G, Hart IR. 2000. Best evidence medical education. Adv Health Sci Educ: Theory and Pract 5(1):71–90.

Hart IR, Harden RM. 2000. Best evidence medical education (BEME): A plan for action. Med Teach 22(2):131–135.

Horn C, Plazas Snyder B, Coverdale JH, Louie AK, Roberts LW. 2009. Educational research questions and study design. Acad Psychiatry 33(3):261–267.

Hoschl C, Fialova L, Moller H-J, Balon R. 2012. European Psychiatric Association guidance on the conflicts of interest. Eur Psychiatry 27(2):142–146, (author erratum, 2012, Eur Psychiatry 27:308).

International Committee of Medical Journal Editors. 2009. Uniform requirements for manuscripts submitted to biomedical journals: Ethical considerations in the conduct and reporting of research: Authorship and contributorship. [Accessed 29 March 2012] Available from http://www.icmje.org/ethical_1author.html

Inui TS, Frankel RM. 1991. Evaluating the quality of qualitative research: A proposal pro tem. J Gen Intern Med 6(5):485–486.

Louie A, Coverdale J, Edenharder K, Roberts LW. 2006. Publishing a manuscript. In: Roberts LW, Hilty DM, editors. Handbook of career development in academic psychiatry and behavioral sciences. Washington, DC: American Psychiatric Publishing. pp 241–250.

Maggio LA, Tannery NH, Kanter SL. 2011. Reproducibility of literature search reporting in medical education reviews. Acad Med 86:1040–1054.

McGaghie WC. 2009. Scholarship, publication, and career advancement in health professions education: AMEE Guide No. 43. AMEE Guide 31(7):574–590.

McGaghie WC, Bordage G, Crandall S, Pangara L. 2001. Method: Research design. Acad Med 76:929–931.

McGaghie WC, Crandall S. 2001. Data analysis and statistics. Acad Med 76:936–938.

Moher D, Schulz KF, Altman D, (Consolidated Standards of Reporting Trials) CONSORT Group. 2001. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. JAMA 285(15):1987–1991.

Pangaro L, McGaghie WC. 2001. Relevance. Acad Med 76:927–928.

Prideaux D. 2002. Researching the outcomes of educational interventions: A matter of design. RTCs have important limitations in evaluating educational interventions. Br Med J 324(7330):126–127.

Provenzale JM. 2010. Revising a manuscript: Ten principles to guide success for publication. AJR Am J Roentgenol 195(6):W382–W387.

Reed D, Price EG, Windish DM, Wright SM, Gozu A, Hsu EB, Beach MC, Kern D, Bass EB. 2005. Challenges in systematic reviews of educational intervention studies. Ann Intern Med 142(12 Pt 2):1080–1089.

Regehr G. 2001. Presentation of results. Acad Med 76:940–942.

Ringsted C, Hodges B, Scherpbier A. 2011. The research compass: An introduction to research in medical education, AMEE Guide No. 56 ed. Dundee, Scotland: Association for Medical Education in Europe.

Roberts LW, Coverdale J, Edenharder K, Louie A. 2004. How to review a manuscript: A "down to earth" approach. Acad Psychiatry 28:81–87.

Roberts LW, Geppert C, Connor R, Nguyen K, Warner TD. 2001. An invitation for medical educators to focus on ethical and policy issues in research and scholarly practice. Acad Med 76(9):876–885.

Roberts LW, Geppert CM, Coverdale J, Louie A, Edenharder K. 2005. Ethical and regulatory considerations in educational research. Acad Psychiatry 29(1):1–5.

Sackett DL, Wennberg JE. 1997. Choosing the best research design for each question. Br Med J 315(7123):1636.

Shaw D. 2011. The ICMJE's definition of authorship is illogical and unethical. Br Med J 343:d7192.

Sullivan GM. 2011. A primer on the validity of assessment instruments. J Grad Med Educ 3:119–120.

Medical Teacher

ISSN: 0142-159X (Print) 1466-187X (Online) Journal homepage: https://www.tandfonline.com/loi/imte20

Twelve tips for getting your manuscript published

David A. Cook

To cite this article: David A. Cook (2016) Twelve tips for getting your manuscript published, Medical Teacher, 38:1, 41-50, DOI: 10.3109/0142159X.2015.1074989

To link to this article: https://doi.org/10.3109/0142159X.2015.1074989

2016, 38: 41–50 MEDICAL TEACHER

TWELVE TIPS

Twelve tips for getting your manuscript published

DAVID A. COOK[1,2]

[1]Mayo Clinic Online Learning, USA, [2]Mayo Clinic College of Medicine, USA

Abstract

The author shares twelve practical tips on how to navigate the process of getting a manuscript published. These tips, which apply to all fields of academic writing, advise that during the initial preparation phase authors should: (1) plan early to get it out the door; (2) address authorship and writing group expectations up front; (3) maintain control of the writing; (4) ensure complete reporting; (5) use electronic reference management software; (6) polish carefully before they submit; (7) select the right journal; and (8) follow journal instructions precisely. Rejection after the first submission is likely, and when this occurs authors should (9) get it back out the door quickly, but first (10) take seriously all reviewer and editor suggestions. Finally, when the invitation comes to revise and resubmit, authors should (11) respond carefully to every reviewer suggestion, even if they disagree, and (12) get input from others as they revise. The author also shares detailed suggestions on the creation of effective tables and figures, and on how to respond to reviewer critiques.

Introduction

Scholarly writing and research reporting are increasingly common in all areas of medicine, not least in health professions education. The rising number of advanced training programs (Tekian et al. 2014) suggests that soon even more education scholars will enter the field with training in research and an expectation to publish. Much has been written about how to plan and conduct a research study or scholarly project (Bordage & Dawson 2003; Beckman & Cook 2007; Ringsted et al. 2011), the elements of research reporting in general (Bordage 1989; Parsell & Bligh 1999; Coverdale et al. 2013), and required reporting elements for specific study types (von Elm et al. 2007; Moher et al. 2009; Schulz et al. 2010; O'Brien et al. 2014). Less has been written about how to navigate the publishing process itself.

The purpose of the present article is to share twelve practical tips (Table 1) on how to successfully navigate the process of getting a manuscript published in a peer-reviewed journal. While these tips reflect the personal approach of a medical education researcher, I believe they apply broadly to all domains of academic writing. I intend to complement rather than repeat others' suggestions regarding effective writing and the editorial process (Bordage 1989; Gopen & Swan 1990; Huth 1999; Parsell & Bligh 1999; Bordage 2001; McGaghie 2009; Coverdale et al. 2013; Azer et al. 2014).

My immediate target audience is lead authors (first authors and corresponding authors), although others including coauthors, editors, and reviewers will also find this useful. A junior first author would ideally implement these tips in collaboration with an experienced mentor. I will assume that readers are familiar with the process of journal submission and peer review, that essential steps such as goal clarification, study design and execution, and reflective critique (Glassick 2000) have already occurred, and that the immediate challenge is to publish.

Getting the manuscript ready

Tip 1

Plan early to get it out the door

Perhaps the greatest challenge facing novice writers is the imperative to overcome writing inertia. To paraphrase Newton's first law, a manuscript in preparation will remain in preparation indefinitely unless acted upon by a motivated author. It will not get published as long as it sits on your desk! You need to get the manuscript off of your to-do list and onto someone else's (namely, the editor's or reviewer's). To make this happen, you need to write regularly, set ambitious goals, and refine the manuscript in stages.

Some people believe that to write effectively they need blocks of time (several days in a row, with several hours each day). This works in some cases, but all too often those big blocks of time fail to materialize, or get consumed by other tasks. It can also be cognitively overwhelming to write episodically, to the point that some writers come to dread their writing day. Research suggests that those who write daily, even for short periods (e.g., 15 minutes per day) are substantially more productive than those who postpone writing until they have a large chunk of time (Boice 1989). Another drawback to allowing time to lapse between writing sessions is the loss in continuity and efficiency – it

Correspondence: David A. Cook, MD, MHPE, Division of General Internal Medicine, Mayo Clinic College of Medicine, Mayo 17-27W, 200 First Street SW, Rochester, MN 55905, USA. Tel: +1 507 266 4156. Fax: +1 507 284 5370. E-mail: cook.david33@mayo.edu

ISSN 0142-159X print/ISSN 1466-187X online/16/010041–10 © 2015 Taylor & Francis
DOI: 10.3109/0142159X.2015.1074989

Table 1. Twelve tips for getting your manuscript published.

GETTING THE MANUSCRIPT READY
1. Plan early to get it out the door
2. Address authorship and writing group expectations up front
3. Maintain control of the writing
4. Ensure complete reporting
5. Use electronic reference management software
6. Polish carefully before you submit
7. Select the right journal
8. Follow journal instructions precisely
WHEN YOU ARE REJECTED
9. Get it back out the door quickly!
10. Take seriously all reviewer and editor suggestions
WHEN YOU ARE INVITED TO REVISE
11. Respond carefully to every suggestion, even if you disagree
12. Get input from others as you revise
9 (revisited). Get it back out the door quickly!

takes time to remember where you left off and what issues needed to be addressed. Occasionally, if you are having trouble articulating a particularly difficult idea, it helps to step away for a period and then return with fresh eyes. But in general it helps to write regularly – ideally daily, even if for only short periods.

It helps to set ambitious yet realistic deadlines for yourself and your coauthors. Deadlines should push you to work hard and quickly; conservative or comfortable deadlines will not accomplish this purpose.

As Strunk and White noted, "Revising is part of writing. Few writers are so expert that they can produce what they are after on the first try." (Strunk & White 2000, p. 72). Using outlines and multiple progressively refined drafts will optimize organization and minimize the amount of wasted effort (e.g., polishing text that is subsequently revised or deleted). It is far easier to visualize and manipulate the flow of logic using short bullet-points than full sentences and paragraphs. Thus, I agree with McGaghie (2009) in advocating the liberal use of outlines – especially when drafting text that allows flexible organization (e.g., a research manuscript's Introduction and Discussion, or a non-research scholarly article). Once the outline is complete, converting bullet-points into sentences and paragraphs is straightforward. For this first "rough idea draft" I try to clearly articulate my thoughts but I do not attempt to fine-tune each sentence. In subsequent drafts I iteratively refine my thinking, shorten the text, clarify meaning, and adjust individual words. Different parts of the manuscript may progress at different rates (e.g., I usually do not write the Discussion until the other sections are nearly complete). At each stage (outline, rough idea draft, first final draft, and each fine-tuning draft) I enlist the aid of my coauthors with explicit instructions and objectives for that stage (e.g., "Focus on broad-stroke ideas today; do not worry about fine-tuning, and ignore the Discussion"). Using outlines can help in later stages as well, such as when shortening the manuscript or making major revisions in response to reviewer comments. Converting a fully-written Introduction and Discussion back into outline form allows one to easily find and eliminate redundancy, rearrange and merge paragraphs, and focus on the central message.

Tip 2

Address authorship and writing group expectations up front

Publishing a manuscript is usually a team effort. Decisions about authorship – including the order of authorship and the expectations and responsibilities of each coauthor – should be made as early as possible. A full discussion of the issues surrounding qualifications for authorship (Wislar et al. 2011; Bordage et al. 2015) is beyond the scope of this article, but current guidelines indicate that authors must make *substantial* contributions to the conception, design, or data analysis and interpretation, and *critically* revise the manuscript for *important* intellectual content (International Committee of Medical Journal Editors 2013). Lesser contributions, including administrative, political, and funding support and simple data analysis, should be recognized with an acknowledgement.

It helps to set clear expectations for each author before writing begins. This would ideally include a rough timeline and deadline (see Tip #1), the expected turnaround time for each draft, and any specific contributions (e.g., manuscript sections). First, second, and senior author positions should be finalized after these initial discussions.

Tip 3

Maintain control of the writing

One person – usually the first author – must be ultimately responsible for the final manuscript and all changes made during editing. Thus, when I am the lead author, I maintain a single copy of a master document and edit this document to incorporate suggestions from coauthors.

Collaboratively editing documents online (e.g., using Google Docs or Dropbox) is increasingly popular, and is useful for many activities in education, research, and administration. Yet, I have found collaborative editing to be counterproductive when writing for publication because inconsistencies, redundancies, and omissions inevitably creep in when multiple authors jointly contribute untracked edits to a single, shared document.

Using tracked changes avoids this, but simply accepting changes suggested by coauthors is not advisable. First, coauthor-suggested edits frequently contain small errors in spelling, punctuation, or grammar (often encouraged and obscured by the tracked changes format). Second, most suggestions require judgment and editing to prioritize the use of limited space, reconcile conflicts, avoid redundancy, trim words, integrate with other ideas, and avoid inaccuracies.

To avoid these challenges, I distribute a manuscript copy to all coauthors via email or a file sharing tool such as Dropbox, and collect suggestions as tracked changes and embedded comments (or, occasionally, as written comments on paper). I then integrate these responses and make needed edits to the *separate* master document.

Tip 4

Ensure complete reporting

It is often difficult for investigators, even as a team, to identify all of the information needed or expected by readers. For this reason, reporting guidelines have been developed for most common study designs including randomized trials (Schulz et al. 2010) (including an extension for non-pharmacologic studies, Boutron et al. 2008), non-randomized trials (von Elm et al. 2007), assessment studies (Bossuyt et al. 2003), qualitative research (O'Brien et al. 2014), and systematic reviews (Moher et al. 2009). Additional reporting standards can be found at www.equator-network.org. Reporting guidelines will not salvage a poorly planned or poorly executed study, but they can certainly improve the odds of success for a well-done study (especially for authors with less experience).

Existing guidelines provide detailed suggestions for the Abstract, Methods, and Results, but less direction for the Title, Introduction, and Discussion. I will, therefore, offer brief suggestions for the latter three. First, the title is the shortest possible abstract (Cook et al. 2007; Bordage et al. 2015). It is almost always the first thing a potential reader reads, and it may be the last if it does not catch his or her attention. Rather than being written as a last-minute afterthought, the title should be the product of thoughtful effort by all authors. I typically create a list of key words reflecting the manuscript's central message, use permutations of these to generate over a dozen tentative titles, and then solicit impressions from coauthors and other colleagues before selecting the best title.

The Introduction sets the stage for all that follows (Beckman & Cook 2007). It starts with a very broad problem and then focuses this problem through the lens of a *conceptual framework* or theory (Bordage 2009) and a summary of prior work (*literature review*) into a focused *problem statement*. The problem statement clearly identifies one important aspect of the broad problem that remains incompletely understood, and highlights how a better understanding of this issue would advance the field as a whole. The problem statement is immediately followed by a statement of study intent – a research question, hypothesis, goal, or purpose – that explains how the subsequently described study will fill that knowledge gap and thereby accomplish the needed advance. For example, "Although these studies suggest that adding more questions does not necessarily enhance learning, the interpretations are confounded by the simultaneous variation in the types of questions, which in turn suggests the need for further research [*problem statement*]. . . .We therefore sought to answer the question: Does varying the number of self-assessment questions affect knowledge outcomes in [Web-based learning] for medical residents? [*research question*]." (Cook et al. 2014) These four elements – conceptual framework, literature review, problem statement, and statement of study intent – combine to not only articulate a question, but also to convince the reader that the answer remains unknown and that finding the answer is important (McGaghie et al. 2001).

The Discussion section of most research reports is far longer than needed, and thereby consumes valuable space that could be used in other sections. "Brevity is the soul of wit," and writing a focused yet informative Discussion is an oft-neglected part of manuscript preparation (Docherty & Smith 1999; Clarke et al. 2002). I have found three practices that help me stay focused and succinct. First, I imagine that no one except the editors and reviewers will read the entire Discussion. Most readers read the first paragraph and many read the last, but I suspect that everything in between has a limited audience. As such, I focus my efforts on crafting a concise summary of the study findings (first paragraph), highlighting well-supported implications (last paragraphs), and addressing issues that would concern editors and reviewers. Second, the Discussion section should focus on the study's objective findings and immediate, justified implications. Limited speculation on the interpretation and application of findings is appropriate, but this is not the place to editorialize on issues only tangentially related to the study, even if germane to the study topic. Third, to operationalize the first two suggestions I organize the Discussion into four sections, with subheadings for the latter three:

(1) *Summary* (no heading): One or two paragraphs that succinctly and objectively summarize the key findings without further elaboration (i.e., no citations to other sources, no interpretation).
(2) *Limitations*: More than just a *list* of limitations; an examination of how the study scope and design might *influence* the interpretation and application of findings.
(3) *Integration with prior work*: One to three paragraphs linking my findings with both theory and empiric research.
(4) *Implications for practice and research*: Three to five paragraphs highlighting what readers can do differently now that they know these results. These implications should derive directly from the findings of the present study as interpreted in light of the limitations and integration, and could anticipate effects on theory, practice, or future research. The "Implications" replace the need for a separate Conclusion section; the implications *are* the conclusion and should leave readers with a clear sense of how this study will influence their practice.

Tip 5

Use electronic reference management software

Writing a manuscript without the support of electronic reference management software is like building a bookcase using a hand drill instead of a power drill – sure, you can do it, but it is a lot easier with a power tool! With reference management software you create a list (library) of articles and books you wish to cite and then place a placeholder (field code) representing each reference in desired location(s) in your manuscript. The software formats these placeholders into the reference style required by the selected journal. Advantages include the abilities to re-use references in a new manuscript without re-typing or copy-and-pasting, insert a new reference near the top of the manuscript without manually renumbering the subsequent references, share reference libraries with colleagues, make personal notes about a reference, and adjust the format

style (e.g., to meet another journal's requirements) with the click of a button. Wikipedia currently lists over 30 software packages including purchased products such as EndNote and Reference Manager, and numerous free options such as Mendeley and Zotero (see http://en.wikipedia.org/wiki/Comparison_of_reference_management_software).

Beyond urging you to "just do it" I suggest the following power-user tips:

- Learn some skills beyond the rudimentary insert and format functions. Useful tasks include tweaking the formatting styles (in my experience, journal-specific templates don't match perfectly with actual journal requirements), adding text such as page numbers to the in-text citation, finding and eliminating duplicate entries, and exporting libraries to share with a colleague or import into another program such as Excel.
- Use PubMed and journal websites to add articles to your library rather than entering these by hand.
- Use self-entered "Notes" to organize, sort, and search for references within your library.
- Copy and share references or entire libraries with others. CiteULike is a free Internet service dedicated to promoting such sharing.
- Consider turning off the instant formatting feature offered by several programs (e.g., EndNote's "Cite While You Write"). While useful if you are the sole author, it gets messy if collaborators delete or move text or if you track changes.
- Correct any errors in the reference list at their source (usually the library entry or journal-specific template) rather than making edits directly to the reference list in your manuscript. If the same reference appears twice in the reference list there are probably two library entries for that reference. Missing information or other problems with the reference list format could indicate an error in that reference's library entry or in the journal-specific formatting template.

As a final step before submission, you will need to convert the document with dynamic placeholders to a placeholder-free, text-only document ("convert to plain text" in EndNote). Save a separate copy of the final, unformatted placeholder version to use as a starting point when making revisions.

Tip 6

Polish carefully before you submit

Editors will help you tailor your message to their audience, but you should not expect them to correct poor grammar or typographical errors. Errors or inconsistencies in writing or formatting impair understanding, suggest (rightly or wrongly) sloppiness in the scientific rigor, and will require extra effort from the editorial team – none of which will improve your chances of manuscript acceptance! The notion of polishing need not conflict with the tip to get the manuscript out the door. Polishing does not necessarily add a lot of time, but it does require deliberate planning and effort. The following tips will prevent common errors:

- Use a consistent font style for each level of subheading (e.g., bold and all capitals for main section headings [level 1], bold and first letter capitalized for level 2 subheadings, etc.). Word processing "styles" can help maintain this consistency. Some journals have specific subheading format requirements.
- Use abbreviations and acronyms sparingly. Many abbreviations commonly used in a narrow specialty, and all abbreviations invented by the research team, will be unfamiliar to most readers. Define such abbreviations, or better yet eliminate them altogether.
- Review the title of each table and figure to ensure that it accurately and completely describes the information contained therein, and carefully proofread table/figure footnotes and legends. Ensure that footnotes define all abbreviations, clarify any apparent inconsistencies (e.g., percentages that do not sum to 100%, or response rates that vary across table cells), and explain key analyses. Table 2 contains additional tips for constructing tables and figures.
- Confirm that data and other details in the abstract match those in the main text. Similarly verify the match between tables and figures and the main text and abstract.
- Verify the format and accuracy of each reference, including adherence to the journal's formatting style.
- Remove all comments and resolve all tracked changes.

I always read the entire final draft, including abstract and tables, *out loud* because I have found this helps me identify awkward sentences and grammatical errors. It is also helpful to ask a non-coauthor colleague to read the manuscript to identify awkward sentences, logical inconsistencies, missing information, and simple errors. For those writing in a non-native language, a skilled native speaker should always proofread the manuscript (note that just because someone speaks English does not mean he or she is a good writer or good proofreader). McGaghie (2009) enumerated several additional suggestions for those writing in a non-native language.

Tip 7

Select the right journal

You will naturally want your work published in the best possible journal, but the best journal is not necessarily the one with the highest impact factor or the greatest prestige (Azer et al. 2014). In selecting a journal, I first consider the journal's readership: Is this journal followed by the people who will read, understand, apply, and cite my work? If not, then regardless of the journal's prestige my work may not get the visibility it deserves. Journals usually publish "Instructions for authors" that describe their target audience, scope, and preferences for specific topics and study types. Perusing back issues of a potential journal can further clarify the type and topic of articles published. Experienced colleagues can also offer field-specific tips.

Second, I consider the quality and potential impact of my own work. Let us face it – some articles are stronger than

Table 2. Tips for effective tables and figures.

Tips for effective tables

- Consider whether the table is really necessary
  - The ''information density'' (information per square inch) should be greater than just putting this information in the main text; data that would require fewer than 2 columns and rows should be presented in the text rather than a table
  - The information should complement rather than duplicate information available elsewhere in the manuscript
- Create a brief but explanatory title
- Carefully select the data (not all data are equally important) and thoughtfully organize the data to communicate a clear message
- Ensure that the table can be easily interpreted without reference to the main text
- Ensure that the table accurately reflects both the data and the intended message
  - Organize and order information to facilitate accurate and intuitive interpretation; emphasize relationships (differences, patterns, interactions) that most clearly communicate the central message
  - Provide sufficient information to allow interpretation in context (e.g., historical data, comparison group, reference range); clearly indicate factors (e.g., design changes or historical events) that might influence data
- Keep the table simple, clean, and free of extraneous detail
- Distinguish counts (tallies of items or events) and measurements (readings or scores on a continuous or semi-continuous scale)
  - Counts can be reported in absolute numbers (e.g., the number of events) or proportions (e.g., the number of a given response from a larger number of observations); proportional counts are typically summarized with numerator, denominator, and percentage (ideally all three); absolute counts benefit from a reference range or explanation of context to enable meaningful interpretation (e.g., is 247 website visits per month a lot or a little?)
  - Measurements are typically summarized with an estimate of central tendency (e.g., mean, median), an estimate of variance (e.g., standard deviation, interquartile range), and the number of observations; again a reference range or context is helpful
  - A separate column is usually used for count vs measurement data, although mixed-data columns can be created if changes in cell contents are clearly indicated (e.g., with row labels or footnotes)
  - The total number of observations can be reported in the column heading (if consistent for that column)
- Verify the accuracy of all data
- Create the table using your word processor Table tool rather than using tabs and hard returns; each piece of data should be contained in its own cell (this facilitates the publication process)
- Create a column heading for every column; use row labels as needed for additional clarity
- Identify probability level values (*p* values) in the table cells or using footnotes
- Resolve or explain all ambiguities and perceived incongruities (e.g., changes in the number of responses leading to ''shifting denominators'')
- Explain all abbreviations; special use of italics, parentheses, and dashes; special symbols; and empty cells
  - Example: a cell entry ''46/50 (92%)'' should have a column heading or footnote explaining that this means ''No./N (%)"
  - Keep abbreviations consistent with the main text; define all abbreviations using footnotes (so that the table can stand alone)
- Consistently apply formatting
  - Within a table: consistently use emphasis (bold, italics), line spacing, abbreviations, and column and row labels
  - Across tables: apply similar formatting for all tables in the manuscript
- Follow all journal-specific instructions on table creation
  - Use footnote symbols conforming to journal instructions (e.g., *, †, ‡, § or [a], [b], [c], [d])
  - Look at recent back issues of the journal for examples
- If the table or its data are from another source, cite the original source
- Refer to the table in the text
- Place the table in the manuscript according to journal instructions (i.e., appended at the end, embedded in the main text, or submitted in a separate document)

For additional information on table preparation, see the Purdue Online Writing Lab (owl.english.purdue.edu) and Wainer (1984), Morgan (1985), and Schriger et al. (2006).

Tips for effective figures

- Consider whether the figure is really necessary
  - The ''information density'' (information per square inch) should be greater than just putting this information in the main text or in a table. Note that the effective visual display of information can efficiently communicate key relationships, but often at the sacrifice of potentially useful information (e.g., specific numeric results); such trade-offs should be carefully considered and minimized as much as possible
  - The information should complement rather than duplicate information available elsewhere in the manuscript
- Create a brief but explanatory legend or caption
- Carefully select the data (not all data are equally important) and thoughtfully organize the data to communicate a clear message
- Follow established guidelines and norms for specific figure types (e.g. participant flow diagram for experimental studies (Schulz et al., 2010), or study flow diagram for systematic reviews (Moher et al., 2009))
- Ensure that the figure can be easily interpreted without reference to the main text
- Ensure that the visual metaphor of the figure accurately reflects both the data and the intended message; avoid perceptual distortions
  - The visual representation of numbers should be directly proportional to the numerical quantity
  - Ensure that that all scales (e.g., *x* and *y* axes) are consistently used and correctly proportioned; disproportionate scaling (e.g., scales that vary irregularly along the axis, scales that do not start at 0, and nonlinear scales) can be misleading, and should be used with restraint and always made explicit (verbally or visually) in the legend or in the figure itself
  - Organize and order information to facilitate accurate and intuitive interpretation; emphasize relationships (differences, patterns, interactions) that most clearly communicate the central message
  - Provide sufficient information to allow interpretation in context (e.g., historical data, comparison group, reference range); clearly indicate factors (e.g., design changes or historical events) that might influence data
- Keep the figure simple, clean, and free of extraneous detail; avoid using special effects (e.g. 3-D effects, shading, and layered text)
  - The ''data-to-ink ratio'' (Tufte, 2001) provides one approximation of the clarity of presentation (high ink [which leads to a lower ratio] suggests unnecessary clutter that can obscure the message)
- Verify that all data are accurate and are plotted accurately
- Use lettering that is dark enough and large enough to read, and compatible in size with the rest of the figure
- Explain all line, symbol, and color styles; text emphasis (bold, italics); and abbreviations in the legend or caption
  - Keep abbreviations consistent with the main text; define all abbreviations (so that the figure can stand alone)
- Consider using grayscale (or black-and-white) rather than color

(*continued*)

- Color should only be used if the print version of the figure will appear in color
- Even if the print figure is published in color, remember that it will usually be black-and-white if printed by a reader
- Varying the symbol (triangle, circle, square; solid or open) or line style (solid, dashed, dotted; light or heavy) may be more clear than varying colors

• Consistently apply formatting
- Within a figure: consistently use line, symbol, and color styles; text emphasis (bold, italics); and abbreviations
- Across figures: prepare parallel (i.e., adjacent) or equally important figures according to the same scale; apply similar formatting for all figures in the manuscript

• Follow all journal-specific instructions on figure creation, including figure resolution and file format (e.g., JPEG, TIFF, PNG)
• Refer to the figure in the text
• Place the figure in the manuscript according to journal instructions (i.e., appended at the end, embedded in the main text, or submitted as a separate file)

For additional information on figure preparation and the visual display of data, see the Purdue Online Writing Lab (owl.english.purdue.edu) and Tufte (2001), Wainer (1984), and Schriger and Cooper (2001).

others, and will merit publication in a higher-impact journal. You do not want to sell short your work, but repeatedly aiming too high will only result in repeated rejections. Each submission–rejection cycle delays your paper's appearance in print, reflects wasted effort on your part (and the journal's), requires you to re-immerse yourself in the details of a project from which you have moved on, and generates increasing frustration and discouragement for you and your coauthors. An honest appraisal of your work's merit, which may require input from a non-coauthor colleague, will reduce both time and frustration. I will occasionally aim one step higher than my realistic estimation for my first submission – hoping to catch a lucky break. But if that fails, I immediately shift to a journal that I believe will be a reasonable match.

Third, I consider the impact and prestige of the potential journals. This is notoriously difficult to define. Quantitative metrics (Rizkallah & Sin 2010; Carpenter et al. 2014) such as the journal impact factor, immediacy index, eigenfactor, and article influence score all attempt to estimate impact, but all have notable deficiencies and norms are field-specific (i.e., the same impact factor value might be considered low in one field and high in another). Once again, experienced colleagues can help navigate journal prestige.

Other considerations include the time from submission to acceptance and from acceptance to publication, publication fees, recognition of the journal by local peers, and restrictions on words or figures. Some journals have several-month backlogs, while others will publish articles "online ahead of print" within weeks of acceptance. Local opinions about journal prestige should not be ignored, but remember that people may change their opinion when presented with new information (e.g., data supporting the prestige of a journal within a field), and moreover if your work is discovered and cited by others this will typically carry more weight than the journal in which it was published.

Tip 8

Follow journal instructions precisely

As a journal editor and reviewer, I am frequently dismayed at authors' disregard for journal instructions concerning topics of interest, article type, manuscript length and required elements, abstract structure and length, reference citation format, and more. Failing to follow instructions creates an unfavorable first impression with the editorial team, and in some cases may trigger automatic rejection. Thus, the last thing I do before submitting is re-read the author instructions and verify that I have fully complied. The only exception is that sometimes I will exceed the maximum number of references (although I have occasionally had to shorten that list during revisions). Most journals limit the number of words, tables, and figures. In general, do not exceed stated limits without express permission from the editorial staff (and document this in your cover letter).

Most journals still require a cover letter. However, this can be very brief and focused. A concise but complete cover letter might consist of three short paragraphs – one each for authorship, article summary, and potential reviewers. The only required element in most letters is a brief statement about authorship, conflicts of interest, and prior publication. The two to four sentence summary paragraph should not recapitulate the abstract (which the Editor is going to read soon enough), but rather should focus on the importance of this topic, the anticipated impact of these findings on the field, and why this manuscript is a good match for the chosen journal. It is usually helpful to suggest some potential reviewers, even if not required. Research suggests little difference in the quality of review from author-suggested versus editor-suggested reviewers (Schroter et al. 2006), and most editors welcome suggestions because it saves them time and because you will probably be more familiar with subject experts than they.

When you are rejected (because you will be)

By following the first eight tips, you successfully got your manuscript out the door. Unfortunately, since most journals accept fewer than 20% of the manuscripts they receive, your manuscript will most likely be rejected. You will naturally feel discouraged when you get that rejection letter. However, rejection is simply part of the publishing game. Most of the papers I have published were rejected by the first journal, and several had two or three rejections before finally finding a home. Yet, I have seen colleagues hesitate to submit their manuscript to another journal because they feel discouraged after the first rejection. You should *never* give up after the first rejection! You have already invested substantial effort in getting the manuscript to this stage; that effort is wasted if you stop now.

Tip 9

Get it back out the door quickly!

The rejected manuscript does you no good sitting on your desk. You need to get it back onto someone else's as quickly as possible. Make whatever changes you are going to make (see Tip #10), identify the next journal, and submit. To expedite this process (and to make the rejection less painful), at the time of initial submission I often list the two or three journals to which I will next submit the manuscript if (when!) rejection occurs.

As a corollary: It is rarely helpful to argue with editors, especially if they feel the topic is not a good fit for their journal. Never resubmit a manuscript to the same journal without getting express permission from the editor.

Tip 10

Take seriously all reviewer and editor suggestions

In most cases of rejection, you will receive feedback from the editor, several reviewers, or both. Even though you are not *required* to address all of these suggestions (and I've heard some experts suggest ignoring the feedback on rejected manuscripts), I believe it is foolish to ignore this free advice. Moreover, if the same reviewer is asked to review the manuscript again by the next journal, he or she may feel offended if suggestions are entirely ignored.

After a short cooling-down period (rejection is always hard!) I carefully consider all reviewer suggestions, prioritize each as (a) essential (e.g., errors, omissions, or ambiguities), (b) high-yield, (c) easy and useful, or (d) other (e.g., low-yield or erroneous), and revise the manuscript to incorporate those in the first three categories. Tip #11 contains additional suggestions for responding to reviewer feedback.

When you are invited to revise (because you will be, eventually)

Manuscripts are (almost) never accepted without revisions, so when you get an email stating, "We cannot accept it in its present form, but we would be willing to consider it again if you revise it to address the reviewer comments," that is very good news. However, you still have a lot of work ahead as you make the required revisions.

Tip 11

Respond carefully to every suggestion, even if you disagree

The reviewers are always right (Eva 2009). Even if you disagree with their opinion or believe they are flagrantly wrong, there is always something that you can clarify, justify, or explain in response to their critique, and these changes will nearly always improve the manuscript.

Reviewer comments typically vary (Fiske & Fogg 1990), but just because they address different issues does not necessarily mean that they disagree. Outright disagreements between reviewers are infrequent. More often, reviewers simply focus on different issues that individually are all important and collectively will substantially improve the quality of the work.

In contrast to rejection, following which you can selectively respond to high-priority suggestions, in responding to an invitation to revise it is essential to respond to every comment and suggestion. I classify reviewer comments into five types, namely (a) recognition of poor writing; (b) identification of an error; (c) suggestion to elaborate on a theme; (d) opinion without suggestions, and (e) compliment. Each of these requires a different response. These classifications, and my response approach, are listed in Table 3.

Some journals request that changes be tracked or otherwise highlighted. If not, be sure to again remove all tracked changes and comments prior to resubmission.

Be humble and respectful in the response letter. Remember that the reviewers gave freely of their time to read your manuscript and provide comments, and that they are (nearly always) trying to be constructive rather than disparaging. Even if you feel the reviewer is off base, it is usually possible to avoid a direct confrontation by finding and emphasizing areas of agreement, making a relatively inconsequential change in the manuscript text, requesting input from the Editor, or playing one reviewer off another ("Reviewer 2 suggested we shorten the Introduction, but Reviewer 1 indicated it was just the right length; we have elected to make no change for now, but would be willing to do so if the Editor believes it would be helpful."). To create a more favorable tone, it helps us to set aside the letter for a day or two, ask a collaborator to read it from the perspective of the editor or reviewer, or imagine that the reviewer is a good friend and will see this letter (which might actually be true).

Additional tips for an effective response letter include the following:

- Respond to every comment individually (except purely complimentary comments).
- Use white space (indented text), tables, and italics or bold font to distinguish reviewer comments from your response.
- Quote modified passages in full, and/or refer to the page number in the final manuscript where the text can be found.

Tip 12

Get input from others as you revise

Do not try to complete the revision on your own. Share reviewer comments with your coauthors and ask for their help in addressing concerns – especially the comments you find particularly challenging. If needed, contact a non-author colleague for help.

If you have concerns that cannot be resolved though discussion with your coauthors or other colleagues, you may wish to contact an editor for guidance on how to proceed. Remember that the editor is your friend – he or she invited you to revise and resubmit your work, and wants your response to be successful. However, be sure to follow proper channels

Table 3. Types of reviewer comments, and appropriate response.

Type of reviewer comment	Example*	Discussion and explanation	Response
Recognition of poor writing or omission	''The sentence on page 4, line 3, is difficult to understand.'' ''I could not tell how many repetitions each student performed.''	The reviewer is right – always. Even if you think it was clearly written, the reviewer had a hard time understanding. You might question his or her intelligence for not being able to understand your writing, but the reviewer is probably smarter (and investing more effort) than most other readers.	Clarify this area of ambiguity in the manuscript.
Identification of an error or limitation	''There appears to be an inconsistency between the data reported in the main text and in Table 2.'' ''The investigators used the *t*-test, but the Wilcoxon rank sum test would have been more appropriate.'' ''There are several other studies addressing this question, including work by [author]. These should be cited in the Introduction.'' ''The claim that the results apply to practicing physicians is not justified because it extrapolates beyond the data.''	The reviewer might be correct or incorrect about the issue. If he or she is incorrect, the mistake often arises because of an omission or ambiguity in your writing.	First double check your work, then seriously consider: Is the reviewer correct? • If you believe the reviewer is wrong, did his or her error arise because of ambiguous writing? If so, fix it. It helps to be humble, and take as much responsibility as possible. • If the reviewer is correct, fix or address the error. Tactfully explain your rationale for change or no change in the response letter. ''We neglected to report that we verified the assumptions for the use of parametric tests such as the t-test. We have clarified this point in the Methods, and continue to use the same statistical test.''
Suggestion to elaborate on (or trim) a theme	''It would be good for the authors to elaborate on the finding that ____.'' ''In discussing this point, the authors may wish to draw in the work by [author].'' [note this suggestion is less forceful than the ''error'' quoted above] "The authors spend too much time talking about ____, which is only tangentially related to this topic. This should be deleted.''	It is essential to distinguish errors, which must be fixed, from suggested elaborations (or deletions), which are optional. Errors affect the rigor or correctness of the methods, reporting, or interpretations, whereas elaborations affect only the scope and completeness of the inferences and implications. If you don't make a given elaboration the paper will still be just as rigorous and correct, although it might be incomplete. There is never enough room to say everything one would like to say. This applies to both ideas originating from the authors and those arising from the reviewers. Just because something is interesting or true or relevant does not mean it has to be said in this manuscript. You as the author must make tough choices about what is essential. Criticisms about the appropriateness of the conceptual framework or research question should be addressed as possible errors (see above).	For suggested elaborations, seriously consider: Is the message strengthened if you follow this advice? Or, conversely, does the suggestion confuse the issue, dilute the message, or open you to criticism? Based on this, decide whether to incorporate or defer the suggestion. • If you opt to incorporate a suggested elaboration, it is often appropriate to keep it short. • If you choose not to make a change, defend your decision by stating something like, ''This is an excellent suggestion, and we agree with the reviewer. However, due to space constraints we are not able to address this point fully." For suggested deletions, it is usually appropriate to trim text, although perhaps not as aggressively as the reviewer advises. Editor suggestions to trim overall length (e.g., to achieve a specific word limit) should always be followed, but you can exercise discretion in what text to trim.
Opinion without suggestions	''It is interesting to note that this issue arises in the work on cognitive load theory as well.''	If a reviewer discusses a topic without suggesting any changes (i.e., expressing	First carefully consider: Is a specific suggestion hidden in this opinion?

	''I disagree with the interpretation of finding ____.'' [a suggestion to change is implied] ''I found it particularly fascinating that these authors identified ____.''	platitudes, opinions, and prior knowledge), then strictly speaking no response is required. However, sometimes such narratives contain a partially-obscured suggestion that should be addressed. Even if no suggestion is requested or implied, the topic may be worth discussing (but see issues regarding suggested elaborations, above).	• If yes, then respond accordingly. • If no, then treat this as a suggested elaboration (above) or compliment (below).
Compliment	''This is an important question and a very timely study.'' ''The randomized design adds rigor.''	Compliments do not require a specific response, and can be ignored or addressed at your discretion.	Acknowledge this graciously and without elaboration in the response letter. If there are a lot of compliments, I usually just start off by saying ''We received a number of favorable comments and do not comment on these,'' and then ignore them.

*All example quotes are fictitious.

(e.g., working through the journal's editorial desk rather than contacting the editor directly) and be respectful of his or her time.

Tip 9 (revisited). Get it back out the door quickly

(Note the theme here!) You are so close. This is the final step – victory is within reach. Take this last leg of the race at a sprint, and get published!

Notes on contributor

DAVID COOK, MD, MHPE, is Associate Director of Mayo Clinic Online Learning, Research Chair for the Mayo Multidisciplinary Simulation Center, and Professor of Medicine and Medical Education, Mayo Clinic College of Medicine; and a Consultant in the Division of General Internal Medicine, Mayo Clinic, Rochester, MN, USA.

Acknowledgements

The author thanks Jonathan Ilgen, MD, and Ryan Brydges, PhD, for constructive comments on early drafts of this manuscript.

Declaration of interest

The author reports that they have no conflicts of interest.

References

Azer SA, Dupras DM, Azer S. 2014. Writing for publication in medical education in high impact journals. Eur Rev Med Pharmacol Sci 18: 2966–2981.

Beckman TJ, Cook DA. 2007. Developing scholarly projects in education: A primer for medical teachers. Med Teach 29:210–218.

Boice R. 1989. Procrastination, busyness and bingeing. Behav Res Ther 27: 605–611.

Bordage G. 1989. Considerations on preparing a paper for publication. Teach Learn Med 1:47–52.

Bordage G. 2001. Reasons reviewers reject and accept manuscripts: The strengths and weaknesses in medical education reports. Acad Med 76: 889–896.

Bordage G. 2009. Conceptual frameworks to illuminate and magnify. Med Educ 43:312–319.

Bordage G, Dawson B. 2003. Experimental study design and grant writing in eight steps and 28 questions. Med Educ 37: 376–385.

Bordage G, McGaghie WC, Cook DA. 2015. Title, authors, and abstract. In: Durning SJ, Carline JD, editors. Review criteria for research manuscripts. Washington, DC: Association of American Medical Colleges. pp 148–156.

Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HCW. 2003. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. Ann Intern Med 138:40–44.

Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P, for the CONSORT Group. 2008. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: Explanation and elaboration. Ann Intern Med 148:295–309.

Carpenter CR, Cone DC, Sarli CC. 2014. Using Publication metrics to highlight academic productivity and research impact. Acad Emerg Med 21:1160–1172.

Clarke M, Alderson P, Chalmers I. 2002. Discussion sections in reports of controlled trials published in general medical journals. JAMA 287: 2799–2801.

Cook DA, Beckman TJ, Bordage G. 2007. Abstracts and titles of research reports in medical education omit essential information: A systematic review. J Gen Intern Med 22(suppl 1):78.

Cook DA, Thompson WG, Thomas KG. 2014. Test-enhanced web-based learning: Optimizing the number of questions (a Randomized Crossover Trial). Acad Med 89:169–175.

Coverdale JH, Roberts LW, Balon R, Beresin EV. 2013. Writing for academia: Getting your research into print: AMEE Guide No. 74. Med Teach 35:e926–e934.

Docherty M, Smith R. 1999. The case for structuring the discussion of scientific papers. BMJ 318:1224–1225.

Eva KW. 2009. The reviewer is always right: Peer review of research in medical education. Med Educ 43:2–4.

Fiske DW, Fogg LF. 1990. But the reviewers are making different criticisms of my paper! Diversity and uniqueness in reviewer comments. Am Psychol 45:591–598.

Glassick CE. 2000. Boyer's expanded definitions of scholarship, the standards for assessing scholarship, and the elusiveness of the scholarship of teaching. Acad Med 75:877–880.

Gopen GD, Swan JA. 1990. The science of scientific writing. Am Sci 78: 550–558.

Huth EJ. 1999. Writing and publishing in medicine. Baltimore, MD: Williams and Wilkins.

International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. 2013. [Accessed 1 December 2014] Available from www.icmje.org/.

McGaghie WC. 2009. Scholarship, publication, and career advancement in health professions education: AMEE Guide No. 43. Med Teach 31: 574–590.

McGaghie WC, Bordage G, Shea JA. 2001. Problem statement, conceptual framework, and research question. Acad Med 76:923–924.

Moher D, Liberati A, Tetzlaff J, Altman DG. 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. Ann Intern Med 151:264–269.

Morgan PP. 1985. Building a table: 2. CMAJ 133:839.

O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. 2014. Standards for reporting qualitative research: A synthesis of recommendations. Acad Med 89:1245–1251.

Parsell G, Bligh J. 1999. AMEE Guide No. 17: Writing for journal publication. Med Teach 21:457–468.

Ringsted C, Hodges B, Scherpbier A. 2011. 'The research compass': An introduction to research in medical education: AMEE Guide no. 56. Med Teach 33:695–709.

Rizkallah J, Sin DD. 2010. Integrative approach to quality assessment of medical journals using impact factor, eigenfactor, and article influence scores. PLoS One 5:e10204.

Schriger DL, Cooper RJ. 2001. Achieving graphical excellence: Suggestions and methods for creating high-quality visual displays of experimental data. Ann Emerg Med 37:75–87.

Schriger DL, Sinha R, Schroter S, Liu PY, Altman DG. 2006. From submission to publication: A retrospective review of the tables and figures in a cohort of randomized controlled trials submitted to the British Medical Journal. Ann Emerg Med 48:750–756, 756 e751-721.

Schroter S, Tite L, Hutchings A, Black N. 2006. Differences in review quality and recommendations for publication between peer reviewers suggested by authors or by editors. JAMA 295: 314–317.

Schulz KF, Altman DG, Moher D. 2010. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomized trials. Ann Intern Med 152:726–732.

Strunk Jr W, White EB. 2000. The elements of style. Upper Saddle River, NJ: Pearson.

Tekian A, Roberts T, Batty HP, Cook DA, Norcini J. 2014. Preparing leaders in health professions education. Med Teach 36: 269–271.

Tufte ER. 2001. The visual display of quantitative information. Cheshire, CT: Graphic Press.

von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, for the STROBE Initiative. 2007. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies. Ann Intern Med 147: 573–577.

Wainer H. 1984. How to display data badly. Am Stat 38:137–147.

Wislar JS, Flanagin A, Fontanarosa PB, Deangelis CD. 2011. Honorary and ghost authorship in high impact biomedical journals: A cross sectional survey. BMJ 343:d6128.

Medical Teacher

ISSN: 0142-159X (Print) 1466-187X (Online) Journal homepage: https://www.tandfonline.com/loi/imte20

Looking back to move forward: Using history, discourse and text in medical education research: AMEE Guide No. 73

Ayelet Kuper, Cynthia Whitehead & Brian David Hodges

To cite this article: Ayelet Kuper, Cynthia Whitehead & Brian David Hodges (2013) Looking back to move forward: Using history, discourse and text in medical education research: AMEE Guide No. 73, Medical Teacher, 35:1, e849-e860, DOI: 10.3109/0142159X.2012.748887

To link to this article: https://doi.org/10.3109/0142159X.2012.748887

Published online: 21 Dec 2012.

2013; 35: e849–e860 MEDICAL TEACHER

WEB PAPER
AMEE GUIDE

Looking back to move forward: Using history, discourse and text in medical education research: AMEE Guide No. 73

AYELET KUPER[1,2,3], CYNTHIA WHITEHEAD[3,4,5] & BRIAN DAVID HODGES[3,6,7]

[1]Sunnybrook Health Sciences Centre, [2]Department of Medicine, [3]The Wilson Centre, [4]Women's College Hospital, [5]Department of Family and Community Medicine, [6]Department of Psychiatry, [7]University Health Network, UHN/University of Toronto, all at the University of Toronto, Canada

Abstract

As medical education research continues to diversify methodologically and theoretically, medical education researchers have been increasingly willing to challenge taken-for-granted assumptions about the form, content and function of medical education. In this AMEE Guide we describe historical, discourse and text analysis approaches that can help researchers and educators question the inevitability of things that are currently seen as 'natural'. Why is such questioning important? By articulating our assumptions and interrogating the 'naturalness' of the *status quo*, one can then begin to ask *why* things are the way they are. Researchers can, for example, ask whether the models of medical education organization and delivery that currently seem 'natural' to them have been developed in order to provide the most benefit to students or patients – or whether they have, rather, been developed in ways that provide power to faculty members, medical schools or the medical profession as a whole. An understanding of the interplay of practices and power is a valuable tool for opening up the field to new possibilities for better medical education. The recognition that our current models, rather than being 'natural', were created in particular historical contexts for any number of contingent reasons leads inexorably to the possibility of change. For if our current ways of doing things are not, in fact, inevitable, not only can they be questioned, they can be made better; they can changed in ways that are attentive to whom they benefit, are congruent with our current beliefs about best practice and may lead to the production of better doctors.

> *Let us give the term genealogy to the union of erudite knowledge and local memories which allow us to establish a historical knowledge of struggles and to make use of this knowledge tactically today.*
>
> (Foucault 1980, p. 83)

Practice points

- Making meaningful change in medical education requires questioning taken for granted assumptions about what medical education currently is and what it should be.
- Historical, discourse and text analysis approaches, which are widely and successfully used outside medical education research, can enhance our field by helping us to 'make strange' things heretofore accepted as 'normal' or 'natural'.
- History is not a singular linear development towards progressive improvement but rather a fluid construction incorporating multiple contextual perspectives.
- Discourse analysis enables researchers to understand the effects and relations of language, practices and power in our current assumptions about medical education.
- Text analysis, while often used in conjunction with discourse analysis, can also be useful for conducting many other types of qualitative research.

Introduction

As medical educators we strive continually to improve the form and content of the education and training we provide for future physicians. While this is a noble aspiration, as a medical education community we are often limited in our ability to make meaningful change because we assume that large components of our current system are rational and inevitable. However, history shows us that the structures of medical education are instead arbitrary and contingent. Questioning the many things that we take for granted within medical education can give medical educators the freedom to re-imagine what medical education could be. Such questioning is often difficult because our ways of teaching and learning appear to be so natural that it is difficult for us to think that they could be undertaken in any other way.

We begin this AMEE Guide with an approach we call 'making strange'. This is a way of gaining new, even startling, perspectives about things that we would otherwise accept as 'normal', because they are so familiar, so engrained in routine, so naturalized, that it becomes difficult to imagine that the

Correspondence: A. Kuper, The Wilson Centre, University Health Network, University of Toronto, 200 Elizabeth Street, 1ES-565 Toronto, ON M5G 2C4, Canada. Tel: 416-480-5495; fax: 416-480-5977; email: ayelet94@post.harvard.edu

ISSN 0142–159X print/ISSN 1466–187X online/13/O10849–12 © 2013 Informa UK Ltd.
DOI: 10.3109/0142159X.2012.748887

world could be organized in any other way. The notion of 'making strange' has been attributed to twentieth century German playwright Bertold Brecht. Sometimes called the 'distancing effect' after the German term '*Verfremdungseffekt*', Brecht wanted the audience of his plays to have a particular experience. Rather than being swept up in the drama in a way that allowed observers to lose themselves passively in the characters and the setting, Brecht instead wanted audience members to be critical observers and to be conscious of their critical observer position (Brecht 1964). To do this, he often crafted his stage settings to reveal the act of construction itself. Rather than trying to disguise the constructed nature of the play, he amplified it. In this way, the observer continues to participate in the drama, but in a way that allows them to remain conscious at all times that it is a drama – a drama that has been constructed in a particular and deliberate way.

Brecht's approach is helpful in thinking about medical education. Much of medical education is also constructed in particular ways. Faculty, students and patients play particular roles. Hospitals, clinics and classrooms are set in particular ways. We follow tradition and ritual in much of this but rarely reflect on their constructed nature. For example, for decades, medical school consisted of years spent in a lecture theatre, with a professor delivering wisdom to rows of students dutifully taking notes. Then, in about the 1960s, a new notion of small group, problem-based learning appeared. What had been taken for granted and assumed to be 'normal' for so long suddenly appeared, if not strange, at the very least worthy of questioning – a rather arbitrary construction, one that perhaps could be re-examined or changed.

How do we 'make strange'?

Practices that once appeared 'normal' may gradually begin to seem strange with the passage of time as scientific, social and political practices evolve. However, for the researcher, there are at least two effective strategies that are helpful in 'making strange' in a more deliberate fashion: applying historical and cultural lenses (Kuper & Hodges 2010). Simply examining a taken-for-granted practice from the perspective of another culture is often an effective way of throwing its constructed nature into relief. Many such examples have been described in the medical education literature. Student evaluation of teachers seems 'normal', for example, until one spends time with Japanese educators who explain that evaluation of an elder is culturally inappropriate. Assessing professionalism as an individual behaviour seems 'natural' until one studies a Confucian culture and learns that the behaviour of individuals is considered less important than the behaviour of the collective (Ho et al. 2011). Having examinations seems to be a taken-for-granted aspect of medical education until one visits Denmark and discovers that assessment is thought to foster competition which is thought incompatible with professional behaviour.

Applying a cultural lens is a fascinating way of making taken-for-granted elements of medical education seem strange. Several researchers in medical education today are productively pursuing this line of work. In this Guide, however, we will not discuss further the use of cultural lenses, but will delve into the second approach – applying historical lenses – in more detail. Examining practices in medical education at different periods of history is a very effective way of illustrating the constructed and therefore changeable nature of much in medical education.

Applying a historical lens to make strange

We begin with the premise that there is not a single 'true' history of anything. Every historical event that has been interesting enough to be retold, be it a revolution, a war or a political transition, can be seen from different perspectives. Placing emphasis on different elements of an event, on the perspectives of different participants in that event and on their different causal motivations results in rather different tellings and retellings of the event. And as time passes, new perspectives on the event may lead to further retelling. Thus, history is fluid. A simple version of this is the observation that, in war, 'history is told by the victor'. So too in medical education, a widely cited example being the much-recounted history of the Flexnerian reforms of medical education in North America. Following the release of Flexner's (1910) report by the Carnegie Foundation, medical education throughout North America was changed. But the way this bit of medical education history is told is subject to some very disparate interpretations (Hodges 2005).

The Flexnerian history is told variously as: an heroic accomplishment that led medicine to finally develop a scientific base and relocate its education in universities; a discriminatory turning point that led to the closure of medical schools for blacks and women (Strong-Boag 1981) and/or the beginning of an century of conflict of interest between the medical profession and corporate interests (Brown 1979). Interestingly, even the documentation of Flexner's own words does not come to us as a unified history (Whitehead 2010). Flexner himself critiqued the results of his own report some 15 years later (Flexner 1925).

The uses of history

From this perspective, history cannot be about the telling of a singular truth from the past or sketching a long and uninterrupted line of progress towards a better world. Rather history is about the different ways in which events have been, or could have been, recorded. Taking this approach to history is decidedly optimistic: rather than placing emphasis on the fixity of life, attention is paid to what is changeable. This constructivist approach allows us to question the underlying assumptions of recorded history, wondering not only why it was written in a certain way, in a certain place and at a certain time, but also how it might be written differently.

Studying the different ways in which history is told is called 'historiography' (Breisach 1983, p. 487). Examining a historical event through a deliberate juxtaposition of its multiple retellings opens up the potential for a 'critical' analysis. That is, questions can be raised about *who* told (or tells) which historical version, which *individuals or organizations* stand to gain or lose power or prestige from certain ways of recounting history, and what goals are advanced by emphasizing those particular historical details and interpretations. We can see that certain versions of history are more prominent than others, and

in particular times and places some become the dominant ways of understanding reality while others are suppressed. These different ways of seeing the world rest on conceptual systems and ways of speaking that together are termed 'discourses' (as explained in more detail below). Untangling this can be a tricky business because some discourses dominate today (at the time of current reading) just as potentially different discourses dominated in the past (at the time of the original writing).

Let us look at an example in more detail. As we have described, a critical version of the history of the Flexnerian reforms focuses on the concomitant closing of medical schools for women. Seen through twenty first century eyes an author might well interpret these events as examples of gender discrimination. However, a discourse of gender equity/discrimination was not very commonplace in 1910 in North America (such a discourse existed, but it was not dominant; for examples see: Jacobi 1891). Feminist historians, writing in the late twentieth and twenty first century have been able to show, through study of historical documents, that in 1910 the mainstream discourse construed women as intellectually inferior, inadequately adapted to the study of science and unable to cope with the demands of medical education and practice (Clevenger 1987). Discourses that supported more equitable admission of women to medical schools did not become prominent until the 1970s (Cooke et al. 2010).

The history of admission of women to medical schools could be told simply as a linear story of the advancement of equity and of women's rights: in the nineteenth century there was discrimination and by the end of the twentieth century there was great attention to gender equity. However that is only one telling, and it would miss a very important nuance. Prior to the Flexner reforms there were actually many women studying medicine in North America. The closure of medical schools was an attempt to address the multitude of 'proprietary' (for profit) institutions that Flexner argued had very low standards. The new medical schools that would emerge, based on the Johns Hopkins model, would be much more exclusive and much more expensive. Exclusion was about gender, but it was at least as much about socio-economic status. Were one to take an interest today in rising medical school tuitions, rising student debt and the proliferation of for-profit medical schools around the world, another look at the history of medical education and the Flexnerian reforms could be taken from such an economic point of view. Yet it is only in recent decades that an economic history of the medical profession has been told (Starr 1982).

We can see that the 'science-revolution' version of the Flexnerian history compares in interesting ways to the 'feminist-discrimination' version which can be juxtaposed with the 'economic' version. It could be argued that all are 'true' in some sense. However, the telling of the history takes on a different character depending on where emphasis is placed. Today, as we embrace a vigorous discussion about the appropriate criteria for admission to medical school, the ways in which we tell and retell medicine's grappling with this topic in the past are very relevant. An old adage, attributed to Santayana (1905, p. 284), holds that '*those who cannot remember the past are condemned to repeat it*'. The question that deserves our attention as medical education historiographers is, *which* histories are we doomed to repeat? Whitehead (2011) has illustrated that within the medical education discourse and literature the same topics and arguments recurring continually, albeit through the different lenses of successive historical periods. For example, the notion that medical knowledge has 'exploded' in such a way as to overwhelm the curriculum has been articulated in nearly every decade since 1910 (Whitehead in press). Understanding this 'repetition compulsion' must surely be important in moving forward.

Methodology

Discourse

Discourse is a concept that is becoming increasingly recognized in the medical education field. Like many popular terms, while its meaning is often assumed, it is actually used to express a range of different constructs. Mills (1997) provides a very useful summary of differences in the use of the term in different contexts and disciplines, including linguistics, sociology and psychology. In general, discourse relates to language, texts and the contexts in which language and texts are used and put into practice. In some forms of discourse analysis, this includes how they shape and are shaped by power structures and relations.

With a goal of 'making strange', so as to better understand our taken-for-granted assumptions in medical education, we have found that critical discourse analysis provides an effective and relevant approach for questioning such assumptions. It has been used in many disciplines to explore how language relates to the social construction of phenomena (Hodges 2009). Phillips and Hardy (2002) suggest that the different forms of discourse analysis can be categorized as focusing more on constructivist or critical approaches, depending on whether they highlight social construction processes or power dynamics. Types of discourse analysis can be further classified according to whether they focus more on text or on context (Phillips & Hardy 2002). Critical discourse analysis, in this schema, is a critical, context-focused approach.

Critical discourse analysis examines the way that discourse makes certain statements appear inevitable and not open to questioning or doubt. As described by Rogers et al. (2005, p. 371), critical discourse analysis is characterized by the 'movement from description and interpretation to explanation of how discourse systematically constructs versions of the social world'. Critical discourse analysis, therefore, focuses on the relation of language and practices and power. Parker (2002) has developed a very useful framework to guide researchers in distinguishing discourses. Box 1 provides a summary of this framework as presented by Shaw and Greenhalgh (2008).

Foucauldian critical discourse analysis

Many researchers who engage in critical discourse analysis from a historical perspective draw upon the works of Michel Foucault. Foucault did not offer a unified theoretical approach to history, but instead provided a number of concepts and

Box 1. Overview of Parker's framework.

Criteria for distinguishing discourses	Description
Discourse is realized in texts	As the world around us is textual, we need to treat objects of study (e.g. documents) as texts which are described and put into words
A discourse is historically located	Discourses are embedded in history and should be considered in relation to time. We need to explore how and where discourses emerge and describe how they change
A discourse is a coherent system of meanings	Discourse is made up of groups of statements that present a particular reality of the world. The task of the analyst is to map the world as discourse represents
A discourse is about objects	Using language means referring to objects and representing them in particular ways. Hence, we unpick what objects are referred to and how they are talked about
A discourse contains subjects	As discourse addresses us in particular ways and allows us to perceive ourselves in certain roles, we need to identify the rights we have to speak in relation to any discourse
A discourse refers to other discourses	Describing discourses necessarily involves the use of other discourses. Contrasting different ways of speaking helps to disentangle this
A discourse reflects on its own way of speaking	Each discourse comments upon the terms it employs, referring to other texts to elaborate. Hence, there is a need to reflect on the terminology used
Discourses support institutions	Discursive practices involve the reproduction of institutions. Analysis involves identifying institutions that are reinforced or subverted when a discourse is used
Discourses reproduce power relations	Discourse and power are intimately related so we need to look at which categories of person gain and lose from employment of a discourse
Discourses have ideological effects	Different versions of how things *should* proceed can coexist and compete within discourse. Hence, there is a need to show a discourse connects with other discourse to sanction control

Note: Parker (2002) as presented in Shaw and Greenhalgh (2008). Reproduced with permission.

theoretical lenses which can be combined to explore issues of knowledge and power as they vary across different historical periods. Foucault set out to study that which appears obvious or self-evident to us today, in contrast to what appeared to be self-evident to others in the past. He described this as unearthing the 'history of the present' (Foucault 1995, p. 31). In his examinations of madness, prisons and hospitals (Foucault 1980, 1988, 1995) he showed the ways that particular discourses are made possible, arise, change, become dominant and later disappear. Foucault focused on the analysis of discursive shifts (i.e. shifts between discourses), which he called *discontinuities* or *ruptures*. Several Foucauldian concepts, those of *archaeology*, *genealogy* and *serial history*, are particularly relevant to 'making strange' in medical education and unearthing aspects of its history; these will be discussed in detail in the sections that follow.

Archaeology. We commonly think of archaeology as digging up ancient pottery shards in order to help us reconstruct long-lost civilizations and how they worked. Foucault's use of the term archaeology similarly describes a way to metaphorically dig up bits of language in order to reconstruct the ideas and practices (i.e. the discourses) of the past as well as of the present. Foucault's concept of archaeology is helpful as it focuses attention on the way our ideas of 'truth' have been embedded in the different language that has been used in different ways in different times. It also requires us to analyse our current assumptions about accepted forms of knowledge since, for Foucault (2000, p. 132), 'Truth' is to be understood as a system of ordered procedures for the production, regulation, distribution, circulation and operation of statements'. By taking an archaeological approach, changes, or *discontinuities*, in the kinds of statements that are being made become extremely important, as these signal a shift in ways of thinking and in the rules governing discourse production. As described by Davison (1986, p. 223), '*new statements which seem to be mere incremental additions to scientific knowledge are in fact only made possible because underlying rules for the production of discourse have significantly altered*'. An archaeological approach probes something that might appear to be 'natural' and shows various factors that influence, affect and shape its emergence. Archaeology thus makes visible the confluence of forces allowing a discourse to emerge and the way the discourse operates. It '*attempts to isolate the level of discursive practices and formulate the rules of production and transformation for these practices*' (Davidson 1986, p. 227). By so doing, the '*conditions of possibility*' (Foucault 1994, p. xxii) are shown. Certain statements and ways of thinking are made possible; others are made impossible. Certain voices are heard and valued; others are not.

While discourses are characterized by particular ways of talking and thinking, they also encompass a number of other discrete but interrelated elements. These elements include roles for people to play, institutions to govern and have power, and objects (both real and conceptual) that are made possible by particular discourses. The Foucauldian historian tries to unearth as many of these discursive elements as possible, assembling them into a developing understanding of the discourse of which they are constituent parts.

Foucault's study of madness is classic in demonstrating that the twentieth century discourse of madness as illness is completely different from previously existing notions of madness as spiritual possession or social deviancy (Foucault 1988). Once madness is understood as mental illness, care of the insane becomes the job of doctors and hospitals, rather than clergy and churches or jailors and prisons.

Similarly, in *discipline and punish*, Foucault (1995) demonstrated a dramatic conceptual shift, as crime became something for which to be imprisoned rather than something to be punished by torture. Foucault showed that prison reforms, considered by proponents in nineteenth century as 'humanitarian' and 'progressive,' led to a marked change in disciplinary techniques. Instead of 'brutal but unfocused physical punishment' of the body of the criminal, there is

Box 2. Discursive changes in the *good doctor* in medical education.

The *good doctor* as a Flexnerian *scientist*:
The discourse of the *scientist* physician formed the basis of Abraham Flexner's proposals for reform. Flexner's *scientist* was an erudite and incisive thinker, who incorporated various forms of knowledge into his approach to his care for his patients. Flexner's *scientist* was generally socially well-placed, white and male

The good doctor as a man of character:
Flexner's notion of the *scientist* physician was not adopted with the changes to medical schools that followed his report. Instead, science became curricular content and the discourse of the *good doctor* as a man of *character* became prominent

The good doctor as a compilation of characteristics:
In the late (1950s) the discourse of the *good doctor* shifted from *character* to *characteristics*. Psychometric measures were increasingly used to dissect the medical student into his component parts

The good doctor as roles-competent:
This discourse considers the *good doctor* as *competent* in the performance of various *roles*. Discourses of *production* combine with competency language to depict a manufacturing model of medical training

Note: Material derived from Whitehead (2011).

instead 'intrusive psychological control' (Gutting 2005, p. 81). Self-control, self-discipline and self-surveillance are all products of this discourse. Implications of such different ways of thinking for society more broadly can be profound.

Box 2 presents a worked-out example of discursive changes within medical education derived using Foucauldian discourse analysis.

Genealogy. Foucault used the term genealogy not, as in common usage, to describe the discovery of individual family trees but rather to link knowledge and power. Whereas archaeology, in this framework, describes the specific discourses and their elements as they exist at particular points in time, genealogy is a study of the evolution of these discourses and the ebbs and flows of their relationships to each other. These ebbs and flows are not random; rather, they are animated by shifts in how power is enacted. Power is taken to be a force like electricity that is present in every interaction, every communication and every moment, and so does not lie in particular individuals or institutions. Using a particular discourse perpetuates a particular arrangement of power linked to that discourse, which in turn perpetuates the discourse itself. Genealogy thus examines the relationship between power and discursive practices, providing a 'history of the present' (Foucault 1995, p. 31). Foucault did not see knowledge and power as separable, meaning that shifts in what is considered to be 'true' are also inevitably shifts in power relations. In this framework, knowledge and power are interchangeable.

Foucault explicitly linked power and truth, describing *regimes of truth* that are made possible by certain discourses:

> [T]ruth isn't the reward of free spirits, the child of protracted solitude, nor the privilege of those who have succeeded in liberating themselves. Truth is a thing of this world: it is produced only by virtue of multiple forms of constraint. And it induces regular effects of power. Each society has its regime of truth, its 'general politics' of truth: that is, the types of discourse which it accepts and makes function as true; the mechanisms and instances which enable one to distinguish true and false statements, the means by which each is sanctioned; the techniques and procedures accorded value in the acquisition of truth; the status of those who are charged with saying what counts as true. (Foucault 1980, p. 131)

As we try to understand the effects and relations of language, practices and power in our current assumptions in medical education, genealogy helps to show how the relation of language, practices and power creates *regimes of truth*.

Power, most importantly, is not only something that is repressive, but is also very much a productive force. For Foucault:

> Power must be analysed as something which circulates, or rather as something which only functions in the form of a chain. It is never localised here or there, never in anybody's hands.... Individuals are the vehicles of power, not its points of application. (Foucault 1980, p. 98)

Serial history. Mapping of shifting discourses allows for an understanding of changing *regimes of truth*. In medical education, this allows us to see the way our assumptions change over time, and the implications and effects of these changes. Foucault clearly distinguished the difference between his notion of serial history and a linear history. Linear history seeks to explain events in terms of causal factors, are generally designed to demonstrate progress, and examine the past as a way to justify and explain the present (Foucault 1999, p. 423). A serial history, in contrast, does not take current conceptions or ideas for granted but seeks to understand how they came to be, examining the various factors and relations that allow new ways of speaking and thinking to be adopted:

> Serial history makes it possible to bring out different layers of events as it were, some being visible, even immediately knowable by the contemporaries, and then, beneath these events that form the froth of history, so to speak, there are other events that are invisible, imperceptible for the contemporaries, and are of completely different form. (Foucault 1999, pp. 427–428)

Serial history, therefore, offers a powerful way to focus on changes and shifts in language, and the way that such changes

in language connect to the construction and conception of other related ideas.

Elements of discourse: Example of physician competence

In summary, discourse consists of a variety of elements, analysis of which can demonstrate the connections between language, practices and power. These elements can best be described by using specific examples. A very current example of a discourse in medical education is that of physician competence (Hodges 2012). Competence, within this framework, is our discursive *object*. One discourse of competence is that of competence as *knowledge*. If a competent physician is one with appropriate knowledge, then the *role* for the student is to memorize facts. The role for teachers is to be a fount of knowledge, often delivering such knowledge through large group didactic teaching sessions. Compilations of facts, such as textbooks or lecture handouts, are provided to students for memorization and reproduction. Multiple-choice exams allow assessment of appropriate memorization; hence testing centres are dominant *institutions*. This knowledge accumulation approach draws upon the monastic tradition, in which the student is a passive recipient of knowledge approved by higher order experts.

In contrast, another discourse of competence is that of competence as *performance*. In this discourse, Miller's (1990) pyramid focuses our attention not just on knowing but on showing. Hence, a student's *role* is no longer that of memorizer, but instead becomes that of actor and performer. The teacher becomes an observer and demonstrator of skills. Instead of multiple-choice exam questions, the student is assessed through the use of Objective Structured Clinical Examinations or simulations. Standardized patient centres and simulation labs become dominant *institutions* in this discursive framing. In this discourse, which draws upon behaviourist understandings, observation of performance provides proof of competence.

Competence, in outcomes-based models, is positioned as a discourse of *production*. In this discourse, the student's *role* becomes one of raw material to be shaped and moulded by teachers who themselves take on the role of assembly-line management. Efficiency, accountability, quality assurance and standardized measures are valued, and this discourse of production draws on capitalist and corporate language. Interestingly, even when medical educators use outcomes-based models, current testing methods still hearken back to previous discourses of knowledge and performance in assessment, since nobody has yet found a foolproof way to measure outcomes. The emerging discourse of competence as *reflection* is appearing in conjunction with discourses of production, yet it builds upon very different foundations. The *role* of the student shifts from one of raw material to one of self-analyst; the *role* of the teacher moves from one of production-line manager to one of mentor and guide. This discourse draws on a belief that self-reflection and self-assessment provide a path to competence. Portfolios provide the measure of competency assessment in this framing.

As this example shows, when we approach the idea of physician competence from a variety of discursive lenses, we find that very different beliefs and values are at play in each. Teachers and learners assume different roles. Different institutions take on more prominent positions. Different people or institutions gain and lose power as these discourses privilege one or another form of teaching, learning and assessment. There are real and practical effects of discourse change.

While it is important to analyse each discursive strand separately, discourses do not, of course, appear in isolation. Instead, different discourses interact, with certain discourses rising to prominence at different times. Discourses coexist, sometimes clashing, sometimes bouncing off each other, and sometimes subtly transforming each other. Foucault notes that as a discourse is taken up in a different setting it:

> [C]irculates, is used, disappears, allows or prevents the realization of a desire, serves or resists various interests, participates in challenge and struggle, and becomes a theme of appropriation or rivalry. (Foucault, 1972, p. 105)

Looking at the intersection of discourses allows us to examine the various threads that are coming together in our daily practices.

Discourses in practice

Now let us see how we can use our understanding of discourse in a practical way, using the discourses of competence just discussed. Medical school admissions criteria are one obvious place these discursive framings can be helpful. If we wish to select students who will absorb and regurgitate large quantities of knowledge, marks in pre-medical school subjects that are taught and assessed in this way will be a good guide. If we wish to admit students who will perform on simulations, tools such as the Multi-Mini Interview, which is increasingly being used in North America for medical school admissions (Eva et al. 2004), should be a better marker of success. If we wish students who self-reflect, we might be able to better consider such abilities through an essay or interview. What it would be unwise to do would be to use marks in a biochemistry class as a measure of reflective capacity.

In practice we often find a muddled mix of discourses. The CanMEDS competency framework (Frank 2005), for example, uses the terminology of *roles* to describe its competencies and draws together performance and production discourses in so doing. Behaviourist roles are combined with outcomes-based statements. Assessment of competency in outcomes-based models generally combines examinations of knowledge, reflective exercises, performance measures and standardized checklists. Hence, these outcomes-based assessments are drawing upon knowledge, performance and reflection discourses of competence. It is very important to be aware of the different discursive threads that are being woven together in combining these assessment tools and the history of the development of each. After all, each of these discourses (knowledge, performance, production and reflection) is based

on different assumptions and is an expression of different values and practices. Taking this historical analytic view, it should be no surprise that we sometimes end up with combinations of disparate elements that may not actually make much sense together. By understanding the disparate elements, and the ways they fit (or do not fit) together, we may be better able to shape our tools in the future.

Method

For the researcher interested in taking these ideas about history and discourse and putting them into practice, the next step is to be deliberate and rigorous about identifying, collecting and analysing the appropriate data sources for this kind of research. Text analysis naturally aligns with the discourse analysis approach we have been discussing above. However because text analysis can be a useful method for many types of qualitative research, the following section also provides a more general overview of text analysis.

Why texts?

As for most qualitative methodologies, there are four major methods that can be used to gather data for discourse analysis: interviews, focus groups, observation and text analysis. Of these, observation is intrinsically limited to the study of events that are occurring in the present or that will somewhat predictably occur in the near future. Interviews and focus groups are somewhat more flexible in that they can be used to gather perspectives on events and occurrences from the recent past. These methods are, however, firmly limited by the life spans of the potential research participants: one could imagine contemporary researchers interviewing Admissions Committee members from the 1980s but of course not from the 1880s.

There are also theoretical issues inherent in gathering current perspectives on the past. Individual opinions and understandings shift over time. These shifts are often slow and subtle enough as to be imperceptible but may, over a prolonged period, become quite radical. As different discourses become dominant, different ways of thinking about the same questions become natural and obvious. These ways of thinking will colour participants' recollections and descriptions of past thoughts, decisions and actions. Thus, current interviews about past events are likely to be more useful indicators of current discourses than of the discourses that were in circulation at the time of those events. In order to access discourses contemporary to a particular period in the past, it becomes necessary to use data that was created in that period – that is, to gather and analyse texts.

What are texts?

The term 'text' encompasses a wide variety of physical objects that contain and convey meaning. Texts are most commonly taken to mean written documents but can also include such media as: visual arts including photographs, paintings and sculptures; graphic design; textiles; music and film. (See Box 3 for an example from medical education of a discourse analysis of a particular graphic design, the CanMEDS diagram as in Figure 1.) However, given the nature of our own expertise as well as the predominant textual medium currently taken up in discourse analyses, our focus in this Guide is on texts composed of written words.

Even within written texts there is a broad variety of different kinds of texts that offer different research possibilities. There are, for example, texts that were intentionally written to be read by many others (e.g. books, magazine and journal articles and blog postings), texts that were meant for a limited audience (e.g. letters and e-mails) and texts that were intended only for private consumption (e.g. diaries and notebooks). As another example, there are texts that are currently considered to be authoritative (e.g. articles in the *New England Journal of Medicine*), texts whose authority may be contested or denied (e.g. patient narratives posted in online communities) and texts which may be seen as illicit for transgressing ethical boundaries (e.g. medical student narratives about patients posted in online communities).

Which texts?

The selection of texts, often called 'delimiting the corpus', is a key step in textual analysis. This choice often begins with a research question, ideally situated within a particular theoretical and methodological framework, and proceeds with the identification of relevant texts. A researcher studying the spread of a particular discourse within a field of practice might, for example, want to focus on authoritative public texts in that field like journal articles or textbooks, albeit with a clear understanding of the limitations that this focus may engender. Another researcher interested in the impact of that same discourse on the medical student experience may need to search for other, less readily available texts like diaries or readily available but non-authoritative texts like blogs.

Once the general category of texts that are relevant to a research question has been identified, the researcher then selects specific texts according to his or her particular methodological approach. Some methodologies require a more rigid, predetermined delimitation of the texts to be studied, whereas others are more fluid or eclectic in their collection of textual data, but all require setting some sorts of boundaries around the texts to be studied. Examples of boundaries that might be considered in selecting particular texts are listed in Table 1. So, for example, if a researcher was studying changing admissions practices to the University of Toronto's Faculty of Medicine, she might delimit her corpus to include all documents produced by and for the admissions committee for that medical school, including its meeting agendas, its meeting minutes and its reports, between 1945 and 2010 (thus delineating boundaries of time, place, institution, authors and intended audiences). The time span would have to be justified (e.g. with respect to known changes in admissions to higher education in North America after World War II), as would the choice of institution. She would also have to justify other potential types of texts she had considered and chosen not to include (e.g. student newspaper articles about medical school admissions processes). Note that not all boundaries will be addressed in every situation; in this case, content, genre and language are not part of the formal

Box 3. A discourse analysis of a graphic design: the CanMEDS daisy.

This discourse analysis examined the CanMEDS diagram (Figure 1). Textual documents used in this analysis included the archives from the Educating Future Physicians for Ontario (EFPO) project (University of Toronto, Thomas Fisher Rare Books Library). The EFPO project developed a series of roles, which were modified by the Royal College of Physicians and Surgeons of Canada and organized into the daisy-shaped CanMEDS diagram. The discourse analysis aimed to understand the graphic design by identifying the discourses at play during roles development

The EFPO project began as a response to a strike by physicians in Ontario, Canada. Project leaders aimed to better align physician education with societal needs by defining a series of roles that physicians ought to play. The project involved extensive public consultations with physicians, educators, students, other health care professionals and members of the public, including representatives from multicultural groups, disabled persons groups, women's groups, AIDS groups and seniors' groups

The principal author examined all documents in the EFPO archive. One prominent discourse identified was a discourse of threat (to physician expertise, status and authority) and need to protect the profession from these threatening forces. A second discourse was that of societal need. The discourse of societal need was repeatedly invoked in the discussion of roles. The proposed use of 'roles' appeared in the earliest EFPO documents as the way to achieve societal needs. However, nowhere in the archive was the relationship between roles and societal need explained. Instead, the two were placed side by side in sentences, and their connection rhetorically assumed by their direct and recurrent juxtaposition

Visual images are not simply aesthetic, but convey messages that are value-laden (Zibrowski et al. 2009). The visual structure of the CanMEDS Framework is an innocent daisy, in which medical expertise is surrounded and encased by petals. Understanding the discourse of threat and a need to protect the profession's expertise that pervaded the EFPO documents, one possible interpretation of the graphic design could be that the 'petal roles' are functioning to 'armour' medical expertise

Note: Material derived from Whitehead et al. (2011a).

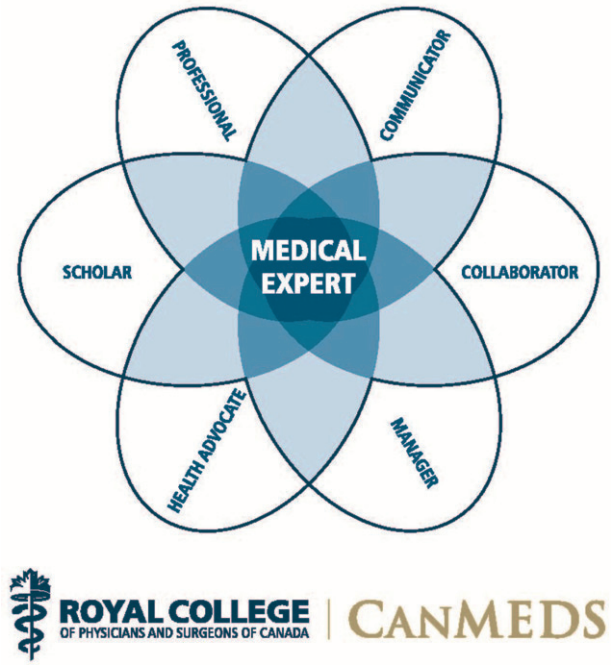


Figure 1. The CanMEDS diagram.
Copyright © 2009 The Royal College of Physicians and Surgeons of Canada. http://rcpsc.medical.org/canmeds. Reproduced with permission.

boundaries of the corpus but rather are determined by possibilities allowed by the other boundaries.

It should be noted that although in medical education research this particular directionality, coming from a research question to a text, predominates, there are disciplines in which the text sometimes or often comes first. That is, there are many domains of research in which a researcher may begin with a text (or a group of texts), determining relevant research questions based on the nature, content and/or context of that particular text. A classic example of this is the discipline of literary studies, where a researcher may begin by wanting to study a particular novel. In order to delineate a research question she would then immerse herself in everything previously written about that novel and potentially, depending on her theoretical and methodological orientation, about its author and the rest of that author's oeuvre, about the time the novel was written, about its literary antecedents, etc. Throughout this process the novel would be the central focus of her work and her jumping-off point into other writings and various ideas. Other disciplines in which this sort of text-centred process often occurs include art, rhetoric and history.

Primary vs. secondary texts

The texts that are selected as being relevant to a particular research question – the texts that are actually under study – are usually referred to as primary texts. Many methodologies also make use of secondary texts. These are texts that are outside the delimited boundaries of the corpus and are not being used to directly answer the research question, but that are nonetheless helpful in understanding the phenomenon under study. For example, the aforementioned researcher studying changing admissions practices at one medical school between 1945 and 2010, whose primary texts are its admissions committee documents, might also need to gather a selection of medical education journal articles, government documents, university policy documents and student information leaflets (among other things) to get a fuller picture of medical school admissions during that period. Primary and secondary texts, then, are not defined by the nature of the texts themselves but by the uses to which they are put in the research process; thus, the primary texts for one study might be the secondary texts for another, and vice versa.

Delimiting a Foucauldian archive

As described above, there are many decisions to be considered when deciding how to set appropriate boundaries and reasonably limit the texts that will be analysed. While the basic issues are similar (we all need a rigorously reasoned and well-described rationale for inclusion and exclusion of texts appropriate to our research questions and methodologies, with sufficient openness to needs and issues that emerge as the research is being conducted to make reasonable adjustments), Foucauldian critical discourse analysis employs some specific

Table 1. Boundaries that may be considered in selecting particular texts for analysis.

Boundary	Examples of what might be specified	Examples of selected texts
Time	• Weeks • Months • Years • Centuries, etc.	• The medical charts produced on a ward over a 6-week period • The programmes of a health professions education conference over a 10-year period
Place	• Neighbourhood • City • Region • Country, etc.	• The reports about health care from a city's newspapers • The government regulatory documents related to a country's nursing education policies
Institution	• University • Medical school • Hospital • Hospital ward • International non-governmental organization • Committee, etc.	• The diversity committee documents from a university • The policy documents governing clinical work on a ward
Language	• Dominant international languages • Minority languages • Language(s) known to the researcher, etc.	• The academic journal articles written in English about the globalization of medical education • The academic journal articles written in Chinese and Arabic about the globalization of medical education
Content	• Subjects • Theories, etc.	• The blog postings about the medical school application process from Canadian and British websites • The academic journal articles about professionalism in the nursing education literature
Genres	• Prose fiction • Poetry • Academic journal articles • Academic books • Blog posts • Popular press articles, etc.	• The memoirs of medical school experiences published as mass-market paperbacks in English since 1970 • The poems published in the arts sections of high-impact general medical journals
Authors	• Educators • Learners • Clinicians • Patients • Researchers, etc.	• The stories written by patients for an online literary health care journal • The reports written by medical school Deans during accreditation reviews
Intended audiences	• Educators • Learners • Clinicians • Patients • Researchers, etc.	• The patient-information leaflets about the presence of medical trainees within a hospital • The grant applications about knowledge translation submitted to a national health research funding body

terms and approaches. Unlike some other forms of text analysis, a Foucauldian approach requires bi-directionality. That is, a researcher does not just choose her set of texts and move from text to discourse. Instead, there is a process of back and forth between text and discourse. This may at first seem confusing, but if we look back at Parker's framework (Box 1) we see the importance of institutions, power relations and links to other discourses in a critical discourse analysis. So, for example, if a researcher wanted to undertake the study described above of the University of Toronto medical school admissions criteria from 1945 to 2010 as a critical discourse analysis, she would need to choose an initial set of texts as a starting point. While reading those texts, she would begin to locate key statements about admissions processes. From these, she would start to identify how these statements are constructed: who is saying them, for what purpose, and in what contexts. She would very likely find that she needed to look beyond the originally chosen documents in order analyse how these discourses were being legitimized and made possible. As she then moved to position the discourses she had identified within the broader social context she might need to examine additional texts to see how these statements relate to and are reinforced by specific practices, institutions and power relations. Obviously, since the researcher would not know what discourses would be identified when beginning the analysis, the initial choice of texts is a starting point rather than a rigidly defined archive. A descriptions of Foucauldian methodology will articulate the processes used and choices made by the researcher in delimiting her archive over the course of the research process (Hodges 2009, pp. 50–51).

Organizing the materials

As will by now have become clear, most forms of text analysis (including Foucauldian critical discourse analysis) require collecting, organizing and analysing a large volume of text. The practicalities of managing this amount of data can be disconcerting and potentially overwhelming. There is no one 'right' way to do this, but there are several factors researchers might want to consider in their decision-making. In recent years, the availability of many documents electronically has

Table 2. Questions to keep in mind when analysing texts.

General question	More specific questions that may be important depending on the text, the theory/methodology being used and the research question
What is the text?	Is it a book, journal article, blog post, letter, dairy, photograph, painting, film, etc?
Who wrote the text?	What is his/her/their gender, age, ethnicity, country of origin, country of residence, socio-economic status, class, educational level, profession, job, etc?
When was the text written?	In what century, decade, year, etc?
Where was the text written?	In what country, city, type of institution, specific institution, etc?
How was the text written?	In what language, genre, form, etc? Using what key words, metaphors, symbols, etc? Making what key arguments?
Why was the text written in that way at that place and time by that person/group?	Was it commissioned, submitted, self-published, secret, paid for directly by a funder, funded indirectly, a plea for funding, supported, a plea for support, required, forbidden, authoritative, contested, transgressive, etc?
How does the text relate to the research question?	
How does the text relate to other primary texts already analysed?	
How does the text relate to secondary texts relevant to the study as a whole?	

dramatically shifted organizational paradigms, and those whose corpus or archive is available in electronic format can make use of any one of many available software programs to store and manage them. Even when texts are not available electronically, database or referencing software can be used to maintain lists of texts being used in hard copy. However, no matter how sophisticated the software being used, it is still only an organizational tool. While researchers (and reviewers) can be seduced by fancy software, it can be just as effective to use index cards to keep lists of texts, to sort documents into piles and to identify key points with sticky notes and highlighters.

How to read and analyse a text

However a particular text is selected, and however the data drawn from it will be organized, the next step in using a text for research it is of course to read it and to analyse it. Those two steps, reading and analysis, are intimately bound up with each other. It is only to be expected that a researcher reading a text with a particular research question in mind will immediately start thinking about how it relates to that question, to other research texts she has already read and to her analytic understanding thus far of those texts. As well, the particular questions shaping themselves in her mind as she read the text (and the notes she would be taking as she read) would be guided by her theoretical and methodological orientation. This is no more or less true for text analysis than for the analysis of observations, interviews and focus groups; theory and methodology will orient the researcher to the relevance of different facets of her data and enable her to enter into the analytic process.

For example, a Foucauldian studying discourses of medical training – of what it's like 'to become a doctor' – might read her archive's tenth mass-market memoir from the 1980s about the medical training experience looking for words linked to discourses she had already begun to identify through her ongoing research, for moments when discourses interacted or even clashed, for examples of groups and institutions that gained or lost power within a particular dominant discourse. On the other hand, a critical feminist studying gender differences in descriptions of physicians in the 1980s might read the same memoir paying particular attention to who wrote it and their social location (gender, age, ethnicity, socio-economic status, class, etc.), specific words and phrases that were used in it to describe male and female physicians, the social locations of the character(s) in the memoir to which those descriptions were attributed, etc. Despite these theoretical nuances, there are certain basic questions that can be usefully kept it mind when reading texts; these are listed in Table 2.

This particular approach to text analysis is grounded in our particular expertise as medical education researchers who use history to make visible the contingent aspects of contemporary medical education, and more generally in our disciplinary affiliations as social scientists. Our goal is to reveal possibilities for change. Others who study texts, such as rhetoricians or social linguists, might pay even closer attention to phraseology, grammar or even punctuation; however, like Shaw and Greenhalgh (2008, p. 2519), '*although our analysis is not focused on the micro-analysis of texts, wherever possible we draw attention to concrete language use*' as part of our research data. Still others who study texts, such as literary scholars, might focus on intertextuality, character development or narrative structure; such interests are well-represented in the literature and medicine community, and occasionally cross over into medical education research. All of these approaches are useful; they simply draw on different theoretical and methodological armamentaria to answer different types of research questions. As well, they all share a common understanding: that analysing texts is not about a particular type of coding, about the software using to organize the textual data, or about coming to a single incontrovertible truth, but rather about considered thought, methodologically informed meaning-making and theoretically grounded interpretation.

Conclusion

We began this AMEE Guide with an approach we called 'making strange' and discussed how such an approach can produce unexpected insights about things we would otherwise accept as normal or natural. Our goal is to illustrate the utility of this approach for the medical education researcher. In undertaking such work, it is important to be aware that this approach, which questions the foundations of people's assumptions, can sometimes be perceived as provocative. We have all three occasionally encountered this reaction to our own work (see, e.g. Whitehead et al. 2011a, 2011b; Sherbino et al. 2011). We would never advocate avoiding controversial topics. However, in reframing currently accepted 'truths', the wise researcher might want to take into account that these 'truths' may be touchstones for some of their readers and should aim to gently lead their readers towards a more nuanced understanding rather than to shock them into a different awareness of particular issues.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

Notes on contributors

AYELET KUPER, MD, DPhil, is a Scientist at the Wilson Centre for Research in Education (University Health Network/University of Toronto) and an Assistant Professor in the Department of Medicine at the University of Toronto. She is an active Staff Physician in the division of General Internal Medicine at Sunnybrook Health Sciences Centre in Toronto. Her research interests include the nature of legitimate knowledge production within medical education research, the history of medical education research, and the roles of the medical humanities and social sciences within medical education.

CYNTHIA WHITEHEAD, MD, PhD, is the Vice-Chair, Education and an Associate Professor in the Department of Family and Community Medicine at the University of Toronto. She is an active staff physician at Women's College Hospital in Toronto. She is an Educator Researcher at the Wilson Centre for Research in Education (University Health Network) and Education Scientist at the Centre for Ambulatory Care (Women's College Hospital) at the University of Toronto. Her areas of interest include primary care education, outcomes-based education and the history of medical education. Her research focuses on critical discourse analysis of various aspects of health professions education.

BRIAN DAVID HODGES, MD, PhD, is a Professor in the Faculty of Medicine and the Faculty of Education (OISE/UT) at the University of Toronto, the Richard and Elizabeth Currie Chair in Health Professions Education Research at the Wilson Centre for Research in Education (University Health Network/University of Toronto) and Vice President Education at the University Health Network (Toronto General, Toronto Western, Princess Margaret and Toronto Rehab Hospitals). He leads the AMS Phoenix Project: A Call to Caring, an initiative to rebalance the technical and compassionate dimensions of health care. His research focuses on using Foucauldian-inspired discourse analysis to study various aspects of health professional education and practice: competence, assessment, professionalism and globalization.

References

Brecht B. 1964. Brecht on theatre: The development of an aesthetic. Willet J, translator and editor. New York, NY: Hill and Wang.

Breisach E. 1983. Historiography: Ancient, medieval & modern. Chicago: The University of Chicago Press.

Brown RE. 1979. Rockefeller medicine men. Berkeley, CA: University of California Press.

Clevenger MR. 1987. From lay practitioner to doctor of medicine: Woman physicians in St. Louis, 1860-1920. Missouri Historical Society Gateway Heritage, 8(3). The Bernard Becker Medical Libarary Digital Collection, Washington University in St Louis. [Accessed 30 March 2012] Available from http://beckerexhibits.wustl.edu/mowihsp/articles/practitioner.htm

Cooke M, Irby D, O'Brien B. 2010. Educating physicians: A call for reform of medical school and residency. San Francisco, CA: Jossey-Bass.

Davidson AI. 1986. Archaeology, genealogy, ethics. In: Hoy Couzens D, editor. Foucault: A critical reader. Oxford: Blackwell. pp 221–235.

Eva KW, Reiter HI, Rosenfeld J, Norman GR. 2004. An admissions OSCE: The multiple mini-interview. Med Educ 38(3):314–326.

Flexner A. 1910. Medical education in the United States and Canada: A report to the Carnegie Foundation for the advancement of teaching. New York: Carnegie Foundation.

Flexner A. 1925. Medical education: A comparative study. New York, NY: Macmillan.

Foucault M. 1972. The archaeology of knowledge and the discourse on language. Sheridan Smith AM, translator. New York, NY: Pantheon Books.

Foucault M. 1980. Power/knowledge: Selected interviews and other writings, 1972–1977. Gordon C, editor. Toronto: Random House of Canada.

Foucault M. 1988. Madness and civilization: A history of insanity in the age of reason. New York, NY: Vintage Books, Random House.

Foucault M. 1994. The order of things: An archaeology of the human sciences. New York, NY: Vintage Books, Random House.

Foucault M. 1995. Discipline and punish: The birth of the prison. 2nd ed. Sheridan A, translator. New York, NY: Vintage Books, Random House.

Foucault M. 1999. Aesthetics, method, and epistemology. Hurley R, translator. New York, NY: The New Press.

Foucault M. 2000. Power. Hurley, R, et al., translators. New York, NY: The New Press.

Frank J. 2005. The CanMEDS 2005 physician competency framework. Better standards. Better physicians. Better Care. Ottawa: The Royal College of Physicians and Surgeons of Canada.

Gutting G. 2005. Foucault: A very short introduction. Oxford: Oxford University Press.

Ho MJ, Yu KH, Hirsh D, Huang TS, Yang PC. 2011. Does one size fit all? Building a framework for medical professionalism. Acad Med 86(11):1407–1414.

Hodges B. 2005. The many and conflicting histories of medical education in Canada and the United States: An introduction to the paradigm wars. Med Educ 39(6):613–621.

Hodges BD. 2009. The objective structured clinical examination: A socio-history. Berlin: LAP Press.

Hodges BD. 2012. The shifting discourses of competence. In: Hodges BD, Lingard L, editors. The question of competence: Reconsidering medical education in the twenty-first century. Ithaca: Cornell University Press. pp 14–42.

Jacobi MP. 1891. Women in medicine. In: Myer Nathan, A, editors. Women's work in America. New York, NY: H. Holt & Co. pp 139–206.

Kuper A, Hodges BD. 2010. Medical education in its societal context. In: Dornan T, Mann KV, Scherpbier Ajja, Spencer J, editors. Medical education: Theory and practice. London: Elsevier. pp 39–49.

Miller G. 1990. The assessment of clinical skills/competence/performance. Acad Med 65(9):S63–S67.

Mills S. 1997. Discourse. London: Routledge.

Parker I. 2002. Critical discursive psychology. Basingstoke: Palgrave Macmillan.

Phillips N, Hardy C. 2002. Discourse analysis: Investigating processes of social construction. Thousand Oaks, CA: Sage.

Rogers R, Malancharuvil-Berkes E, Mosley M, Hui D, O'Garro JG. 2005. Critical discourse analysis in education: A review of the literature. Rev Educ Res 75(3):365–416.

Santayana G. 1905. Reason in common sense. New York, NY: Charles Scribner's Sons.

Shaw SE, Greenhalgh T. 2008. Best research – for what? Best health – for whom? A critical exploration of primary care research using discourse analysis. Soc Sci Med 66(12):2506–2519.

Sherbino J, Frank J, Flynn L, Snell L. 2011. "Intrinsic Roles" rather than "armour": Renaming the "non-medical expert roles" of the CanMEDS framework to match their intent. Adv Health Sci Educ Theory Pract 16(5):695–697.

Starr P. 1982. The social transformation of American medicine: The rise of a sovereign profession and the making of a vast industry. New York, NY: Basic Books.

Strong-Boag V. 1981. Canada's women doctors: Feminism constrained. In: Shortt SED, editor. Medicine in Canadian Society: Historical perspectives. Montreal: McGill-Queen's University Press. pp 207–235.

University of Toronto (UTL), Thomas Fisher (TF) Rare Books Library. Ms. Coll. 152. Associated Medical Services (AMS) Archives.

Whitehead CR. 2010. Recipes for medical education reform: Will different ingredients create better doctors? A commentary on Sales and Schlaff. SocSci Med 70(11):1672–1676.

Whitehead CR. 2011. The good doctor in medical education 1910–2010: A critical discourse analysis. Dissertation, University of Toronto. [Accessed 30 March 2012] Available from http://hdl.handle.net/1807/32161.

Whitehead CR. In press. Scientist or science-stuffed? Discourses of science in North American medical education. Med Educ.

Whitehead CR, Austin Z, Hodges BD. 2011a. Flower power: The armoured expert in the CanMEDS competency framework. Adv Health Sci Educ Theory Pract 16(5):681–694.

Whitehead CR, Austin Z, Hodges BD. 2011b. Intentions versus unintended discursive consequences: Reflections upon Sherbino et al.'s commentary on "Flower Power". Adv Health Sci Educ Theory Pract 16(5):699–701.

Zibrowski EM, Singh SI, Goldszmidt MA, Watling CJ, Kenyon CF, Schulz V, Maddocks HL, Lingard L. 2009. The sum of the parts detracts from the intended whole: Competencies and in-training assessments. Med Educ 43:741–748.

Medical Teacher

ISSN: 0142-159X (Print) 1466-187X (Online) Journal homepage: https://www.tandfonline.com/loi/imte20

Quantitative and qualitative methods in medical education research: AMEE Guide No 90: Part I

Mohsen Tavakol & John Sandars

To cite this article: Mohsen Tavakol & John Sandars (2014) Quantitative and qualitative methods in medical education research: AMEE Guide No 90: Part I, Medical Teacher, 36:9, 746-756, DOI: 10.3109/0142159X.2014.915298

To link to this article: https://doi.org/10.3109/0142159X.2014.915298

2014, 36: 746–756 MEDICAL TEACHER

AMEE GUIDE

Quantitative and qualitative methods in medical education research: AMEE Guide No 90: Part I

MOHSEN TAVAKOL[1] & JOHN SANDARS[2]
[1]The University of Nottingham, UK, [2]The University of Leeds, UK

Abstract

Medical educators need to understand and conduct medical education research in order to make informed decisions based on the best evidence, rather than rely on their own hunches. The purpose of this Guide is to provide medical educators, especially those who are new to medical education research, with a basic understanding of how quantitative and qualitative methods contribute to the medical education evidence base through their different inquiry approaches and also how to select the most appropriate inquiry approach to answer their research questions.

Introduction

Over the past few decades, major advances have occurred in both the understanding and practice of medical education. Medical education research has contributed considerably to these advances by adding reliable new knowledge to an existing body of educational knowledge to produce 'best evidence' that can help medical educators to make better decisions about important areas of medical education, such as teaching and learning, effective curriculum design and assessment. Through research, data can be collected and analysed to better understand the teaching and learning process (Norman 2002) and also to inform decision making about how well a particular programme, practice, procedure or policy is operating (Tavakol & Gruppen 2010). However, there is often little interest by clinicians in medical education research, possibly as a result of a lack of training in education research methods, and with many clinical educators also feeling less confident in the application of qualitative research approaches (Tavakol et al. 2008). This could be due to the fact that the nature of qualitative studies in comparison with quantitative methods has not been recognised (Morse 2005), especially since medical educators tend to gather empirical data that are grounded in objective rather than subjective reality (Buckley 1998). However, the contribution of qualitative studies in evidence-based practice has increasingly been recognised in both healthcare systems and educational research (McEwan et al. 2004; Ong & Richardson 2006; Bower & Scambler 2007).

The purpose of this Guide is to provide medical educators, especially those who are new to medical education research, with a basic understanding of how quantitative and qualitative methods contribute to the medical education evidence base through their different inquiry approaches. It also provides readers with the primary steps of the research process and an understanding of how to select the most appropriate inquiry approach to answer their research questions

Practice points

- Quantitative and qualitative studies are not contradictory, but complementary. Both develop new knowledge for solving research problems.
- Quantitative research has a positivist paradigm, in which the world to be researched is viewed as an objective reality, but qualitative research has a naturalistic paradigm, in which the world to be researched is viewed as a socially constructed subjective reality.
- Qualitative research provides an opportunity to generate and explain models and theories inductively, whereas quantitative research provides an opportunity to test theories deductively.
- When there is little knowledge about the phenomenon of interest, qualitative approaches are suggested to explore and understand the phenomenon.
- In quantitative research, the accuracy of the research results depends on the validity and reliability of the measurement tools, whereas in qualitative research the trustworthiness of the research findings heavily relies on the researcher as a tool, and hence participants should verify their findings.
- Quantitative researchers rely on numerical values obtained from statistical procedures and their corresponding p values, whereas qualitative researchers rely on excerpts from the actual voice of participants to describe and support the identified themes.
- All research must consider essential ethical principles to ensure that participants are not harmed, either in the process of data collection or by the presentation of results.

Correspondence: Dr. Mohsen Tavakol, Medical Education Unit, University of Nottingham, Nottingham NG7 2UH, UK. Tel: +44(0)115 823 0014; E-mail: mohsen.tavakol@nottingham.ac.uk

ISSN 0142-159X print/ISSN 1466-187X online/14/90746–11 © 2014 Informa UK Ltd.
DOI: 10.3109/0142159X.2014.915298

What is medical education research?

Research is '. . . investigation or experimentation aimed at the discovery and interpretation of facts, revision of accepted theories or laws in the light of new facts, or practical application of such new or revised theories or laws' (Merriam-Webster 2013). The ultimate goal of research is to gain new knowledge that can then be added to a body of existing knowledge in order to develop new insights and create more useful knowledge to solve a problem. Medical education research is a careful or systematic study designed to answer the fundamental questions raised by medical educators in order to make educational decisions that can be based on rigorous research-based findings rather than personal experiences.

Understanding of the knowledge construction process

Paradigms

Medical educators are always faced with questions in the real world, such as why do students struggle to learn genetics or does using a video improve learning? They use specific methods based on their own views of the world in order to find out the best answers to these questions. In the terminology of research, a paradigm is a comprehensive belief system or a worldview that provides a general perspective or framework to guide an understanding of the phenomenon under investigation. Hence, paradigms direct medical education researchers to employ the optimal methodological techniques given the nature of the phenomenon under study. According to Guba (1990), the paradigms that are adopted by educators respond to three questions: (a) What is the nature of reality (known as ontology or metaphysics)? (b) What is the nature of knowledge, its limitations and its relationship to the researcher (known as epistemology)? (c) How should the researcher go about finding out knowledge (known as methodology, by which the researcher chooses to conduct the investigation of the phenomenon)? For example, researchers may employ a cross-sectional design with one of the quantitative research traditions to answer their questions. The ontological and the epistemological questions focus on philosophical issues underlying research paradigms. In medical education research, there are a variety of research questions that may be raised by medical educators. Answering these questions requires the use of different types of research paradigms. Two main paradigms that guide disciplined inquiry in medical education, the positivist paradigm and the naturalistic paradigm, are discussed below.

The positivist paradigm

Epistemologically, over the last few centuries, the positivist paradigm was dominant to create new knowledge. It was introduced by Descartes in 1637. For Descartes, there is an objective reality that is directly observable and this can be measured using mathematical models that can predict future events. He believed that the researcher must distance him/herself from the participant to avoid any distortion of the interpretation of the findings from the study. Positivists believe that objective collection of data and its analysis must be independent of the opinions of the researcher. In Table 1, the ontological, epistemological and methodological assumptions of the positivist approach to research are summarised.

Within positivism, a hypothesis is derived from a theory and then empirically tested and replicated by a neutral researcher. Based on the result of a statistical hypothesis test, the researcher identifies the relationship between cause and effect within a value-free inquiry (Denzin & Lincoln 2011). According to positivism, there is an objective knowledge that is to be discovered and human beings cannot socially construct this knowledge. The generalisation of the study results to another situation is possible within a positivism paradigm; the nature of reality is fixed, it is single (i.e. the study results either support or reject a particular hypothesis), and it is measurable (Denzin & Lincoln 2011; Rubin & Rubin 2012). A positivist researcher will neutrally rely on statistical inferences, and if the

Table 1. Some assumptions of the positivist and constructivist approaches.

Type of assumption	Positivist approach (quantitative)	Constructivist approach (qualitative)
Epistemology (what is the relationship between the researcher and knowledge)	Knowledge is uncovered by detached scientific observations. The reality is independent of any opinions of the researcher. The researcher tries to minimise subjectivity and to maximise objectivity	Knowledge is socially constructed through interaction of the researcher with research participants. The values of both the researcher and the research participants contribute to knowledge, with there is a lack of neutrality and objectivity.
Ontology (what is the nature of reality)	The reality is singular. Reality is constructed based on cause and effect inferences.	Multiple realities exist. Each study participant has a different view on the phenomenon being studied.
Methodology (What is the research process?)	Deductive reasoning: Statistical hypothesis testing Objective and measurable Validation of theories Prediction and estimation Identifying associations between variables Generalization from samples to population Rule-bound Statistical analyses Internal and external validity Sample is large or random	Inductive reasoning: theory or hypothesis construction Subjective and non-measurable Explore participants' experiences Provide rich description of the phenomenon being investigated Generate hypothesis or theory Generalisation does not matter Context-bound Sample size is small

study results do not fit a theory/model, then the theory/model can be modified and subsequently tested using statistical procedures.

Positivism was criticised by post-positivists and in the last quarter of the twentieth century was rapidly deflated (Alvesson & Skoldberg 2009). Although proponents of the positivist approach believe that there is a fixed and objective reality that can be investigated, post-positivists argue that the absolute reality of knowledge can never be obtained, it is only estimated. Post-positivism considers that the reality is captured based on multiple methods and researchers seek to test, verify and refine theories to understand the world (Creswell 2014). Post-positivist researchers begin with a theory, then collect data in order to either support or refute the theory, and then make necessary changes and collect further data to check on whether the theory is supported or refuted (Denzin & Lincoln 2011; Creswell 2014). Thus the assumptions of post-positivist mostly support quantitative inquiry approaches rather than qualitative inquiry approaches.

The naturalistic paradigm

A different perspective to understand the world is the naturalistic view, otherwise known as constructivism, and is associated with qualitative inquiry approaches. This perspective is typically considered as an approach to qualitative inquiry (Lincoln & Guba 1985; Creswell 2014). The constructivist paradigm, a social movement opposed to positivism, began with writers such as Weber and Kant (Polit & Beck 2014). According to the constructivist epistemology, 'knowledge is the result of a dialogical process between the self-understanding person and that which is encountered, whether a text, a work of art, or the meaningful expression of another person' (Smith 1990). For constructivist researchers, individuals do not passively receive knowledge, but they actively construct knowledge through engagement with each other and the social world they are living in. Uncovering and understanding the social world of individuals thus leads to the production of meaningful knowledge. The epistemological perspective of constructivism has a focus on the 'meaning-making activity of an individual's mind' (Crotty 1998; Ritchie et al. 2013). In Table 1, the ontological, epistemological and methodological assumptions of the constructivist paradigm to research are summarised.

Ontologically, from the constructionist perspective, reality is not a single (one knowable) reality, but there are multiple realties which are constructed by those who participate in the study and this reality is negotiated with the research participants. As previously mentioned, positivist researchers believe that the individual parts of reality are not interrelated and can be separated into separate dimensions. However, naturalists believe that the reality is a whole and is not divided into its parts. For example, the parts of a 'whole cloth' (as a reality) are interrelated and inseparable. By removing part of the cloth, we actually destroy the cloth (its meaning) (Erlandson et al. 1993). From an epistemological perspective, constructivist researchers collect data subjectively to explore a single overall dimension of a phenomenon so that its true meaning is captured rather than reducing it in a number of different individual dimensions of the phenomenon. Additionally, constructivist researchers consider that inquiry and knowledge are value-laden since the researcher's beliefs highly influence the interpretation of knowledge (Tashakkori & Teddlie 1998; Griffin & Museus 2011). In contrast to positivists, which have a deductive approach to analysis (i.e. they choose a theory first and based on that theory they formulate hypotheses to test), the approach of constructivist researchers is inductive (i.e. the researcher begins with the participant's perspective and then a hypothesis or theory is created which is grounded in the real-life experiences of participants). This emergent theory illuminates the phenomenon under investigation.

Inductive and deductive approaches

Knowledge is created, based on either the inductive or deductive approach. As previously pointed out, qualitative researchers use the inductive approach to generate knowledge whereas quantitative researchers use the deductive approach to generate knowledge. The process of the inductive approach (a bottom-up method of analysis) begins with exploring the specific details of participants' experience and then gradually moves to more general principles of the phenomenon being investigated (Liehr & Smith 2002). For example, suppose a clinical educator is interested in exploring the experiences of medical students in problem-based learning (PBL). The clinical educator can use an inductive approach and by using qualitative research methods, such as interviewing students who have experienced the PBL approach, he/she can generate new understanding and theory about the PBL experience. Each student has their own specific experiences but interviewing several students will identify several common themes across the students. The process of the deductive approach, on the other hand, begins with formulating a research hypothesis about the phenomenon of interest (which is usually based on a theory). The hypothesis is then tested using statistical procedures to support or refute this hypothesis. Causal explanations, generalisation and prediction may be made based on these statistical procedures. For example, suppose the same clinical educator has formulated a hypothesis entitled 'the PBL group will score higher than the non-PBL group in the communication skills course'. The clinical educator needs to consider theories that have been previously developed by qualitative researchers to develop the assessment questions and then use quantitative research methods to collect and analyse data. Therefore, quantitative researchers test theories in order to generalise the study results to the target population. Both approaches are important for generating knowledge and the choice is based on the question being investigated.

The role of theory in research

A theory is 'a set of interrelated constructs (concepts), definitions, and propositions that present a systematic view

of phenomena by specifying relations among variables, with the purpose of explaining and predicting the phenomena' (Kerlinger 1970). Fundamentally, a theory is an idea, a guess, or a speculation, which may account for reality. Theories should guide the research process both in qualitative and quantitative research methods (Morse & Field 1995). Qualitative researchers, use the inductive approach to research and explore the observed data for 'the patterns and relationships and then develops and tests hypotheses to generate theory or uses developed theories to explain the data' (Morse & Field 1995). In this approach, research questions are created by the qualitative researcher, and then data are collected in the participant's setting. Data analysis is inductively built from specific (particulars) to general themes (generating categories and themes). The themes are finally interpreted by the researcher (Creswell 2014). However, sometimes qualitative researchers use a deductive approach in the initial stage of qualitative data analysis. They can develop a template (codebook) that uses a theoretical framework in order to organise the qualitative dataset for interpretation. However, these categories may not accurately reflect the participants' views of the phenomenon under investigation. An interesting example of a hybrid approach of inductive and deductive approach was well illustrated by Fereday & Muir-Cochrane (2006).

On the other hand, quantitative researchers formulate a research hypothesis deductively from an existing theory, and then the research hypothesis is tested by gathering data. Based on the statistical procedures used for the hypothesis testing, the existing theory is either revised or supported. For example, based on the humanistic education theory (by Carl Rogers), a researcher may hypothesise that small group teaching is more effective than large group teaching and this hypothesis can be tested using statistical procedure (Lodico et al. 2010).

Quantitative researchers who do not employ a theoretical framework for their own research study, particularly those who wish to establish cause-effect relationships, may struggle to explain why some independent variables influence the dependent variables (Kawulich 2009). In quantitative studies, theory-driven investigations are essential for the generalisation of the study results. Unfortunately, in many medical educational papers, the research question or hypothesis is not connected to a theoretical framework.

Concepts, constructs and variables

In qualitative studies, the building blocks of a theory are called concepts (Brown 2010; Polit & Beck 2014). Concepts are abstractions of particular characteristics of human behaviour, such as empathy, motivation and pain (Polit & Beck 2014). Researchers are unable to directly observe concepts in the real world but can measure them indirectly as a construct. For empathy to be measured as a construct, researchers need to identify the behavioural manifestations of empathy that can be considered as proxies of empathy, such as summarising the feelings that are expressed by participants. The terms of concept and construct are often used interchangeably in research.

A variable is a concept, and as its name suggests, is something that is likely to vary. From a quantitative point of view, a concept is observable and measurable and takes different values. For example, age, gender, teaching methods all are variables as they vary from one individual to another. Quantitative researchers are interested in investigating how or why phenomena vary, and also how the variation in a variable is explained by the variation in another variable. As an example, consider the variable of 'learning' and a research study that wishes to investigate what factors can affect student learning. Motivation as a variable may be investigated as a learning factor by the researchers. Quantitative researchers quantify student performance, for example, ranging from 0 to 100. It is noteworthy to mention that if every student obtained a mark of 60, student performance would not be a variable, it would be a constant. Qualitative researchers do not quantify a variable. For example, student performance could be reported using qualitative words, such as inferior, poor, borderline, satisfactory, good and excellent.

Dependent and independent variables

Quantitative researchers make a link between the basic building blocks of theory and the basic unit of scientific studies in order to establish the cause and effect relationships between variables. For example, does an educational intervention produce improvement in the reliability of OSCEs? In this example, researchers face a cause and effect relationship between educational intervention and the improvement of OSCEs. The presumed *cause* is the independent variable (sometimes called the exposure or predictor) whereas the presumed *effect* is the dependent variable (sometimes called the response or outcome). Quantitative researchers are interested in knowing how the independent variable causes the change in the dependent variable, especially if the independent variable predicts the dependent variable (Brown 2010). Sometimes it is very difficult to decide which of two variables in a study is the independent variable and which is the dependent variable (McBurney & White 2010). As an example, consider there is an association between drug education programmes and medication compliance. It is very difficult to conclude whether drug education programmes cause medication adherence or whether a predisposition to medication adherence causes people to adhere to a medication regimen. In medical education research, there are many confounding factors that can influence the dependent variable (outcome). Sometimes researchers are unable to manipulate independent variables in order to see its effect on the dependent variable. Examples include age, gender and year on a medical school programme. Qualitative researchers, on the other hand, are not interested in quantifying associations and relationships, or in seeking cause and effect connections. They are interested in similarities and differences in patterns of association in order to explore the underlying meanings of the phenomena under investigation.

Debate over the quality of qualitative and quantitative research

At the beginning of this Guide, we discussed philosophical perspectives of qualitative and quantitative research methods. When we are speaking of research methods, most medical educators and clinicians think of research studies that have a large sample size, and are randomly taken from the population of interest. They think how to randomly assign their study participants to groups (intervention and non-intervention group). They also think of gathering numerical data in order to use statistical procedures to produce study results. Although research studies that follow these steps in the research process can be useful, it is not enough to produce knowledge about reality, especially where situations are examined through the eyes of the participants (Cohen et al. 2008). For example, how clerkship students interact with the parents of unconscious children in hospital, 'what are the processes and strategies of clinical reasoning used by the students to produce treatment?' (Khatami et al. 2012) or 'medical students understanding of empathy' (Tavakol et al. 2012). Such social situations can be explored best by a researcher who integrates him/herself in the situation and obtains ideas, feelings, expectations, perceptions, experiences and behaviour patterns from the participants' point of view (Brown 2014). Perhaps more importantly, when researchers have little knowledge about new phenomena or new meanings of phenomena, qualitative inquiry methods are the best for gaining a deeper understanding of the phenomenon from the participant's perspective (Trice & Bloom 2014). In addition, when a theory is missing, qualitative studies can be used to generate theory (Leedy & Ormrod 2005). Quantitative researchers need to be guided by theories that are developed by the qualitative researchers. Qualitative study results can shed light on phenomena that are not accurately understood in teaching and practice. In addition, qualitative research methods are 'the most humanistic and person-centred way of discovering and uncovering thoughts an action of human beings' (Halloway & Biley 2011).

The underlying distinctions between quantitative and qualitative methods and their epistemological and ontological considerations have contributed to a better understanding of research issues. However, qualitative inquiry approaches have often been criticised by quantitative researchers (who view phenomena independent of the behaviour of the researcher), who consider quantitative research results to be more objective and value-free. In medicine, 'qualitative research continues to be devalued, and is considered to be 'subjective, biased, and opinion based' (Morse 2006, 2011). Quantitative researchers also argue that qualitative research does not have a strong design, and hence they do not recommend it for funding (Morse 2006). There are arguments against the quality of the knowledge produced by qualitative inquiry: 'quality in qualitative research is a mystery to many health services researchers' (Dingwall et al. 1998). Perhaps more importantly, the highest level of evidence has been awarded to quantitative research and the importance of qualitative research has been undermined and ranked at the lowest level of evidence (Cochrane 1972, 1989; Morse 2011). It should be noted that, however, chapter 20 of the Cochrane Intervention handbook has outlined how qualitative studies can contribute to Cochrane Intervention reviews. The handbook stated 'there are many methods of qualitative evidence synthesis that are appropriate to the aims and scope of Cochrane Intervention reviews'. For example, qualitative researchers should systematically review related individual papers in order to address 'important outcomes' and 'questions directly related to the effectiveness review'. For more details, see the handbook (Higgins & Green 2011). Although the number of papers in relation to qualitative evidence syntheses is growing (McInnes et al. 2011; Marshall et al. 2012), systematic reviews of qualitative studies are not well established in comparison with quantitative studies.

Although the issue of the nature of the knowledge produced by qualitative researchers has been raised by quantitative researchers, the discrimination against qualitative methods continues (Morse 2006), 'qualitative researchers can address the issue of quality in their research... its methods, can and do, enrich our knowledge of health and health care' (Mays & Pope 2000). Criticisms about qualitative research methods are based on a lack of understanding of what is the purpose of qualitative research studies to produce knowledge for medical educators and clinicians (Brown 2014). According to Morse 'the fact that so little is generally known about what qualitative inquiry is appalling' (Morse 2005, 2006). While qualitative inquirers gain knowledge of social reality, which is completely different from quantitative inquiries, both approaches can produce a wide range of knowledge about the phenomenon under investigation. There is a growing body of evidence that suggests the combination of qualitative and quantitative methods are important (Ashley & Boyd 2006), although 'it takes time and effort to understand both styles and see how they can be complementary' (Neuman 2003). Although quantitative and qualitative inquiry methods each have different underlying epistemological and ontological assumptions about the generation of knowledge and reality, their differences do not make one better or worse than the other. They are complementary rather than contradictory. Complex research questions require complex answers which can be achieved through the integration of qualitative and quantitative approaches. Qualitative researchers have now recognised that the perspectives of participants are not enough *per se* and multiple forms of evidence are essential. Likewise, quantitative researchers have realised that the perspectives of participants can play an essential role in quantitative results (Taylor 2013). It is perhaps for that reason that mixed methods research is grown increasingly popular among researchers. We need to think first that research is about inquiry and that distinctions between quantitative and qualitative approaches 'are arguably reified more by a need to label approaches than by true differences in in purpose' (Newman & Hitchcock 2011). Approaches should act as 'servants' rather than as 'rulers' and can be considered as different tools to be used when most appropriate to answer a research question (Silverman 2010). From a learning perspective, therefore, medical educators should learn techniques that are used in both quantitative and qualitative methods.

Steps of the research process

The quantitative research processes have a linear sequence, and consists of different steps, beginning with the identification of research questions and ending with a statement answering those questions (Nieswiadomy 1998; Polit & Beck 2014). The qualitative research process, on the other hand, tends to have a nonlinear sequence (or an iterative, repeating or recursive process). For example, qualitative researchers collect data and analyse them concurrently. They immediately begin data collection and analysis with the first interview. The next interview will be planned based on this interview. This process (i.e. data collection and analysis) continues until they sense that data saturation is achieved, that is until no new themes are identified. Unlike qualitative researchers, quantitative researchers analyse their data after all of the data are gathered. Due to the nature of qualitative studies, they may have different research process flow diagrams. Figure 1 illustrates the main steps in a quantitative research study.

Defining the problem

As seen in Figure1, researchers first need to identify the problem under investigation (Ary et al. 2006). Researchers should clearly state why they want to conduct a particular study. What is the knowledge gap in the field of study that needs to be closed? What is the importance of the problem? Researchers need to provide a rational for the study that they intend to undertake. The research problem should be logically developed into a discussion of the reasoning behind the study of interest, and end with a statement of the research question. Carefully constructed research questions will facilitate the search for a solution. It is worth mentioning that some researchers state the purpose of the study instead of the research question, with the aim of the study at the end of the introduction.

Literature review

Reviewing the related literature is an important step of the research process and the research report. Literature review informs us about the feasibility of researching the study topic before proper research begins (Hart 2005). Table 2 shows how a literature review can contribute to the study topic.

Quantitative researchers conduct a literature review to gather information on what is already known about the topic and the methods that have been used to study the topic before any data are collected. This will enable the researchers to provide the rationale for the study that they are planning. Consulting the literature can be useful for both quantitative and qualitative studies without considering the researcher's paradigm (Mertens 2010). Although some qualitative researchers acknowledge the importance of doing a review literature prior to commencing their research, some believe that qualitative researchers should not review the literature

Table 2. The contribution of the literature review to the study topic (Gillis & Jackson 2002; Hart 2005; Aveyard 2010).

- Ensures a comprehensive, relevant, clear picture of the studies available on the study topic;
- Identifies the main issues related to the study topic and hence establishes the importance of the study topic;
- Identifies areas where there are consistency or inconsistency in research results;
- Determines what is known and unknown about the study topic (the knowledge gap);
- Identifies experts in the related fields;
- Helps construct a theoretical/conceptual framework for the study topic;
- Helps the researcher to plan methods;
- Identifies how other researchers have measured and analysed their data;
- Discovers instruments or tools that can be used to measure the study variables;

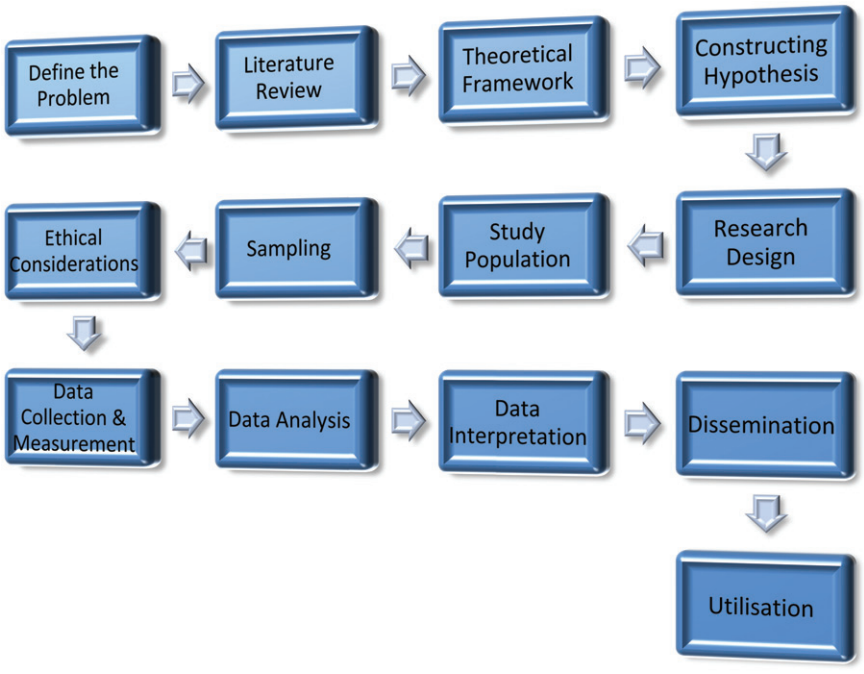


Figure 1. The main steps in a quantitative research study.

before proceeding with data collection. Opponents argue that by reviewing the literature, the conceptualisation of the phenomena under investigation might be contaminated. They believe that this should be explored based on participants' perspectives rather than prior findings (Morse & Field 1995; Polit & Beck 2014). However, others believe that qualitative researchers should be initially aware of what is already known about the phenomena under investigation, but that an additional literature review is also required during the research to make sense of the data (Marshall & Rossman 2006).

Develop a theoretical framework

Another step in the research process (Figure 1) is to develop a theoretical framework to the research study. Theories operate as a 'lens' through which to view the phenomena of interest (Sclater 2012). Theories are generated to describe, predict and understand the relation between two or more different concepts in order to construct universal laws. A theoretical framework (sometimes called a conceptual framework) is a part or a brief explanation of a theory that researchers can verify by hypothesis testing or seek answers to research questions that are driven from theory. A theoretical framework underpins the research problem under investigation, formulates the research questions or hypotheses, guides the data collection process, explains and predicts the underlying cause the phenomena under study (Reeves et al. 2008; Creswell 2014). Moan and Rise, for example, tested the use of the Theory of Planned Behaviour (Ajzen 2005) for explaining and predicting students' intentions to quit smoking and their subsequent behaviour six months later (Moan & Rise 2005).

In quantitative studies, the theoretical framework is deductively established before data are collected. Quantitative researchers report explicitly the theoretical framework of their studies in the introduction section, immediately after addressing the research questions or hypotheses. Qualitative researchers, on the other hand, generate, explain and understand a theory inductively during the research study. Theories and hypotheses are inductively generated after data analysis has begun (Morse & Field 1995). Carefully designed studies use a theoretical framework in order to guide the phenomenon to be studied. Quantitative researchers often do not explicitly discuss the theoretical framework in their reports in comparison to qualitative researchers. However, there are several educational theories that are relevant for medical education research, and these can be explored further through the work by (Kaufim 2003).

Constructing hypotheses

A hypothesis predicts the relationship between the independent variable and the dependent variable. Some quantitative studies explicitly address one or more research hypotheses but qualitative studies, on the other hand, do not have research hypotheses. This is because 'qualitative researchers want the inquiry to be guided by participants' viewpoints rather than by their own hunches' (Polit & Beck 2014). As previously stated, hypotheses are sometimes formulated from theories and often these are formulated from a large body of evidence. For example, a study hypothesised that 'women will show higher levels of empathy than will men' as this hypothesis is consistent with the previous studies (Toussaint & Webb 2005). Descriptive studies do not have a hypothesis.

A hypothesis contains the population, the independent variable, the dependent variable and a predicted relationship between them. Hypotheses are dichotomised into two groups: directional or non-directional. In a directional hypothesis, researchers can predict the direction of the association, either positively or negatively. In a non-directional hypothesis, researchers do not specify the direction of the association. Table 3 shows some examples of directional and non-direction hypotheses.

You may recall from statistics courses that there are two types of hypotheses: null hypotheses and alternative hypotheses (sometimes called research hypothesis). Researchers want to know whether or not their theories can be supported when subjected to the rigors of scientific investigations (Daniel 2005). The null hypothesis is a hypothesis of no difference (i.e. there is no difference between the independent and the dependent variables). The null hypothesis is either rejected or accepted by statistical procedures. If the null hypothesis is rejected, the alternative hypothesis is supported as the available data are incompatible with the null hypothesis (Daniel 2005). Hypotheses neither are proved nor disproved, but they are either supported (accepted) or rejected.

Study results are not always definite and researchers maybe unable to prove or disprove research hypotheses (Polit & Beck 2014). For example, consider the hypothesis that tall medical students show more empathy than shorter students. If a sample of students shows that tall medical students have higher levels of empathy than short ones, we cannot conclude that height is related to a student's empathy since in realty there is no relationship between height and empathy with patients. There are also other influences, including sources of measurement error that can influence statistical inferences, such as the accuracy of measures and factors that are not under the control of the researchers.

Table 3. Directional and non-directional hypotheses.

Type of hypothesis	Hypothesis
Directional	PBL students are better able than non-PBL students in disclosing bad news to patients with life-threatening illness
Non-directional	There is an association between PBL student and non-PBL students in disclosing bad news to patients with life-threatening illness
Directional	OSCEs better measure medical students' clinical performance than do mini-CEXs
Non-direction	There is a relationship between OSCEs and mini-CEXs with respect to measuring medical students' clinical performance.
Directional	Female medical students have more positive attitudes towards epidemiology training than male medical students.
Non-directional	Female medical students differ from male medical students with respect to epidemiology training.

Quantitative research designs

Quantitative research designs differ from qualitative research. Quantitative research designs are classified into three groups: experimental designs, quasi-experimental designs and surveys (Creswell 2013). Experimental designs explain the cause and effect relationship between the independent and dependent variables. Three important features of experimental designs (or Randomised Controlled Trials, RCTs) are: control, manipulation or intervention, and randomisation. These features help quantitative researchers to ensure that the study outcome is caused by a particular intervention rather than by other variables. Researchers control variables (major factors) which may influence the study outcome as they want to ensure that the study outcome is caused by the intervention rather than other variables (sometimes called extraneous or confounding variables). Quantitative researchers also manipulate the independent variable (cause) and then they measure its consequence on the dependent variable (effect). Another feature of experimental designs is randomisation. Researchers randomly assign study participants into experimental or control groups. The experimental group is exposed to a treatment, but the control group does not receive any treatment. Randomisation means that each participant has an equal chance of being selected to either group. By taking these features into account, researchers are able to generalise the study outcome to the population of interest. To illustrate, consider a medical educator investigating the effect of high-fidelity simulation (independent variable) on medical students' knowledge and clinical performance (dependent variable). In this example, simulation may influence knowledge and clinical performance. The medical educator can assign students to two groups randomly (i.e. each student has an equal chance of getting the experiment or control condition). Before the intervention was experienced, the educator can assess (pre-test) the knowledge and clinical performance of students in each group. Subsequently, students in the experimental group can be exposed to the simulators whereas students in the control group are exposed to a traditional intervention which was similar to the information covered in the simulator. The educator then can reassess the knowledge acquisition and clinical performance in both groups (post-test). Finally, the educator can compare the differences between the post-test scores of the two groups in order to identify the effect of the simulator on the knowledge acquisition and clinical performance, while the educator considers other variables (Figure 2). This experimental design is called a randomised control-group pre-test-post-test design (sometimes called an RCT design).

Sometimes in medical education, educators are unable to assign participants randomly to experimental and control groups, but they want to assess a particular intervention. If this is the case, educators need to follow a quasi-experimental design. Sometimes this design is called trials without randomisation. There are different types of quasi-experimental designs (Shadish et al. 2002; Harris et al. 2006) but two important designs will be discussed here: the non-equivalent control group pre-test-post-test design and one group pre-test-post-test design.

The non-equivalent control group pre-test-post-test design (sometimes called a comparison study) compares two or more groups of participants before and after a particular intervention without assigning participants to the experimental and control groups. This design is the same as the pre-test-post-test experimental design (see above), except participants are not grouped (the experimental and control groups) randomly (Polit & Beck 2014). As an educational researcher, you can compare students at medical school A as the experimental group with students at medical school B as the control group. The experimental group attend clinical simulation (intervention) activities whereas the control group attend the normal clinical programme for three weeks. Before the intervention is implemented, the clinical performance of students in two medical schools is assessed as a baseline. Data on clinical performance in both medical schools after 3 weeks – when the intervention is made – are collected to see the effect of the simulation on clinical performance.

The second quasi-experimental design is the one group pre-test-post-test design. Here, as in the previous example, the educator assesses the knowledge and clinical performance of all students before simulation training. This time, however, all students are exposed to simulation training. The difference between the pre-post test scores may be an indication of the change in the use of simulators.

As groups (the experimental and control/comparison groups) are not randomly allocated, it is impossible to say the groups are equivalent at the beginning of the study. Hence, the study findings of quasi-experimental designs are less decisive in comparison with experimental design. In the other words, if the clinical performance of students in groups is not equivalent, the effects of the intervention will not be clear.

Sometimes researchers are not able to randomise participants into groups or they cannot manipulate the independent

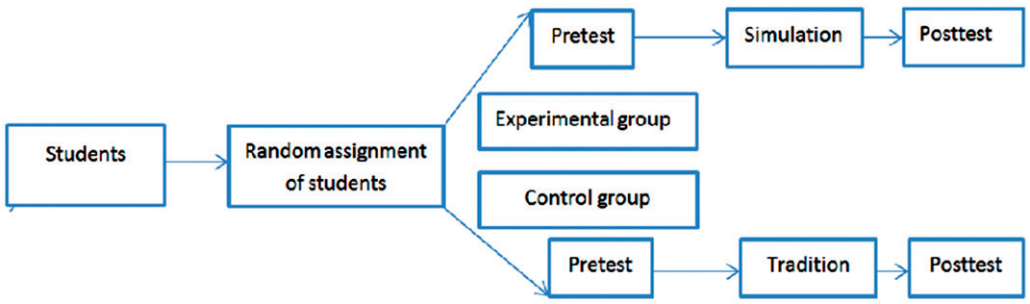


Figure 2. Randomised control-group pre-test–post-test design.

variable in order to observe its effect on the dependent variable. This could be due to ethical considerations or to factors out of the control of the researcher. For example, consider gender as an independent variable; if researchers want to compare female and male students about a particular phenomenon, they cannot manipulate gender and they cannot randomly assign students to be either female or male. When the researcher has no control on the independent variable, the study is non-experimental (sometimes called an observation design). This design is widely used in medical education research. There are different non-experimental designs, but three study designs which are commonly used in medical education research will be discussed: correlational studies, cross-sectional studies and longitudinal studies.

Sometimes when researchers wish to establish a relationship between the variables in their study, but they cannot design an experimental or quasi-experimental study, they plan for correlational studies. In correlational studies, researchers make a claim about the relationship between variables in theories or models. For example, what is the relationship between student ability and their score on the UK Clinical Aptitude Test (UKCAT)? By calculating the correlation between students marks and UKCAT scores, researchers can address the association between student ability and UKCAT.

In cross-sectional studies, researchers collect data at one point in time or over a short period (Kevin 2006). In this design, independent and dependent variables are identified in a given population and then the associations between them are determined. For example, medical educators may be interested to determine the association between medical student year and empathy. If the researchers have knowledge of the medical student year and the empathy scores of students, they will be able to identify the relationship using statistical procedures. In epidemiological studies, retrospective studies or prevalence studies are usually cross-sectional. Data on the independent and dependent (outcome) variables are collected simultaneously (Polit & Beck 2014).

Sometimes study participants are followed over time and data are collected at multiple follow-up times. This is called a longitudinal study. In other words, the same participants are measured 'at each point of the time scale'. Such studies are always concerned with individual change (Goldstein 1968). As an example, a longitudinal study 'was designed to examine changes in medical students' empathy during medical school and to determine when the most significant changes occurs' (Hojat et al. 2009). In epidemiological studies, prospective studies or incidence studies are usually longitudinal. Cohort studies are longitudinal studies which involve a large sample size. The interested reader is referred to the book 'A study guide to epidemiology and biostatistics' for a discussion of the epidemiological studies (Hebel & McCarter 2012).

Conclusions

This part of the Guide has explained how quantitative and qualitative methods can be used in medical education research to produce new knowledge. From a learning perspective, therefore, medical educators should learn techniques that are used in both quantitative and qualitative methods. Although the philosophical assumptions of quantitative research differ from qualitative research, they certainly do not contradict each other, but they are complementary. From a quantitative perspective, a concept is observable and measurable, and is analysed using statistical procedures. From a qualitative point of view, the phenomenon of interest does not quantify, but the qualitative researcher provides a rich description of the phenomenon to be studied. In quantitative studies, the research process is linear and deductive, whereas in qualitative studies the research process is recursive and inductive. In this part, we have also explained some essential steps in the research process and quantitative methods.

Notes on Contributors

MOHSEN TAVAKOL, PhD, MClinEd, is a Lecturer in Psychometrics. His main interests are in medical education assessment, psychometric analysis (Classical Test Theory, Generalisability theory, Item Response Theory Models), robust statistical methods, multivariate statistics, quantitative and qualitative research methods and communication skills.

JOHN SANDARS, MD, MSc, FRCGP, FAcadMEd, MRCP, Cert Ed., is an Associate Professor at the Leeds Institute of Medical Education, University of Leeds, UK. His main interest is in developing teaching and learning by a scholarship approach, in which real life problems are investigated by both quantitative and qualitative research.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

References

Ajzen I. 2005. Attitudes, personality, and behaviour. New York: Open University Press/MaGraw Hill.

Alvesson M, Skoldberg K. 2009. Reflexive methodology: New vistas for qualitative research. Thousand Okas: CA, SAGE.

Ary D, Jacobs L, Sorensen C, Razavieh A. 2006. Introduction to research in education. Belmont, CA: Wasdsworth.

Ashley P, Boyd B. 2006. Quantitative and qualitative approaches to research in environmental management. Australasian J Environ Manage 13:70–78.

Atkinson P, Pugsley L. 2005. Making sense of ethnography and medical education. Med Educ 39:228–234.

Aveyard H. 2010. Doing a literature review in health and social science care: A practical guide. Glasgow: The McGrawHill Companies.

Bower E, Scambler S. 2007. The contributions of qualitative research towards dental public health practice. Community Dentistry and Oral Epidemiology 35:161–9.

Brown C. 2010. Conceptualizing research. In: Aparasus R, editor. Research methods for pharmaceutical practice and policy. London: Pharmaceutical Press. pp 17–35.

Brown S. 2014. Evidence-based nursing. Burlington, MA: Jones & Bartlett Learning.

Buckley G. 1998. Partial truths-research papers in medical education. Med Edu 32:1–2.

Cochrane A. 1972. Effectiveness and efficiency: Random reflection on health services. London: Nuffield Prvincial Hospitals Trus (original publication, 1972).

Cohen L, Manion L, Morrison K. 2008. Research methods in education. London: Routledge.

Creswell J. 2013. Research design: Qualitative, quantitative, and mixed methods approaches. Thousands Okas: SAGE.

Creswell J. 2014. Research design: qualitative, quantitative, and mixed methods approaches. Thousand Oaks: SAGE.

Crotty M. 1998. The foundations of social research: meaning and perspective in the research process. London: SAGE.

Daniel W. 2005. Biostatistics. Hoboken, NJ: Wiley.

Denzin N, Lincoln Y. 2011. The discipline and practice of qualitative research. In: Denzin N, Lincoln Y, editors. The SAGE handbook of qualitative research. Thousands Oaks: SAGE. pp 1–19.

Dingwall R, Murphy E, Watson P, Greathbatch D, Parker S. 1998. Catching gold fish: quality in qualitative research. J Health Serv Res Policy 3: 167–172.

Erlandson D, Harris E, Skipper B, Allen S. 1993. Doing naturalistic inquiry. London: SAGE.

Fereday J, Muir-Chocrane E. 2006. Demonstrating rigour using thematic analysis: Hybrid approach of inductive and deductive coding and theme development [Online]. [Accessed July 2012] Available from http://www.ualberta.ca/~iiqm/backissues/5_1/pdf/fereday.pdf.

Gillis A, Jackson W. 2002. Research for nurses: Methods and interpretation. Philadelphia: F.A. Davis Company.

Goldstein H. 1968. Longitudinal studies and the measurement of change. Statistician 18:93–117.

Griffin K, Museus S. 2011. Application of mixed-methods approaches to higher education and intersectional analyses. In: Griffin K, Museus S, editors. Using mixed -methods approaches to study intersecrionality in higher education. Ann Arbor: Wiley. pp 15–26.

Guba E. 1990. The alternative paradigm dialog. In: Guba E, editor. The Paradigm dialoge. London: Sage Publication. pp 15–27.

Halloway I, Biley F. 2011. Being a qualitative researcher. Qual Health Res 21:968–975.

Harris A, Mcgregor J, Perencevich E, Furuno J, Zhu J, Peterson D, Finkeistein J. 2006. The use and interpretation of quasi-experimental studies in medical informatics. J Am Med Inform 13:16–23.

Hart C. 2005. Doing a literature review. London. SAGE.

Hebel J, McCarter R. 2012. A study guide to epidemiology and biostatistics. Bulington, MA: Jones and Bartlett.

Higgins J, Green S. 2011. Cochrane handbook for systematic reviews of interventions [Online]. The Cochrane Collaboration. [Accessed November 2013] Available from http://handbook.cochrane.org/index.htm#chapter_20/20_qualitative_research_and_cochrane_reviews.htm.

Hojat M, Maxwell K, Brainard G, Herrine S, Isenberg G, Veloski J, Gonnella J. 2009. The devil is the third year: A longitudinal study of erosion of empathy in medical school. Acad Med 84:1182–91.

Kaufim D. 2003. Applying educational theories in practice. BMJ 25:213–216.

Kawulich B. 2009. The role of theory in research. In: Garner M, Wagner C, Kawulich B, editors. Teaching research methods in social sciences. Surrey: Aghgate Publishing Limited.

Kerlinger F. 1970. Foundations of behavioural research. New Yourk: Holt, Rinehart & Winston.

Kevin K. 2006. Study design III: Cross-sectional studies. Evidence-based Dentistry 7:24–25.

Khatami S, Macentee M, Pratt D, Collins J. 2012. Clinical reasoning in dentistry: A conceptual framework for dental education. J Dent Educ 76:1116–1128.

Leedy P, Ormrod J. 2005. Practical research planning and design. New Jersey: Pearson Merrill Prentice Hall.

Liehr P, Smith M. 2002. Theoretical frameworks. In: Lobiondo-Wood G, Haber J, editors. Nursing research: Methods, critical appraisal and utilization. St. Louis: Mosby. pp 107–120.

Lincoln Y, Guba E. 1985. Naturalistic inquiry. Beverly Hills, CA: SAGE.

Lodico M, Spaulding D, Voegtle K. 2010. Methods in educational research: from theory to practice. San Faransico: John Wiley & Sons.

Marshall C, Rossman G. 2006. Designing qualitative research. Thousand Oaks: Sage Publication.

Marshall I, Wolfe C, Mckevitt C. 2012. Lay perspectives on hypertension and drug adherence: systematic review of qualitative research. BMJ 344: e3953.

Mays N, Pope C. 2000. Assessing quality in qualitative research. BMJ 320: 50–52.

McBurney D, White T. 2010. Research methods. Bamont, CA: Wadsworth.

McEwan M, Espie C, Metcalfe J. 2004. A systematic review of the contribution of qualitative research to the study of quality of life in children and adolescents with epilepsy. Seizure 13:3–14.

McInnes E, Seers K, Tutton L. 2011. Older people's views in relation to risk of falling the need for intervention: A meta–ethnography. J Adv Nurs 67:2525–2536.

Merriam-Webster. 2013. Research [Online]. [Accessed 1 December 2013] Available from http://www.merriam-webster.com/dictionary/research.

Mertens D. 2010. Research and evaluation in education and psychology. London: SAGE.

Moan I, Rise J. 2005. Quitting smoking: Applying and extended version of the theory of planned behaviour to predict intention and behaviour. J Appl Biobehav Res 10:39–68.

Morse J. 2005. What is qualitative research? Qual Health Res 15: 859–860.

Morse J. 2006. Reconceptualising qualitative evidence. Qual Health Res 16: 415–422.

Morse J. 2011. What is qualitative health research. In: Denzin N, Lincoln Y, editors). SAGE handbook of qualitative research. Thousand Oaks: SAGE. pp 401–414.

Morse J, Field P. 1995. Qualitative research methods for health professionals. Thousand Oaks: SAGE.

Neuman W. 2003. Social research methods: Qualitative and quantitative approaches. Boston: Allyan and Bacon.

Newman I, Hitchcock J. 2011. Underlying agreements between qualitative and quantitative research: the art short and tall of it all. Human Recourse Dev Rev 10:381–394.

Nieswiadomy R. Foundations of nursing research. Stamford: Appelton & Lange.

Norman G. 2002. Research in medical education: three decades of progress. BMJ 324:1560–2.

Ong B, Richardson J. 2006. The contribution of qualitative approaches to musculoskeletal research. Rheumatology 45:360–70.

Polit D, Beck C. 2014. Essentials of nursing research. Philadelphia: Wolters Kluwer|Lippincott Williams & Wilkins.

Reeves S, Kuper A, Hodges B. 2008. Qualitative research methodologies: Ethnography. BMJ 337:a1020.

Ritchie J, Lewis J, Cnicholls C, Ormston R. 2013. Qualitative research practice: A guide for social science students and researchers. London: SAGE.

Rothstein W, Phuong L. 2007. Ethical attitudes of nurse, physician, and unaffiliated members of institutional review boards. J Nurs Scholarship 39:75–81.

Rubib H, Rubin I. 2012. Qualitative interviewing: the art of hearing data. London: SAGE.

Sclater M. 2012. Theorizing from bricolage: researching collaboration in art and design education. In: Adams J, Cochrane M, Dunne L, editors. Applying theory to educational research: An introductory approaches with case studies. Oxford: Wiley—Blackwell. pp 157–176.

Shadish E, Cook T, Campbell D. 2002. Experimental and quasi-experimental designs for generalised casual inference. Boston: Houghoton Mifflin.

Silverman D. 2010. Doing qualitative research. London, SAGE.

Smith K. 1990. Alternative research paradigms and the problem of criteria. In: Guba E, editor. The paradigms dialogue. Newbury Park, CA: SAGE. pp 167–1187.

Tashakkori A, Teddlie C. 1998. Mixed methodology: Combining qualitative and quantitative approaches. London: SAGE.

Tavakol M, Gruppen L. 2010. Using evaluation research to improve medical education. Clin Teach 7:192–196.

Tavakol M, Murphy R, Rahemei-Madeseh M, Torabi S. 2008. The involvement of clinicians in medical education research. Qual Prim Care 16:335–40.

Tavakol S, Dennick R, Tavakol M. 2012. Medical students' understanding of empathy: a phenomenological study. Med Edu 46:306–316.

Taylor B. 2013. Mixed method research. In: Taylor B, Francis K, editors. Qualitative research in the health sciences: Methodologies, methods and processes. New York: Routledge. pp 162–176.

Toussaint L, Webb J. 2005. Gender differences in the relationship between empathy and forgiveness. J Soc Psychol 14:673–685.

Trice L, Bloom K. 2014. PICOT, problem statement, research ouestion, hypothesis. In: Boswell C, Cannon S, editors. Introduction to nursing research. Burlington, MA: Jones & Bartlett.

Suggested readings

Alasuutari P, Bickman L, Brannen J. 2008. The SAGE handbook of social research methods. London: SAGE.

Bryant A, Charmaz K. 2010. The SAGE handbook of grounded theory (Paperback Edition). London: SAGE.

Corbin J, Strauss A. 2008. Basics of qualitative research: Grounded theory procedures and techniques. Thousand Oaks: SAGE.

Creswell J, Clark V. 2011. Designing and conducting mixed methods research. London: SAGE.

Creswell J. 2011. Educational research: Planning, conducting, and evaluating quantitative and qualitative research. Boston: Pearson.

Denzin N, Lincoln Y. 2011. The SAGE handbook of qualitative research. London: SAGE.

Flick U. 2013. The SAGE handbook of qualitative data analysis. London: SAGE

Grove M, Overton T. 2011. Getting started in pedagogic research within the STEM disciplines [Online]. The University of Birmingham on behalf of the National HE STEM Programme. [Accessed 1 February 2014] Available from http://www.birmingham.ac.uk/Documents/college-eps/college/stem/getting-started-in-stem-pedagogic-optimised.pdf.

Gruppen L. 2008. Is medical education research 'hard' or 'soft' research? Adv Health Sci Educ Theory Pract 13:1–2.

Harland N, Holey H. 2011. Including open-ended questions in quantitative questionnaires. Int J Theory Rehab 18:482–486.

Johnson B, Christnsen L. 2010. Educational research: Quantitative, qualitative, and mixed approaches. London: SAGE Publications.

Morse J. 2004. Qualitative comparison: Appropriateness, equivalence, and fit Qual Health Res 14:1323–1325.

Morse J. 2009. Mixed method design: principles and procedures. Walnut Creek: Left Coast Press.

Morse J, Dimitroff L, Harper R, Koontz A, Kumra S, Matthew-Maich N, Mihas PCM. 2011. Considering the qualitative-quantitative language divide. Qual Health Res 21:1302–1303.

Norman G, Eva K. 2014. Quantitative research methods in medical education. In: Swanwick T, editor. Understanding medical education: Evidence, theory and practice. Sussex: Wiley-Blackwell. pp 349–369.

Sandelowski M. 2004. Counting cats in Zanzibar. Res Nurs Health 27:215–216.

Silverman D. 2006. Interpreting qualitative data: Methods for analysing talk, text and interaction. London: SAGE.

Tavakol M, Wilcox R. 2013. Medical education research: the application of robust statistical methods. Int J Med Educ 4:93–95.

Teddile C, Tashakkori A. 2009. Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioural. London: SAGE.

Thomas G. 2013. How to do your research project: A guide for students in education and applied social sciences. London: SAGE

Trochim W, Donnely J. 2006. The research methods knowledge base. Ohio: Atomic Dog.

Web Centre for social research methods [Online]. [Accessed 1 February 2014]Available from http://socialresearchmethods.-net/.

Wilcox R. 2011. Modern statistics for the social and behavioural sciences: A practical introduction. Boca Raton: CRC Press.

Medical Teacher

ISSN: 0142-159X (Print) 1466-187X (Online) Journal homepage: https://www.tandfonline.com/loi/imte20

Developing questionnaires for educational research: AMEE Guide No. 87

Anthony R. Artino Jr., Jeffrey S. La Rochelle, Kent J. Dezee & Hunter Gehlbach

To cite this article: Anthony R. Artino Jr., Jeffrey S. La Rochelle, Kent J. Dezee & Hunter Gehlbach (2014) Developing questionnaires for educational research: AMEE Guide No. 87, Medical Teacher, 36:6, 463-474, DOI: 10.3109/0142159X.2014.889814

To link to this article: https://doi.org/10.3109/0142159X.2014.889814

Published online: 24 Mar 2014.

2014, 36: 463–474 MEDICAL TEACHER

AMEE GUIDE

Developing questionnaires for educational research: AMEE Guide No. 87

ANTHONY R. ARTINO, JR.[1], JEFFREY S. LA ROCHELLE[1], KENT J. DEZEE[1] & HUNTER GEHLBACH[2]
[1]Uniformed Services University of the Health Sciences, USA, [2]Harvard Graduate School of Education, USA

Abstract

In this AMEE Guide, we consider the design and development of self-administered surveys, commonly called questionnaires. Questionnaires are widely employed in medical education research. Unfortunately, the processes used to develop such questionnaires vary in quality and lack consistent, rigorous standards. Consequently, the quality of the questionnaires used in medical education research is highly variable. To address this problem, this AMEE Guide presents a systematic, seven-step process for designing high-quality questionnaires, with particular emphasis on developing survey scales. These seven steps do not address all aspects of survey design, nor do they represent the only way to develop a high-quality questionnaire. Instead, these steps synthesize multiple survey design techniques and organize them into a cohesive process for questionnaire developers of all levels. Addressing each of these steps systematically will improve the probabilities that survey designers will accurately measure what they intend to measure.

Introduction: Questionnaires in medical education research

Surveys are used throughout medical education. Examples include the ubiquitous student evaluation of medical school courses and clerkships, as well as patient satisfaction and student self-assessment surveys. In addition, survey instruments are widely employed in medical education research. In our recent review of original research articles published in *Medical Teacher* in 2011 and 2012, we found that 37 articles (24%) included surveys as part of the study design. Similarly, surveys are commonly used in graduate medical education research. Across the same two-year period (2011–2012), 75% of the research articles published in the *Journal of Graduate Medical Education* used surveys.

Despite the widespread use of surveys in medical education, the medical education literature provides limited guidance on the best way to design a survey (Gehlbach et al. 2010). Consequently, many surveys fail to use rigorous methodologies or "best practices" in survey design. As a result, the reliability of the scores that emerge from surveys is often inadequate, as is the validity of the scores' intended interpretation and use. Stated another way, when surveys are poorly designed, they may fail to capture the essence of what the survey developer is attempting to measure due to different types of measurement error. For example, poor question wording, confusing question layout and inadequate response options can all affect the reliability and validity of the data from surveys, making it extremely difficult to draw useful conclusions (Sullivan 2011). With these problems as a backdrop, our purpose in this AMEE Guide is to describe a systematic process for developing and collecting reliability and validity evidence

Practice points

- Questionnaires are widely used in medical education research, yet the processes employed to develop questionnaires vary in quality and lack consistent, rigorous standards.
- This AMEE Guide introduces a systematic, seven-step design process for creating high-quality survey scales fit for program evaluation and research purposes.
- The seven-step design process synthesizes multiple techniques survey designers employ into a cohesive process.
- The survey design process described in this Guide includes the following seven steps: (1) conduct a literature review, (2) carry out interviews and/or focus groups, (3) synthesize the literature review and interviews/focus groups, (4) develop items, (5) collect feedback on the items through an expert validation, (6) employ cognitive interviews to ensure that respondents understand the items as intended and (7) conduct pilot testing.
- This seven-step design process differs from previously described processes in that it blends input from other experts in the field as well as potential participants. In addition, this process front loads the task of establishing validity by focusing heavily on careful item development.

Correspondence: Anthony R. Artino, Jr., PhD, Associate Professor of Preventive Medicine and Biometrics, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Road, Bethesda, MD 20814-4712, USA. Tel: +1-301.295.3693; E-mail: anthony.artino@usuhs.edu

ISSN 0142-159X print/ISSN 1466-187X online/14/60463–474 © 2014 Informa UK Ltd.
DOI: 10.3109/0142159X.2014.889814

for survey instruments used in medical education and medical education research. In doing so, we hope to provide medical educators with a practical guide for improving the quality of the surveys they design for evaluation and research purposes.

A systematic, seven-step process for survey scale design

The term "survey" is quite broad and could include the questions used in a phone interview, the set of items employed in a focus group and the questions on a self-administered patient survey (Dillman et al. 2009). Although the processes described in this AMEE Guide can be used to improve all of the above, we focus primarily on self-administered surveys, which are often referred to as questionnaires. For most questionnaires, the overarching goals are to develop a set of items that every respondent will interpret the same way, respond to accurately and be willing and motivated to answer. The seven steps depicted in Table 1, and described below, do not address all aspects of survey design nor do they represent the only way to develop a high-quality questionnaire. Rather, these steps consolidate and organize the plethora of survey design techniques that exist in the social sciences and guide questionnaire developers through a cohesive process. Addressing each step systematically will optimize the quality of medical education questionnaires and improve the chances of collecting high-quality survey data.

Questionnaires are good for gathering data about abstract ideas or concepts that are otherwise difficult to quantify, such as opinions, attitudes and beliefs. In addition, questionnaires can be useful for collecting information about behaviors that are not directly observable (e.g. studying at home), assuming respondents are willing and able to report on those behaviors. Before creating a questionnaire, however, it is imperative to first decide if a survey is the best method to address the research question or construct of interest. A *construct* is the model, idea or theory that the researcher is attempting to assess. In medical education, many constructs of interest are not directly observable – student satisfaction with a new curriculum, patients' ratings of their physical discomfort, etc. Because documenting these phenomena requires measuring people's perceptions, questionnaires are often the most pragmatic approach to assessing these constructs.

In medical education, many constructs are well suited for assessment using questionnaires. However, because psychological, non-observable constructs such as teacher motivation, physician confidence and student satisfaction do not have a commonly agreed upon metric, they are difficult to measure with a single item on a questionnaire. In other words, for some constructs such as weight or distance, most everyone agrees upon the units and the approach to measurement, and so a single measurement may be adequate. However, for non-observable, psychological constructs, a survey scale is often required for more accurate measurement. Survey scales are groups of similar items on a questionnaire designed to assess the same underlying construct (DeVellis 2003). Although scales are more difficult to develop and take longer to complete, they offer researchers many advantages. In particular, scales more completely, precisely and consistently assess the underlying construct (McIver & Carmines 1981). Thus, scales are commonly used in many fields, including medical education, psychology and political science. As an example, consider a medical education researcher interested in assessing medical student satisfaction. One approach would be to simply ask one question about satisfaction (e.g. How satisfied were you with medical school?). A better approach, however, would be to ask a series of questions designed to capture the different facets of this satisfaction construct (e.g. How satisfied were you with the teaching facilities? How effective were your instructors? and How easy was the scheduling process?). Using this approach, a mean score of all the items within a particular scale can be calculated and used in the research study.

Because of the benefits of assessing these types of psychological constructs through scales, the survey design process that we now turn to will focus particularly on the development of scales.

Step 1: Conduct a literature review

The first step to developing a questionnaire is to perform a literature review. There are two primary purposes for the literature review: (1) to clearly define the construct and (2) to determine if measures of the construct (or related constructs) already exist. A review of the literature helps to ensure the

Table 1. A seven-step, survey scale design process for medical education researchers.

Step	Purpose
1. Conduct a literature review	To ensure that the construct definition aligns with relevant prior research and theory and to identify existing survey scales or items that might be used or adapted
2. Conduct interviews and/or focus groups	To learn how the population of interest conceptualizes and describes the construct of interest
3. Synthesize the literature review and interviews/focus groups	To ensure that the conceptualization of the construct makes theoretical sense to scholars in the field and uses language that the population of interest understands
4. Develop items	To ensure items are clear, understandable and written in accordance with current best practices in survey design
5. Conduct expert validation	To assess how clear and relevant the items are with respect to the construct of interest
6. Conduct cognitive interviews	To ensure that respondents interpret items in the manner that survey designer intends
7. Conduct pilot testing	To check for adequate item variance, reliability and convergent/discriminant validity with respect to other measures

Adapted with permission from Lippincott Williams and Wilkins/Wolters Kluwer Health: Gehlbach et al. (2010). AM last page: Survey development guidance for medical education researchers. Acad Med 85:925.

construct definition aligns with related theory and research in the field, while at the same time helping the researcher identify survey scales or items that could be used or adapted for the current purpose (Gehlbach et al. 2010).

Formulating a clear definition of the construct is an indispensable first step in any validity study (Cook & Beckman 2006). A good definition will clarify how the construct is positioned within the existing literature, how it relates to other constructs and how it is different from related constructs (Gehlbach & Brinkworth 2011). A well-formulated definition also helps to determine the level of abstraction at which to measure a given construct (the so-called "grain size", as defined by Gehlbach & Brinkworth 2011). For example, to examine medical trainees' confidence to perform essential clinical skills, one could develop scales to assess their confidence to auscultate the heart (at the small-grain end of the spectrum), to conduct a physical exam (at the medium-grain end of the spectrum) or to perform the clinical skills essential to a given medical specialty (at the large-grain end of the spectrum).

Although many medical education researchers prefer to develop their own surveys independently, it may be more efficient to adapt an existing questionnaire – particularly if the authors of the existing questionnaire have collected validity evidence in previous work – than it is to start from scratch. When this is the case, a request to the authors to adapt their questionnaire will usually suffice. It is important to note, however, that the term "previously validated survey" is a misnomer. The validity of the scores that emerge from a given questionnaire or survey scale is sensitive to the survey's target population, the local context and the intended use of the scale scores, among other factors. Thus, survey developers collect reliability and validity evidence for their survey scales in a specified context, with a particular sample, and for a particular purpose.

As described in the *Standards for Educational and Psychological Testing*, validity refers to the degree to which evidence and theory support a measure's intended use (AERA, APA, & NCME 1999). The process of validation is the most fundamental consideration in developing and evaluating a measurement tool, and the process involves the accumulation of evidence across time, settings and samples to build a scientifically sound validity argument. Thus, establishing validity is an ongoing process of gathering evidence (Kane 2006). Furthermore, it is important to acknowledge that reliability and validity are not properties of the survey instrument, *per se*, but of the survey's scores and their interpretations (AERA, APA, & NCME 1999). For example, a survey of trainee satisfaction might be appropriate for assessing aspects of student well-being, but such a survey would be inappropriate for selecting the most knowledgeable medical students. In this example, the survey did not change, only the score interpretation changed (Cook & Beckman 2006).

Many good reasons exist to use, or slightly adapt, an existing questionnaire. By way of analogy, we can compare this practice to a physician who needs to decide on the best medical treatment. The vast majority of clinicians do not perform their own comparative research trials to determine the best treatments to use for their patients. Rather, they rely on the published research, as it would obviously be impractical for clinicians to perform such studies to address every disease process. Similarly, medical educators cannot develop their own questionnaires for every research question or educational intervention. Just like clinical trials, questionnaire development requires time, knowledge, skill and a fair amount of resources to accomplish correctly. Thus, an existing, well-designed questionnaire can often permit medical educators to put their limited resources elsewhere.

Continuing with the clinical research analogy, when clinicians identify a research report that is relevant to their clinical question, they must decide if it applies to their patient. Typically, this includes determining if the relationships identified in the study are causal (internal validity) and if the results apply to the clinician's patient population (external validity). In a similar way, questionnaires identified in a literature search must be reviewed critically for validity evidence and then analyzed to determine if the questionnaire could be applied to the educator's target audience. If survey designers find scales that closely match their construct, context and proposed use, such scales might be useable with only minor modification. In some cases, the items themselves might not be well written, but the content of the items might be helpful in writing new items (Gehlbach & Brinkworth 2011). Making such determinations will be easier the more the survey designer knows about the construct (through the literature review) and the best practices in item writing (as described in Step 4).

Step 2: Conduct interviews and/or focus groups

Once the literature review has shown that it is necessary to develop a new questionnaire, and helped to define the construct, the next step is to ascertain whether the conceptualization of the construct matches how prospective respondents think about it (Gehlbach & Brinkworth 2011). In other words, do respondents include and exclude the same features of the construct as those described in the literature? What language do respondents use when describing the construct? To answer these questions and ensure the construct is defined from multiple perspectives, researchers will usually want to collect data directly from individuals who closely resemble their population of interest.

To illustrate this step, another clinical analogy might be helpful. Many clinicians have had the experience of spending considerable time developing a medically appropriate treatment regimen but have poor patient compliance with that treatment (e.g. too expensive). The clinician and patient then must develop a new plan that is acceptable to both. Had the patient's perspective been considered earlier, the original plan would likely have been more effective. Many clinicians have also experienced difficulty treating a patient, only to have a peer reframe the problem, which subsequently results in a better approach to treatment. A construct is no different. To this point, the researcher developing the questionnaire, like the clinician treating the patient, has given a great deal of thought to defining the construct. However, the researcher unavoidably brings his/her perspectives and biases to this

definition, and the language used in the literature may be technical and difficult to understand. Thus, other perspectives are needed. Most importantly, how does the target population (the patient from the previous example) conceptualize and understand the construct? Just like the patient example, these perspectives are sometimes critical to the success of the project. For example, in reviewing the literature on student satisfaction with medical school instruction, a researcher may find no mention of the instructional practice of providing students with video or audio recordings of lectures (as these practices are fairly new). However, in talking with students, the researcher may find that today's students are accustomed to such practices and consider them when forming their opinions about medical school instruction.

In order to accomplish Step 2 of the design process, the survey designer will need input from prospective respondents. Interviews and/or focus groups provide a sensible way to get this input. Irrespective of the approach taken, this step should be guided by two main objectives. First, researchers need to hear how participants talk about the construct in their own words, with little to no prompting from the researcher. Following the collection of unprompted information from participants, the survey designers can then ask more focused questions to evaluate if respondents agree with the way the construct has been characterized in the literature. This procedure should be repeated until saturation is reached; this occurs when the researcher is no longer hearing new information about how potential respondents conceptualize the construct (Gehlbach & Brinkworth 2011). The end result of these interviews and/or focus groups should be a detailed description of how potential respondents conceptualize and understand the construct. These data will then be used in Steps 3 and 4.

Step 3: Synthesize the literature review and interviews/focus groups

At this point, the definition of the construct has been shaped by the medical educator developing the questionnaire, the literature and the target audience. Step 3 seeks to reconcile these definitions. Because the construct definition directs all subsequent steps (e.g. development of items), the survey designer must take care to perform this step properly.

One suitable way to conduct Step 3 is to develop a comprehensive list of indicators for the construct by merging the results of the literature review and interviews/focus groups (Gehlbach & Brinkworth 2011). When these data sources produce similar lists, the process is uncomplicated. When these data are similar conceptually, but the literature and potential respondents describe the construct using different terminology, it makes sense to use the vocabulary of the potential respondents. For example, when assessing teacher confidence (sometimes referred to as teacher self-efficacy), it is probably more appropriate to ask teachers about their "confidence in trying out new teaching techniques" than to ask them about their "efficaciousness in experimenting with novel pedagogies" (Gehlbach et al. 2010). Finally, if an indicator is included from one source but not the other, most questionnaire designers will want to keep the item, at least initially. In later steps, designers will have opportunities to determine, through expert reviews (Step 5) and cognitive interviews (Step 6), if these items are still appropriate to the construct. Whatever the technique used to consolidate the data from Steps 1 and 2, the final definition and list of indicators should be comprehensive, reflecting both the literature and the opinions of the target audience.

It is worth noting that scholars may have good reasons to settle on a final construct definition that differs from what is found in the literature. However, when this occurs, it should be clear exactly how and why the construct definition is different. For example, is the target audiences' perception different from previous work? Does a new educational theory apply? Whatever the reason, this justification will be needed for publication of the questionnaire. Having an explicit definition of the construct, with an explanation of how it is different from other versions of the construct, will help peers and researchers alike decide how to best use the questionnaire both in comparison with previous studies and with the development of new areas of research.

Step 4: Develop items

The goal of this step is to write survey items that adequately represent the construct of interest in a language that respondents can easily understand. One important design consideration is the number of items needed to adequately assess the construct. There is no easy answer to this question. The ideal number of items depends on several factors, including the complexity of the construct and the level at which one intends to assess it (i.e. the grain size). In general, it is good practice to develop more items than will ultimately be needed in the final scale (e.g. developing 15 potential items in the hopes of ultimately creating an eight-item scale), because some items will likely be deleted or revised later in the design process (Gehlbach & Brinkworth 2011). Ultimately, deciding on the number of items is a matter of professional judgment, but for most narrowly defined constructs, scales containing from 6 to 10 items will usually suffice in reliably capturing the essence of the phenomenon in question.

The next challenge is to write a set of clear, unambiguous items using the vocabulary of the target population. Although some aspects of item-writing remain an art form, an increasingly robust science and an accumulation of best practices should guide this process. For example, writing questions rather than statements, avoiding negatively worded items and biased language, matching the item stem to the response anchors and using response anchors that emphasize the construct being measured rather than employing general agreement response anchors (Artino et al. 2011) are all well-documented best practices. Although some medical education researchers may see these principles as "common sense", experience tells us that these best practices are often violated.

Reviewing all the guidelines for how best to write items, construct response anchors and visually design individual survey items and entire questionnaires is beyond the scope of this AMEE Guide. As noted above, however, there are many excellent resources on the topic (e.g. DeVillis 2003; Dillman et al. 2009; Fowler 2009). To assist readers in grasping some of

the more important and frequently ignored best practices, Table 2 presents several item-writing pitfalls and offers solutions.

Another important part of the questionnaire design process is selecting the response options that will be used for each item. Closed-ended survey items can have unordered (nominal) response options that have no natural order or ordered (ordinal) response options. Moreover, survey items can ask respondents to complete a ranking task (e.g. "rank the following items, where 1 = best and 6 = worst") or a rating task that asks them to select an answer on a Likert-type response scale. Although it is outside the scope of this AMEE Guide to review all of the response options available, questionnaire designers are encouraged to tailor these options to the construct(s) they are attempting to assess (and to consult one of the many outstanding resources on the topic; e.g. Dillman et al. 2009; McCoach et al. 2013). To help readers understand some frequently ignored best practices Table 2 and Figure 1 present several common mistakes designers commit when writing and formatting their response options. In addition, because Likert-type response scales are by far the most popular way of collecting survey responses – due, in large part, to their ease of use and adaptability for measuring many different constructs (McCoach et al. 2013) – Table 3 provides several examples of five- and seven-point response scales that can be used when developing Likert-scaled survey instruments.

Once survey designers finish drafting their items and selecting their response anchors, there are various sources of evidence that might be used to evaluate the validity of the questionnaire and its intended use. These sources of validity have been described in the *Standards for Educational and Psychological Testing* as evidence based on the following: (1) content, (2) response process, (3) internal structure, (4) relationships with other variables and (5) consequences (AERA, APA & NCME 1999). The next three steps of the design process fit nicely into this taxonomy and are described below.

Step 5: Conduct expert validation

Once the construct has been defined and draft items have been written, an important step in the development of a new questionnaire is to begin collecting validity evidence based on the survey's content (so-called *content validity*) (AERA, APA & NCME 1999). This step involves collecting data from content experts to establish that individual survey items are relevant to the construct being measured and that key items or indicators have not been omitted (Polit & Beck 2004; Waltz et al. 2005). Using experts to systematically review the survey's content can substantially improve the overall quality and representativeness of the scale items (Polit & Beck 2006).

Steps for establishing content validity for a new survey instrument can be found throughout the literature (e.g. McKenzie et al. 1999; Rubio et al. 2003). Below, we summarize several of the more important steps. First, before selecting a panel of experts to evaluate the content of a new questionnaire, specific criteria should be developed to determine who qualifies as an expert. These criteria are often based on experience or knowledge of the construct being measured, but, practically speaking, these criteria also are dependent on the willingness and availability of the individuals being asked to participate (McKenzie et al. 1999). One useful approach to finding experts is to identify authors from the reference lists of the articles reviewed during the literature search. There is no consensus in the literature regarding the number of experts that should be used for content validation; however, many of the quantitative techniques used to analyze expert input will be impacted by the number of experts employed. Rubio et al. (2003) recommends using 6–10 experts, while acknowledging that more experts (up to 20) may generate a clearer consensus about the construct being assessed, as well as the quality and relevance of the proposed scale items.

In general, the key domains to assess through an expert validation process are representativeness, clarity, relevance and distribution. Representativeness is defined as how completely the items (as a whole) encompass the construct, clarity is how clearly the items are worded and relevance refers to the extent each item actually relates to specific aspects of the construct. The distribution of an item is not always measured during expert validation as it refers to the more subtle aspect of how "difficult" it would be for a respondent to select a high score on a particular item. In other words, an average medical student may find it very difficult to endorse the self-confidence item, "How confident are you that you can get a 100% on your anatomy exam", but that same student may find it easier to strongly endorse the item, "How confident are you that you can pass the anatomy exam". In general, survey developers should attempt to have a range of items of varying difficulty (Tourangeau et al. 2000).

Once a panel of experts has been identified, a content validation form can be created that defines the construct and gives experts the opportunity to provide feedback on any or all of the aforementioned topics. Each survey designer's priorities for a content validation may differ; as such, designers are encouraged to customize their content validation forms to reflect those priorities.

There are a variety of methods for analyzing the quantitative data collected on an expert validation form, but regardless of the method used, criterion for the acceptability of an item or scale should be determined in advanced (Beck & Gable 2001). Common metrics used to make inclusion and exclusion decisions for individual items are the content validity ratio, the content validity index and the factorial validity index. For details on how to calculate and interpret these indices, see McKenzie et al. (1999) and Rubio et al. (2003). For a sample content validation form, see Gehlbach & Brinkworth (2011).

In addition to collecting quantitative data, questionnaire designers should provide their experts with an opportunity to provide free-text comments. This approach can be particularly effective for learning what indicators or aspects of the construct are not well-represented by the existing items. The data gathered from the free-text comments and subsequent qualitative analysis often reveal information not identified by the quantitative data and may lead to meaningful additions (or subtractions) to items and scales (McKenzie et al. 1999).

There are many ways to analyze the content validity of a new survey through the use of expert validation. The best approach should look at various domains where the

Table 2. Item-writing ''best practices'' based on scientific evidence from questionnaire design research.

Pitfall	Survey example(s)	Why it's a problem	Solution(s)	Survey example(s)	References
Creating a double-barreled item	– 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, the respondent may talk to his nurses often but talk to administrative staff much less frequently. If this were the case, the respondent would have a difficult time answering the question. Survey items should address one idea at a time.	When you have multiple questions/ premises 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.	– 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?	Tourangeau et al. 2000; Dillman et al. 2009
Creating a negatively worded item	– 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 ''yes'' in order to mean ''no'' (or ''agree'' in order to ''disagree''), the item is flawed.	Make sure ''yes'' means yes and ''no'' means no. This generally means wording items positively.	– In an average week, how many times do you start class on time? – Should the chief resident be responsible for admitting patients?	Dillman et al. 2009
Using statements instead of questions	I am confident I can do well in this course. • Not at all true • A little bit true • Somewhat true • Mostly true • Completely true	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'' (Schwarz 1999). Only rarely do people engage in rating statements in their everyday conversations.	Formulate survey items as questions. Questions are more conversational, more straightforward and easier to process mentally. People are more practiced at responding to them.	How confident are you that you can do well in this course? • Not at all confident • Slightly confident • Moderately confident • Quite confident • Extremely confident	Krosnick 1999; Schwarz 1999; Tourangeau et al. 2000; Dillman et al. 2009
Using agreement response anchors	The high cost of health care is the most important issue in America today. • Strongly disagree • Disagree • Neutral • Agree • Strongly agree	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 options may encourage respondents to think through their responses less thoroughly while completing the survey.	Use construct-specific response anchors that emphasize the construct of interest. Doing so reduces acquiescence and keeps respondents focused on the construct in question; this results in less measurement error.	How important is the issue of high healthcare costs in America today? • Not at all important • Slightly important • Moderately important • Quite important • Extremely important	Krosnick 1999; Tourangeau et al. 2000; Dillman et al. 2009
Using too few or too many response anchors	How useful was your medical school training in clinical decision making? • Not at all useful • Somewhat useful • Very useful	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.	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.	How useful was your medical school training in clinical decision making? • Not at all useful • Slightly useful • Moderately useful • Quite useful • Extremely useful	Weng 2004

Adapted with permission from Lippincott Williams and Wilkins/Wolters Kluwer Health: Artino et al. 2011. AM last page: Avoiding five common pitfalls in survey design. Acad Med 86:1327.

Pitfall	Solution(s)	References
1. Labeling only the end points of your response options 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. Problematic item: How interesting did you find this clinical reasoning course? not at all interesting ○ ○ ○ ○ extremely interesting	**Verbally label each response option.** 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. Improved item: How interesting did you find this clinical reasoning course? ○ not at all interesting ○ slightly interesting ○ moderately interesting ○ quite interesting ○ extremely interesting	Krosnick, 1999
2. Labeling response options with both numbers and verbal labels Because of the additional information respondents must process, including numbers and verbal labels extends response time. 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." Problematic item: How much did you learn in today's workshop? ○ -2 almost nothing ○ -1 a little bit ○ 0 some ○ 1 quite a bit ○ 2 a great amount	**Use only verbal labels** 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. Improved item: How much did you learn in today's workshop? ○ almost nothing ○ a little bit ○ some ○ quite a bit ○ a great amount	Christian et al., 2009; Krosnick, 1999
3. Unequally spacing your 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. Problematic item: How much did you learn from your peers in this course? ○ almost nothing ○ a little bit ○ some ○ quite a bit ○ a great amount	**Maintain equal spacing between response options.** 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. Improved item: How much did you learn from your peers in this course? ○ almost nothing ○ a little bit ○ some ○ quite a bit ○ a great amount	Dillman et al., 2009
4. Placing non-substantive response options together with substantive response options Placing non-substantive 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. Problematic item: How satisfied are you with the quality of the library services? ○ not at all satisfied ○ slightly satisfied ○ moderately satisfied ○ quite satisfied ○ extremely satisfied ○ not applicable	**Use additional space to visually separate non-substantive response options from the substantive options.** Using additional space to visually separate non-substantive response options from substantive options will align the visual midpoint with the conceptual midpoint thereby reducing measurement error. This recommendation is a beneficial exception to the guidance above about maintaining equal spacing between response options. Improved item: How satisfied are you with the quality of the library services? ○ not at all satisfied ○ slightly satisfied ○ moderately satisfied ○ quite satisfied ○ extremely satisfied ○ not applicable	Dillman et al., 2009

Adapted with permission from Lippincott Williams and Wilkins/Wolters Kluwer Health: Artino AR & Gehlbach H (2012). AM last page: Avoiding four visual-design pitfalls in survey development. *Academic Medicine*, 87: 1452.

Figure 1 Visual-design "best practices" based on scientific evidence from questionnaire design research.

Table 3. Examples of various Likert-type response options.

Construct being assessed	Five-point, unipolar response scales	Seven-point, bipolar response scales
Confidence	• Not at all confident • Slightly confident • Moderately confident • Quite confident • Extremely confident	• Completely unconfident • Moderately unconfident • Slightly unconfident • Neither confident nor unconfident (or neutral) • Slightly confident • Moderately confident • Completely confident
Interest	• Not at all interested • Slightly interested • Moderately interested • Quite interested • Extremely interested	• Very uninterested • Moderately uninterested • Slightly uninterested • Neither interested nor uninterested (or neutral) • Slightly interested • Moderately interested • Very interested
Effort	• Almost no effort • A little bit of effort • Some effort • Quite a bit of effort • A great deal of effort	
Importance	• Not important • Slightly important • Moderately important • Quite important • Essential	
Satisfaction	• Not at all satisfied • Slightly satisfied • Moderately satisfied • Quite satisfied • Extremely satisfied	• Completely dissatisfied • Moderately dissatisfied • Slightly dissatisfied • Neither satisfied nor dissatisfied (or neutral) • Slightly satisfied • Moderately satisfied • Completely satisfied
Frequency	• Almost never • Once in a while • Sometimes • Often • Almost always	

researchers have the greatest concerns about the scale (relevance, clarity, etc.) for each individual item and for each set of items or scale. The quantitative data combined with qualitative input from experts is designed to improve the content validity of the new questionnaire or survey scale and, ultimately, the overall functioning of the survey instrument.

Step 6: Conduct cognitive interviews

After the experts have helped refine the scale items, it is important to collect evidence of *response process validity* to assess how prospective participants interpret your items and response anchors (AERA, APA & NCME 1999). One means of collecting such evidence is achieved through a process known as cognitive interviewing or cognitive pre-testing (Willis 2005). Similar to how experts are utilized to determine the content validity of a new survey, it is equally important to determine how potential respondents interpret the items and if their interpretation matches what the survey designer has in mind (Willis 2005; Karabenick et al. 2007). Results from cognitive interviews can be helpful in identifying mistakes respondents make in their interpretation of the item or response options (Napoles-Springer et al. 2006; Karabenick et al. 2007). As a qualitative technique, analysis does not rely on statistical tests of numeric data but rather on coding and interpretation of written notes from the interview. Thus, the sample sizes used for cognitive interviewing are normally small and may involve just 10–30 participants (Willis & Artino 2013). For small-scale medical education research projects, as few as five to six participants may suffice, as long as the survey designer is sensitive to the potential for bias in very small samples (Willis & Artino 2013).

Cognitive interviewing employs techniques from psychology and has traditionally assumed that respondents go through a series of cognitive processes when responding to a survey. These steps include *comprehension* of an item stem and answer choices, *retrieval* of appropriate information from long-term memory, *judgment* based on comprehension of the item and their memory and finally *selection* of a response (Tourangeau et al. 2000). Because respondents can have difficulty at any stage, a cognitive interview should be designed and scripted to address any and all of these potential problems. An important first step in the cognitive interview process is to create coding criteria that reflects the survey creator's intended meaning for each item (Karabenick et al. 2007), which can then be used to help interpret the responses gathered during the cognitive interview.

The two major techniques for conducting a cognitive interview are the *think-aloud* technique and *verbal probing*. The think-aloud technique requires respondents to verbalize every thought that they have while answering each item. Here, the interviewer simply supports this activity by encouraging the respondent to keep talking and to record what is said for later analysis (Willis & Artino 2013). This technique can provide valuable information, but it tends to be unnatural and difficult for most respondents, and it can result in reams of free-response data that the survey designer then needs to cull through.

A complementary procedure, verbal probing, is a more active form of data collection where the interviewer administers a series of probe questions designed to elicit specific information (Willis & Artino 2013; see Table 4 for a list of commonly used verbal probes). Verbal probing is classically divided into concurrent and retrospective probing. In concurrent probing, the interviewer asks the respondent specific questions about their thought processes as the respondent answers each question. Although disruptive, concurrent probing has the advantage of allowing participants to respond to questions while their thoughts are recent. Retrospective probing, on the other hand, occurs after the participant has completed the entire survey (or section of the survey) and is generally less disruptive than concurrent probing. The downside of retrospective probing is the risk of recall bias and hindsight effects (Drennan 2003). A modification to the two verbal probing techniques is defined as immediate retrospective probing, which allows the interviewer to find natural break points in the survey. Immediate retrospective probing allows the interviewer to probe the respondent without interrupting between each item (Watt et al. 2008). This approach has the potential benefit of reducing the recall bias and hindsight

Table 4. Examples of commonly used verbal probes.

Type of verbal probe	Example
Comprehension/interpretation	''What does the term 'continuing medical education' mean to you?''
Paraphrasing	''Can you restate the question in your own words?''
Confidence judgment	''How sure are you that you have participated in 3 formal educational programs?''
Recall	''How do you remember that you have participated in 3 formal educational programs?''
	''How did you come up with your answer?''
Specific	''Why do you say that you think it is very important that physicians participant in continuing medical education?''
General	''How did you arrive at that answer?''
	''Was that easy or hard to answer?''
	''I noticed that you hesitated. Tell me what you were thinking.''
	''Tell me more about that.''

Adapted with permission from the *Journal of Graduate Medical Education*: Willis & Artino 2013. What do our respondents think we're asking? Using cognitive interviewing to improve medical education surveys. J Grad Med Educ 5:353–356.

effects while limiting the interviewer interruptions and decreasing the artificiality of the process. In practice, many cognitive interviews will actually use a mixture of think-aloud and verbal probing techniques to better identify potential errors.

Once a cognitive interview has been completed, there are several methods for analyzing the qualitative data obtained. One way to quantitatively analyze results from a cognitive interview is through coding. With this method, pre-determined codes are established for common respondent errors (e.g. respondent requests clarification), and the frequency of each type of error is tabulated for each item (Napoles-Springer et al. 2006). In addition, codes may be ranked according to the pre-determined severity of the error. Although the quantitative results of this analysis are often easily interpretable, this method may miss errors not readily predicted and may not fully explain why the error is occurring (Napoles-Springer et al. 2006). As such, a qualitative approach to the cognitive interview can also be employed through an interaction analysis. Typically, an interaction analysis attempts to describe and explain the ways in which people interpret and interact during a conversation, and this method can be applied during the administration of a cognitive interview to determine the meaning of responses (Napoles-Springer et al. 2006). Studies have demonstrated that the combination of coding and interaction analysis can be quite effective, providing more information about the ''cognitive validity'' of a new questionnaire (Napoles-Springer et al. 2006).

The importance of respondents understanding each item in a similar fashion is inherently related to the overall reliability of the scores from any new questionnaire. In addition, the necessity for respondents to understand each item in the way it was intended by the survey creator is integrally related to the validity of the survey and the inferences that can be made with the resulting data. Taken together, these two factors are critically important to creating a high-quality questionnaire, and each factor can be addressed through the use of a well-designed cognitive interview. Ultimately, regardless of the methods used to conduct the cognitive interviews and analyze the data, the information gathered should be used to modify and improve the overall questionnaire and individual survey items.

Step 7: Conduct pilot testing

Despite the best efforts of medical education researchers during the aforementioned survey design process, some survey items may still be problematic (Gehlbach & Brinkworth 2011). Thus, the next step is to pilot test the questionnaire and continue collecting validity evidence. Two of the most common approaches are based on *internal structure* and *relationships with other variables* (AERA, APA & NCME 1999). During pilot testing, members of the target population complete the survey in the planned delivery mode (e.g. web-based or paper-based format). The data obtained from the pilot test is then reviewed to evaluate item range and variance, assess score reliability of the whole scale and review item and composite score correlations. During this step, survey designers should also review descriptive statistics (e.g. means and standard deviations) and histograms, which demonstrate the distribution of responses by item. This analysis can aid in identifying items that may not be functioning in the way the designer intended.

To ascertain the internal structure of the questionnaire and to evaluate the extent to which items within a particular scale measure a single underlying construct (i.e. the scale's uni-dimensionality), survey designers should consider using advanced statistical techniques such as factor analysis. Factor analysis is a statistical procedure designed to evaluate ''the number of distinct constructs needed to account for the pattern of correlations among a set of measures'' (Fabrigar & Wegener 2012, p. 3). To assess the dimensionality of a survey scale that has been deliberately constructed to assess a single construct (e.g. using the processes described in this study), we recommend using confirmatory factor analysis techniques; that said, other scholars have argued that exploratory factor analysis is more appropriate when analyzing new scales (McCoach et al. 2013). Regardless of the specific analysis employed, researchers should know that factor analysis techniques are often poorly understood and poorly implemented; fortunately, the literature is replete with many helpful guides (see, for example, Pett et al. 2003; McCoach et al. 2013).

Conducting a reliability analysis is another critical step in the pilot testing phase. The most common means of assessing scale reliability is by calculating a Cronbach's alpha coefficient. Cronbach's alpha is a measure of the internal consistency of the item scores (i.e. the extent to which the scores for the items on a scale correlate with one another). It is a function of the inter-item correlations and the total number of items on a particular scale. It is important to note that Cronbach's alpha is not a good measure of a scale's uni-dimensionality (measuring a single concept) as is often assumed (Schmitt 1996). Thus, in most cases, survey designers should first run a factor analysis,

to assess the scale's uni-dimensionality and then proceed with a reliability analysis, to assess the internal consistency of the item scores on the scale (Schmitt 1996). Because Cronbach's alpha is sensitive to scale length, all other things being equal, a longer scale will generally have a higher Cronbach's alpha. Of course, scale length and the associated increase in internal consistency reliability must be balanced with over-burdening respondents and the concomitant response errors that can occur when questionnaires become too long and respondents become fatigued. Finally, it is critical to recognize that reliability is a necessary but insufficient condition for validity (AERA, APA & NCME 1999). That is, to be considered valid, survey scores must first be reliable. However, scores that are reliable are not necessarily valid for a given purpose.

Once a scale's uni-dimensionality and internal consistency have been assessed, survey designers often create composite scores for each scale. Depending on the research question being addressed, these composite scores can then be used as independent or dependent variables. When attempting to assess hard-to-measure educational constructs such as motivation, confidence and satisfaction, it usually makes sense to create a composite score for each survey scale than it does to use individual survey items as variables (Sullivan & Artino 2013). A composite score is simply a mean score (either weighted or unweighted) of all the items within a particular scale. Using mean scores has several distinct advantages over summing the items within a particular scale or subscale. First, mean scores are usually reported using the same response scale as the individual items; this approach facilitates more direct interpretation of the mean scores in terms of the response anchors. Second, the use of mean scores makes it clear how big (or small) measured differences really are when comparing individuals or groups. As Colliver et al. (2010) warned, "the sums of ratings reflect both the ratings and the number of items, which magnifies differences between scores and makes differences appear more important than they are" (p. 591).

After composite scores have been created for each survey scale, the resulting variables can be examined to determine their relations to other variables that have been collected. The goal in this step is to determine if these associations are consistent with theory and previous research. So, for example, one might expect the composite scores from a scale designed to assess *trainee confidence for suturing* to be positively correlated with the number of successful suture procedures performed (since practice builds confidence) and negatively correlated with procedure-related anxiety (as more confident trainees also tend to be less anxious). In this way, survey designers are assessing the validity of the scales they have created in terms of their relationships to other variables (AERA, APA & NCME 1999). It is worth noting that in the aforementioned example, the survey designer is evaluating the correlations between the newly developed scale scores and both an objective measure (number of procedures) and a subjective measure (scores on an anxiety scale). Both of these are reasonable approaches to assessing a new scale's relationships with other variables.

Concluding thoughts

In this AMEE Guide, we described a systematic, seven-step design process for developing survey scales. It should be noted that many important topics related to survey implementation and administration fall outside our focus on scale design and thus were not discussed in this guide. These topics include, but are not limited to, ethical approval for research questionnaires, administration format (paper vs. electronic), sampling techniques, obtaining high response rates, providing incentives and data management. These topics, and many more, are reviewed in detail elsewhere (e.g. Dillman et al. 2009). We also acknowledge that the survey design methodology presented here is not the only way to design and develop a high-quality questionnaire. In reading this Guide, however, we hope medical education researchers will come to appreciate the importance of following a systematic, evidence-based approach to questionnaire design. Doing so not only improves the questionnaires used in medical education but it also has the potential to positively impact the overall quality of medical education research, a large proportion of which employs questionnaires.

Glossary

Closed-ended question – A survey question with a finite number of response categories from which the respondent can choose.

Cognitive interviewing (or cognitive pre-testing) – An evidence-based qualitative method specifically designed to investigate whether a survey question satisfies its intended purpose.

Concurrent probing – A verbal probing technique wherein the interviewer administers the probe question immediately after the respondent has read aloud and answered each survey item.

Construct – A hypothesized concept or characteristic (something "constructed") that a survey or test is designed to measure. Historically, the term "construct" has been reserved for characteristics that are not directly observable. Recently, however, the term has been more broadly defined.

Content validity – Evidence obtained from an analysis of the relationship between a survey instrument's content and the construct it is intended to measure.

Factor analysis – A set of statistical procedures designed to evaluate the number of distinct constructs needed to account for the pattern of correlations among a set of measures.

Open-ended question – A survey question that asks respondents to provide an answer in an open space (e.g. a number, a list or a longer, in-depth answer).

Reliability – The extent to which the scores produced by a particular measurement procedure or instrument (e.g. a survey) are consistent and reproducible. Reliability is a necessary but insufficient condition for validity.

Response anchors – The named points along a set of answer options (e.g. *not at all important*, *slightly important*, *moderately important*, *quite important* and *extremely important*).
Response process validity – Evidence of validity obtained from an analysis of how respondents interpret the meaning of a survey scale's specific survey items.
Retrospective probing – A verbal probing technique wherein the interviewer administers the probe questions after the respondent has completed the entire survey (or a portion of the survey).
Scale – Two or more items intended to measure a construct.
Think-aloud interviewing – A cognitive interviewing technique wherein survey respondents are asked to actively verbalize their thoughts as they attempt to answer the evaluated survey items.
Validity – The degree to which evidence and theory support the proposed interpretations of an instrument's scores.
Validity argument – The process of accumulating evidence to provide a sound scientific basis for the proposed uses of an instrument's scores.
Verbal probing – A cognitive interviewing technique wherein the interviewer administers a series of probe questions specifically designed to elicit detailed information beyond that normally provided by respondents.

Notes on contributors

ANTHONY R. ARTINO, Jr., PhD, is an Associate Professor of Preventive Medicine and Biometrics. He is the Principal Investigator on several funded research projects and co-directs the Long-Term Career Outcome Study (LTCOS) of Uniformed Services University (USU) trainees. His research focuses on understanding the role of academic motivation, emotion and self-regulation in a variety of settings. He earned his PhD in educational psychology from the University of Connecticut.

JEFFREY S. LA ROCHELLE, MD, MPH, is an Associate Program Director for the Internal Medicine residency at Walter Reed National Military Medical Center and is the Director of Integrated Clinical Skills at USU where he is an Associate Professor of Medicine. His research focuses on the application of theory-based educational methods and assessments and the development of observed structured clinical examinations (OSCE). He earned his MD and MPH from USU.

KENT J. DEZEE, MD, MPH, is the General Medicine Fellowship Director and an Associate Professor of Medicine at USU. His research focuses on understanding the predictors of medical student success in medical school, residency training and beyond. He earned his MD from The Ohio State University and his MPH from USU.

HUNTER GEHLBACH, PhD, is an Associate Professor at Harvard's Graduate School of Education. He teaches a course on the construction of survey scales, and his research includes experimental work on how to design better scales as well as scale development projects to develop better measures of parents' and students' perceptions of schools. In addition, he has a substantive interest in bringing social psychological principles to bear on educational problems. He earned his PhD from Stanford's Psychological Studies in Education program.

Declaration of interest: Several of the authors are military service members. Title 17 U.S.C. 105 provides that "Copyright protection under this title is not available for any work of the United States Government". Title 17 U.S.C. 101 defines a United States Government work as a work prepared by a military service member or employee of the United States Government as part of that person's official duties.

The views expressed in this article are those of the authors and do not necessarily reflect the official views of the Uniformed Services University of the Health Sciences, the U.S. Navy, the U.S. Army, the U.S. Air Force, or the Department of Defense.

Portions of this AMEE Guide were previously published in the *Journal of Graduate Medical Education* and *Academic Medicine* and are used with the express permission of the publishers (Gehlbach et al. 2010; Artino et al. 2011; Artino & Gehlbach 2012; Rickards et al. 2012; Magee et al. 2013; Willis & Artino 2013).

References

American Educational Research Association (AERA), American Psychological Association (APA) & National Council on Measurement in Education (NCME). 1999. Standards for education and psychological testing. Washington, DC: American Educational Research Association.

Artino AR, Gehlbach H, Durning SJ. 2011. AM last page: Avoiding five common pitfalls of survey design. Acad Med 86:1327.

Artino AR, Gehlbach H. 2012. AM last page: Avoiding four visual-design pitfalls in survey development. Acad Med 87:1452.

Beck CT, Gable RK. 2001. Ensuring content validity: An illustration of the process. J Nurs Meas 9:201–215.

Christian LM, Parsons NL, Dillman DA. 2009. Designing scalar questions for web surveys. Sociol Method Res 37:393–425.

Colliver JA, Conlee MJ, Verhulst SJ, Dorsey JK. 2010. Reports of the decline of empathy during medical education are greatly exaggerated: A reexamination of the research. Acad Med 85:588–593.

Cook DA, Beckman TJ. 2006. Current concepts in validity and reliability for psychometric instruments: Theory and application. Am J Med 119: 166.e7–166.e16.

DeVellis RF. 2003. Scale development: Theory and applications. 2nd ed. Newbury Park, CA: Sage.

Dillman D, Smyth J, Christian L. 2009. Internet, mail, and mixed-mode surveys: The tailored design method. 3rd ed. Hoboken, NJ: Wiley.

Drennan J. 2003. Cognitive interviewing: Verbal data in the design and pretesting of questionnaires. J Adv Nurs 42(1):57–63.

Fabrigar LR, Wegener DT. 2012. Exploratory factor analysis. New York: Oxford University Press.

Fowler FJ. 2009. Survey research methods. 4th ed. Thousand Oaks, CA: Sage.

Gehlbach H, Artino AR, Durning S. 2010. AM last page: Survey development guidance for medical education researchers. Acad Med 85:925.

Gehlbach H, Brinkworth ME. 2011. Measure twice, cut down error: A process for enhancing the validity of survey scales. Rev Gen Psychol 15:380–387.

Kane MT. 2006. Validation in educational measurement. 4th ed. Westport, CT: American Council on Education/Praeger.

Karabenick SA, Woolley ME, Friedel JM, Ammon BV, Blazevski J, Bonney CR, De Groot E, Gilbert MC, Musu L, Kempler TM, Kelly KL. 2007. Cognitive processing of self-report items in educational research: Do they think what we mean? Educ Psychol 42(3):139–151.

Krosnick JA. 1999. Survey research. Annu Rev Psychol 50:537–567.

Magee C, Byars L, Rickards G, Artino AR. 2013. Tracing the steps of survey design: A graduate medical education research example. J Grad Med Educ 5(1):1–5.

McCoach DB, Gable RK, Madura JP. 2013. Instrument development in the affective domain: School and corporate applications. 3rd ed. New York: Springer.

McIver JP, Carmines EG. 1981. Unidimensional scaling. Beverly Hills, CA: Sage.

McKenzie JF, Wood ML, Kotecki JE, Clark JK, Brey RA. 1999. Establishing content validity: Using qualitative and quantitative steps. Am J Health Behav 23(4):311–318.

Napoles-Springer AM, Olsson-Santoyo J, O'Brien H, Stewart AL. 2006. Using cognitive interviews to develop surveys in diverse populations. Med Care 44(11):s21–s30.

Pett MA, Lackey NR, Sullivan JJ. 2003. Making sense of factor analysis: The use of factor analysis for instrument development in health care research. Thousand Oaks, CA: Sage Publications.

Polit DF, Beck CT. 2004. Nursing research: Principles and methods. 7th ed. Philadelphia: Lippincott, Williams, & Wilkins.

Polit DF, Beck CT. 2006. The content validity index: Are you sure you know what's being reported? Critique and recommendations. Res Nurs Health 29:489–497.

Rickards G, Magee C, Artino AR. 2012. You can't fix by analysis what you've spoiled by design: developing survey instruments and collecting validity evidence. J Grad Med Educ 4(4):407–410.

Rubio DM, Berg-Weger M, Tebb SS, Lee ES, Rauch S. 2003. Objectifying content validity: Conducting a content validity study in social work research. Soc Work Res 27(2):94–104.

Schmitt N. 1996. Uses and abuses of coefficient alpha. Psychol Assess 8: 350–353.

Schwarz N. 1999. Self-reports: How the questions shape the answers. Am Psychol 54:93–105.

Sullivan G. 2011. A primer on the validity of assessment instruments. J Grad Med Educ 3(2):119–120.

Sullivan GM, Artino AR. 2013. Analyzing and interpreting data from Likert-type scales. J Grad Med Educ 5(4):541–542.

Tourangeau R, Rips LJ, Rasinski KA. 2000. The psychology of survey response. New York: Cambridge University Press.

Waltz CF, Strickland OL, Lenz ER. 2005. Measurement in nursing and health research. 3rd ed. New York: Springer Publishing Co.

Watt T, Rasmussen AK, Groenvold M, Bjorner JB, Watt SH, Bonnema SJ, Hegedus L, Feldt-Rasmussen U. 2008. Improving a newly developed patient-reported outcome for thyroid patients, using cognitive interviewing. Quality of Life Research 17:1009–1017.

Weng LJ. 2004. Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. Educ Psychol Meas 64:956–972.

Willis GB, Artino AR. 2013. What do our respondents think we're asking? Using cognitive interviewing to improve medical education surveys. J Grad Med Educ 5(3):353–356.

Willis GB. 2005. Cognitive interviewing: A tool for improving questionnaire design. Thousand Oaks, CA: Sage Publications.