2014, 36: 657–674 MEDICAL TEACHER

AMEE GUIDE

Medical education scholarship: An introductory guide: AMEE Guide No. 89

GERALD E. CRITES[1], JULIE K. GAINES[1], SCOTT COTTRELL[2], SUMMERS KALISHMAN[3], MARYELLEN GUSIC[4], BRIAN MAVIS[5] & STEVEN J. DURNING[6]

[1]GRU/UGA Medical Partnership, USA, [2]WVU, USA, [3]University of New Mexico, USA, [4]Indiana University School of Medicine, USA, [5]Michigan State University, USA, [6]Uniformed Services University of the Health Sciences, USA

Abstract

This AMEE Guide provides an overview of medical education scholarship for early career scholars, based upon a summary of the existing literature and pragmatic advice derived from the experience of its authors. After providing an introduction to the principles of scholarship and describing questions that the Guide addresses, the authors offer a conceptual description of the complementary traditions of teaching and educational discovery, and advocate for the development of educational scholars with both traditions. They then describe the attributes of effective mentor–mentee relationships and how early career scholars can identify potential mentors who can fulfill this role. In the subsequent sections, they describe the appropriate development of scholarly questions and other components of a complete scholarly plan, including how to use conceptual frameworks in guiding such plans. From here, they describe methods that align with both the teaching and discovery traditions and provide concrete examples of each. They then provide guidelines for assessing the impact of scholarship, identify the various opportunities for sharing it, and how to effectively interpret and describe it. Additionally, they provide practical advice on how appropriately to demonstrate the scholarship in a promotional packet, including the principle of reflectivity in scholarship. Finally, they address the principles of applied research ethics for educational scholarship and when to consider soliciting approval for scholarly activities by a human research board.

Introduction

In the AAMC's classic 2000 special issue on *Expanding the View of Scholarship*, a group of international scholars from the Council of Academic Societies (CAS) published a series of articles advocating a broader view of scholarship as it relates to Boyer's original classification: The Scholarship of Discovery (research, including educational research), the Scholarship of Application, the Scholarship of Integration and the Scholarship of Teaching (Boyer 1990; Beattie 2000). In the series, the authors maintained that all high-quality scholarship must address six core principles: clear goals, adequate preparation, appropriate methods, significant results, effective presentation and reflective critique (Glassick 2000). These scholars advocated a valuation and promotion system within medical schools that equally weigh the four types of scholarship (Bordage et al. 2001).

Despite the growing acceptance of these four types of scholarship in academic institutions, a few lingering questions remain about the differences between educational discovery and other forms of scholarship, particularly teaching scholarship. Furthermore, relatively few resources address the practical issues related to implementing scholarly activities and how to grow an academic career from such activities.

Practice points

- All scholarship should be guided and judged by Glassick's six core principles of excellence for scholarship.
- The educational discovery (research) and teaching scholarly traditions are based upon different assumptions and utilize different methods, but they address similar educational questions and goals and are equally important for the development of educational scholars.
- Educational scholars should carefully articulate their goals by formulating thoughtful questions and select appropriate methodologies to address these questions.
- Successful scholars purposefully design and implement their scholarly activities and early career scholars should seek guidance of mentors for these activities.
- Effective scholars accurately interpret their scholarship's scope and impact and understand how to appropriately present its results.
- Some educational scholarly activities fall under research subject protections and therefore may require review by a human research board (IRB).

Correspondence: Gerald E. Crites, MD, MEd, Associate Professor of Medicine, GRU-UGA Medical Partnership, 108 Spear Road, Winnie Davis Hall-209, Athens, GA 30602, USA. Tel: +706 713 2192; Fax: +706 713 2222; E-mail: gcrites@uga.edu

ISSN 0142-159X print/ISSN 1466-187X online/14/80657–18 © 2014 Informa UK Ltd.
DOI: 10.3109/0142159X.2014.916791

In the course of this Guide, the authors attempt to provide answers to these and other questions as a compendium for medical educational scholarship (Table 1).

If it is not yet evident, this Guide is primarily written for early career scholars who are attempting to identify a direction for her/his educational scholarship; late career scholars may find some sections too rudimentary to be useful. Also, given the scope of this Guide, the authors attempt to address these questions in a succinct way; however, the reader may find some explanations lacking in certain details, so it is advisable to explore the references to address any unresolved questions. Finally, the reader should also consider how the activities of educational scholarship fit into a cohesive plan for one's professional development (McLean et al. 2008; McGaghie 2009).

The scholarship of educational discovery and the scholarship of teaching

First, clarifying terminology is in order. Using a scholarly approach, as implied by Glassick's (2000), criteria lays the groundwork for scholarship. Scholarly activities are the inquiries guided by an academic tradition and the dissemination of the results of inquiries to allow peer judgment of their merit, erudition and utility; the cumulative description for all these activities is scholarship. When relating to educational activities, the authors use the super-ordinate category of educational scholarship to include both educational discovery (research) and teaching scholarship. The authors also emphasize that their definition of educational scholarship evolves from activities within medical educational systems that impact teachers and learners in these systems; other scholarly traditions that may have educational practices for learners not meeting this definition (e.g. health promotion research and implementation science) will not be covered here.

Table 1. Questions related to medical educational scholarship.

- How do educational discovery and teaching scholarship relate, and how are they different?
- How does one select a mentor and when is specific help needed for a project?
- How does one craft clear scholarly questions that have a good chance of leading to something that advances the field of teaching and learning?
- What are the key components to a scholarship plan? How can models and frameworks serve as conceptual roadmaps for scholarship?
- Which methods should one choose to address scholarly questions? How does one take a scholarly approach to teaching? How does one take a scholarly approach to educational discovery?
- How and where can one share scholarship?
- How does one get academic credit for scholarship? How does one show reflectivity in promotion materials?
- When does educational scholarship have ethical concerns (i.e. require IRB approval)?

Beyond the definition provided, is it really necessary to discriminate between scholarship that results from discovery or teaching? They both must have, when done properly, a scholarly basis (e.g. being theory-based and guided by the existing literature), can study the same set of educational processes and can be equally valued by institutions. There are, however, a few arguments for considering their distinctions. For example, some educational grants require a research-based approach (National Board of Medical Examiners 2013) while others allow a more flexible approach (Health Resources and Services Administration 2013), some conferences split scholarly works and their presentation fora across these lines (AAMC-Group on Educational Affairs 2013) and many institutions have developed separate promotions criteria to represent the distinctions (Albert Einstein College of Medicine of Yeshiva University 2013; University of Ottawa 2013). Ultimately, and as the central theme of this Guide, these distinctions may be helpful when creating an appropriate scholarly plan (to use the "correct" approach when developing an inquiry).

Before engaging in an inquiry, scholars first need to understand the philosophical tradition of a discipline that guides the inquiry. Table 2 provides a few comparisons and contrasts between the educational discovery and teaching traditions as a way to understand where their scholarly approaches differ and yet complement each other. Both the discovery and teaching traditions serve the same overarching purpose; that is, to build understanding of learning experiences. To do this, the two traditions take somewhat different approaches. The discovery tradition (researcher) takes its direction from the scientific method, where a phenomenon under study is presumed to be universal, assumes the phenomenon under study can be objectively investigated,

Table 2. Comparisons and contrasts between the educational researcher and teaching traditions.

Characteristic	Educational researcher tradition	Teaching tradition
Overarching purpose	To understand learning experiences	To understand learning experiences
Methodological foundation for inquiry	Scientific method	Experiential learning
Investigator's role	Objective, detached	Integrated with process (the teacher)
How the learning setting and context is perceived	Setting/context can be replicated	Setting/context is somewhat unique
How participants and educational processes are viewed	Study subjects and processes as sources of differences and variation	Teachers/learners with variable needs/abilities; processes reflect the interaction between teachers and learners
Types of conclusions drawn	Inferences	Evaluations
How conclusions are used	Conclusions can be generalized	Conclusions need to be taken in consideration of their context
Implications of inquiry	Improvement in teaching practices, educational administration and learning	Improvement in teaching practices, educational administration and learning

assumes the setting can be replicated elsewhere, assumes that all relevant study variables (including participants and processes) can be identified and captured, and creates conclusions as inferences derived from datasets that are generalizable. The teaching tradition uses an iterative learning process based upon first hand experiences, assumes that the investigators (teachers) are part of the process under study and their subjectivity is usually useful, assumes learning contexts and settings are somewhat unique, views human participants with an eclectic mix of abilities/needs and the educational processes reflect this diversity, draws conclusions as evaluations using personal experience and data analysis, and requires that any conclusions about generalizability must be considered in the context of the inquiry. Finally, both the discovery and teaching traditions inform the same practical matters, that is, improving teaching practices, educational administration and learning.

In this Guide, the authors do not advocate for one tradition or another; they are both effective approaches to make sense of events that occur in complex learning systems and they are equally valid approaches to consider for any scholarly question. They might be best conceptualized of as different ends of a continuum and, therefore, should not be considered mutually exclusive but complementary. Indeed, in order to become a complete educational scholar, the authors recommend readers consider engaging both approaches for any scholarly project. This is a recurring theme of this Guide.

Getting help

Prior to beginning a scholarly project, early career scholars should consider recruiting a mentor to help guide their development. Mentoring (or coaching) is a topic of growing importance in medical education, but little empirical research to guide the development of mentors and mentees exists; for a comprehensive understanding, the reader is encouraged to review some of better works on this topic (Challis 2000; Ross & Cameron 2007; Hammick et al. 2009; McGaghie 2009). Some resources attempt to provide distinctions between coaching and mentoring to help guide early career scholars in making wise choices; our main distinction between a coach and a mentor refers to the length of time of the relationship (mentoring is longer/longitudinal) and the type of guidance (mentoring provides both general career as well as project specific guidance). If inclined to learn about further distinctions, readers are encouraged to explore the provided references (Garvey et al. 2009; Hicks & McCracken 2009; Macaffee & Garvey 2010). Going forward, the focus in this section will be on getting help with scholarship with an expanded focus on mentoring.

When using the guidance of a mentor, mentees typically experience a positive impact on their personal development, career choice and scholarly productivity (Sambunjak et al. 2006). Other authors propose that mentorship leads to greater mentee career satisfaction, networking within a profession and aiding in stress management (Detsky & Baerlocher 2007). Competing time demands for mentors and inadequate support for the development of the mentor–mentee relationship by some academic institutions are two challenges that will likely be ongoing obstacles for scholarly development (Sambunjak et al. 2006).

The authors offers some suggestions regarding "best practices" below. The reader (from the perspective of a mentee) should be mindful that, due the lack of studies, empirical evidence for these recommendations is not available but are based mostly upon the authors' diverse experience and a synthesis of the qualitative literature:

(1) *Listen*. This is perhaps the most important feature of being a mentor and, for mentees, is an essential quality to seek in a mentor. Mentors who use effective listening and reflection can help mentees determine what kind of career they wish to pursue and identify the best path to this career. For example, a mentee should assure that a mentor is able to inquire about how a mentee likes to spend his/her time prior to structuring specific roles. Also, mentors should be mindful that their mentees may choose not to follow their advice, and mentees may need to occasionally use this privilege to enhance personal growth.

(2) *Spend time on this relationship*. This is becoming more challenging in our current academic climate. Mentees should recognize potential mentors who are willing to commit their time and follow through by setting aside time to meet; mentors who delay or cancel meetings can interfere with relationship building and the development of mutual trust.

(3) *Understand the mentoring role*. There is a key difference between coaching and mentoring. A mentee may have several coaches who help with specific aspects of their scholarly pursuits, but a mentor provides longitudinal support across several pursuits. It is important for mentees to regularly schedule time with their mentors even if there may be no obvious agenda; it is also advisable to return frequently to a discussion of mentees overarching career goals, such as what is the mentee doing, planning and visioning, to inform both intermediate and longer term career goals. The agenda of the meetings should be left to the mentee as much as possible and mentors should provide support by answering questions, reviewing manuscripts and providing general advice.

(4) *Set boundaries*. Being "friends" can compromise the mentor–mentee relationship but this is not to say that the relationship should not be enjoyable. Setting boundaries also allows for needed feedback which may be critical at times and provides both mentors and mentees the freedom to be transparent and honest at all times. Boundaries should also include being explicit about credit for scholarship; this can be challenging for both the mentor as well as the mentee. When working within a mentors program of study the roles are typically understood by both parties, but this relationship can become strained as the mentee's expertise grows and wishes to embark on a program of study of his/her own. Therefore, it is important to be explicit about the roles and responsibilities, including who is going to receive what credit on a scholarly project and when mentees should pursue spin-off projects. When ready, a mentee should be encouraged to pursue her/his own scholarly plan and

both the mentor and the mentee should be aware that the time will come when the mentor, to further a mentee's development, will need to stop sharing scholarly contributions. An alternative (but often less available) option is identifying mentors who do not share in the mentees scholarly activity. Finally, not every mentor–mentee relationship will work perfectly. If and when this occurs, the mentee may need to initiate a discussion about the need for changing roles or possibly changing mentors.

(5) *Provide focus*. A common problem for junior scholars is difficulty in determining one's "program of research" or "scholarly direction" and, therefore, risk getting overextended in multiple scholarship related activities. Mentees should frequently engage mentors about scholarly choices and a good mentor will help a mentee say "no" to projects that are not feasible or are a poor fit for the mentees long-term goals.

(6) *Establish support networks*. Mentors know the experts and the "up and coming" scholars in a field; therefore, mentees should engage their mentors to help them connect with scholars who have similar pursuits. This collaboration enables more work to be done (increasing the capacity for research for the mentee and sharing the workload burden with collaborators) as long as clear guidelines are provided to each member on the project team. A common mistake early career scholars make is not being inclusive enough early on (e.g. identifying potential collaborators the opportunity to meaningfully participate in a project or paper), and a mentor can help with these critical decisions. Establishing support networks also allows the mentor to "fade the scaffolding" of support with the mentee, facilitating their career growth. Mentees are encouraged to request help with networking and collaboration from mentors and good mentors should be familiar with some of the better mentoring practices around collaborative networks (Garvey et al. 2009; McGaghie 2009).

(7) *Be mindful of promotion and tenure requirements*. It is a necessary for mentees to occasionally solicit a mentor's advice for a mentee's long-term success and the advice must take into account short-term milestones. Knowing promotion and tenure guidelines and helping the mentee to navigate this journey are essential for effective mentoring. For example, if one is interested in soliciting letters of support from external referees for a promotion packet, then this may preclude these individuals from collaborating on projects or papers, and vice versa.

(8) *Mentoring as a "journey"*. Viewed from the lens of a "coach", mentoring does not take on specific, predefined roles or responsibilities but provides the needed skill and trait development (e.g. deliberate practice activities) as mentees grow and develop over years in their discipline. Such a view takes on an emphasis whereby scholarly products/outcomes are not as critical as much as the processes of building a trustful relationship, sharing expertise, providing moral support and knowing when to provide a mentee room to branch out on their own (Awaya et al. 2003). We believe that these activities are critical to attend.

Table 3. Useful steps in question development.

- Free write a list of ideas of what is interesting about this topic or issue in education, and then put it down for a while.
- Revisit it and revise.
- Present the list of ideas to a colleague, preferably one who is unfamiliar with the intended scholarly project (if one can explain it to him/her, then it is generally easier to explain it to a broader audience).
- With this feedback, begin to refine ideas into questions. It may be appropriate to judge whether some scholarly questions have similarities and need to be grouped together, and whether some questions are subordinate to broader questions.
- Consider getting some advice from an expert or mentor (especially if the reader is unfamiliar or inexperienced with the conceptual framework or approach, for example the data source or analytic process under consideration).

In order to identify potential mentor candidates, readers should contact senior scholars in their institution or academic supervisors (e.g. department chairs). Additionally there are several academic societies that emphasize mentoring services at a distance and the reader is encouraged to identify opportunities through them if they are available.

Setting clear goals

Once a scholar identifies the individuals needed to support an inquiry, another step for a successful scholarly approach is setting clear goals. Scholars often fail to fully reflect upon the goals of their inquiry; they can address this by being thoughtful about their scholarly questions and the best approaches to answer those questions (Ringsted et al. 2011). Most scholars come from narrow research or teaching traditions and have been exposed to only a few approaches. As such, scholars try to fit their question to a particular approach rather than the other way around. For this reason, the steps found in Table 3 may be useful for question development:

At this stage, a comparison of each question (or groups of questions) to the columns in Table 2 may be helpful. How much effort would it take to allow multiple lines of inquiry (discovery, teaching or both) to emerge from this work? Would multiple lines of inquiry be useful for building a clearer picture for the phenomenon under study? Clearly articulating the different lines of inquiry may also help keep these ideas separate during planning and hopefully result in multiple scholarly products, allowing one question/series of questions to be "counted twice" (Bordage 2010).

Once a list (and groups) of scholarly questions are developed, the reader may want to apply one of the several existing criteria to judge the "survivability" of the scholarly inquiry. One useful set of criteria is the FINER criteria (Seehusen & Weaver 2009) which were developed to judge the likelihood of success of a scholarly activity (Table 4).

The structure of a good question depends on the specific type of scholarship and, therefore, is difficult to address specifically. A general guideline to writing questions is provided below:

- What are the underlying assumptions of the inquiry (not stated within the question proper, but clarified along with it)?

Table 4. The FINER criteria.

- Is this inquiry Feasible? Does the author have institutional support and/or the necessary resources to pull it off, such as protected time or mentoring? Will the plan need financial support and does the author have access to such resources? Are the questions of ''appropriate size'' so one can realistically do them in the intended time frame and required setting(s)? Will changes to program policies or processes be needed for the inquiry to succeed? And if so, are these changes possible/supported by the institution?
- Is it Interesting to others? Although a scholarly question sounds interesting to its originator, it must also be interesting to the broader educational community.
- Is it Novel? Has it been done before (or has it been done in the context that the author is interested in studying)?
- Is it Ethical? Are there any ethical issues that need addressed and is institutional review needed?
- Is it Relevant? Will users of the resultant scholarship find it useful? Will the scholarship provide a foundation upon which others can build? (Gusic et al. 2013)

- How does it add to what is known collectively and/or through the literature? Does it have a theoretical or conceptual framework to guide its development?
- Who or what is the target of the inquiry?
- Which set of circumstances or settings need to be in place to assure that the questions are adequately addressed?
- Is there a comparison group, metric or evaluation standard that needs to be considered when interpreting the results?
- What are the outcomes, and why are they important?

Adequate preparation

Most intended scholars make it this far, but, beyond it, many fail to get their scholarship off the ground. Many times, the main culprit may be the lack of scholarship planning, sometimes referred to as a research plan (if the project is research) (Leedy & Ormrod 2001). The plan itself, the explicit statement of the intentions of a scholarly project, is useful at the beginning of a scholarly project but it will likely change after its initiation. The scholarly planning process should also consist of periodic changes to the plan as each situation dictates. The process of regularly reviewing the plan typically improves a scholar's organizational and time-management skills, but, more importantly, it provides an opportunity to renew his/her commitment to the project and place it on equal parity as other life activities. Internal motivation nourishes this commitment; the odds that the reader will follow through with a scholarly project commitment will improve dramatically if the project really excites her/him. A particularly interesting and exciting project sustains a scholar, over the long haul, through ups and downs.

Any scholarly plan should be devised with advice from a coach or mentor for early career scholars. The explicit plan that comes out of the discussion will have several continuously evolving stories, and these can be documented in several important sections:

- *The scholarly questions.* Establishing the scholarly question(s) is an early and critical step. It is the focal point for the study and will help determine many other decisions (e.g. data collection and analysis) within the scholarly plan.
- *The literature and/or current evidence.* It is crucial that the literature is reviewed before the scholarly plan is created. The literature review allows the reader to see what scholarship has already been done, and to study the results of scholarship close to the reader's topic. The review of the literature provides a framework for establishing the importance of the inquiry and allows the scholars to compare his/her findings to other studies (Creswell 2009). Searching for the literature can be done using many reliable resources at the library from online databases (PubMed, CINAHL, ERIC and PsychInfo) to more traditional resources such as books and journals. Librarians are valuable resources when a scholar needs assistance with a literature review. It is important that the literature review should be approached systematically, and it will possibly be visited more than once during the project cycle.
- *The data sources and collection methods (test, methods how to get to data, permissions).* The sources of data and collection methods will also be driven by the scholarly questions (Creswell 2009). As described in the methods sections below, a variety of data collection methods can be used. The sources of the data will also vary. In some instances, data banks of information may be accessed (e.g. test performance over time). In others, humans may be involved as a direct source of data (e.g. survey responses or direct observations). An important component of a scholarly plan is to project when and how the data will be collected and what preparatory activities will be needed to accomplish it. Once data is collected, the reader should understand her/his institutional policies regarding data storage and security and follow them.
- *The analyses/evaluation to be used.* As with data collection, the analysis or evaluation of the data will be driven by the question. The types of data collected are also a determinant of the kinds of analysis or evaluation that might be done on the data. Later sections in this guide provide specific examples of analysis or evaluation techniques. A scholarly plan should include all the resources needed to do the analyses and whether additional resources, such as analytic software or external consultants, is needed.
- *The criteria for determining measurable outcomes.* For some studies, it is important to set criteria to help with the interpretation of the results. The following are examples of outcome criteria for different types of inquiries: the specific learners' score that is an indicator of adequate mastery knowledge attainment; the evaluation metric which is considered as a measure of course success; the number which indicates an adequate number for a subject sample, and; when data saturation is met.
- *The costs of the inquiry and funding.* Many projects will need no budget, but sometimes funding is needed to provide adequate resources for activities such as protected time, participatory incentives, teaching resources, analytic support or consultants. In these instances,

preparing a budget helps organize and solicit support. When projects do require support, locating and applying for internal and/or external funding or grants may be useful for project completion.

- *The project milestones and timelines*. Scholars should establish milestones and timelines as guides for project implementation; specific tools, such as project management software or simple paper charts, can be valuable resources for establishing and managing project timelines. As you are creating your plan, be sure to consider other things outside your study (e.g. personal demands, travel and other work obligations) that may impact it (Leedy & Ormrod 2001). Be aware that timelines often need changed when dictated by unexpected events.

It is useful to consider using educational theories or conceptual frameworks as guides for the scholarly plan. Identifying potential theories and frameworks early on in the planning process can make the scholarly planning easier by allowing a better understanding of how the literature integrates with your plan and how to modify steps in a scholarly plan.

Conceptual frameworks can be theories that have a body of evidence that support them, models that are built from evidence accrued from studies that integrate several theories, or best practices with some outcome evidence from empirical studies. Scholars can use theories and conceptual frameworks early in the scholarly planning process to guide the development of the research questions and any interventions or curricular activities that support them. Conceptual frameworks have often been tested and the evidence associated with them suggests educational interventions or curricular activities that could be investigated further, adding additional value to planning activities. There are a number of AMEE Guides that address conceptual frameworks and may be useful to review (Sandars & Cleary 2011; Schuwirth & van der Vleuten 2011; Yardley et al. 2012). In addition, there are papers well worth reviewing that provide examples from a wide range of conceptual frameworks and illuminate the way these frameworks strengthen the research question and help study design and methodology (Reeves et al. 2008; Bordage 2009).

There are too many available frameworks to allow full description here; the following are four examples of the application of conceptual frameworks in educational scholarship where the authors used an intentional approach to align the scholarly questions and activities around a framework or theory:

- *Concept maps*. West et al. (2000) applied instruction and application of concept mapping to assess how learners in graduate medical education organize their knowledge at different points throughout their training. The framework of concept mapping identified with this approach was developed by Novak and others as a formative tool; through concept mapping learners show the way they relate ideas among different concepts within one topic or subject area (Novak & Gowin 1984; Novak & Musonda 1991). These authors adapted the concept map to their assessment strategies to address several study questions: (1) Could this group measure difference and change in the conceptual frameworks of learners over time in graduate medical education? (2) Could this measurement predict expected change? (3) Could concept maps be scored in reliable ways? The conceptual framework developed by Novak was an organizer for this study and was adapted for the study intervention.
- *Team-based learning*. Bou Akl et al. (2012) used evidence-based recommendations derived from team-based learning (TBL) assessments and the TBL framework developed by Michaelsen & Sweet (2008a,b). The investigators piloted a clinical pharmacology course using the TBL framework with groups of third and fourth year students using 10 multiple choice clinical pharmacology questions in an Internal Medicine exam to answer the following question: Where should the course be located, in the third year or the fourth year of the curriculum? Comparison of these measures for students in the third year and the fourth year cohorts provided findings that led the authors' to recommend placement of the clinical pharmacology TBL course in the third year of medical student training.
- *Classical test theory*. Crites et al. (2012) used classical test theory to design and develop multiple choice item tests of evidence-based medicine (EBM) and clinical decision-making (CDM). The authors over a period of years collaboratively developed a database with multiple choice questions adapted from the taxonomy models of Bloom and Gagne and guided by principles of instructional design (Smith & Ragan 1999). They applied the taxonomy to the questions to sort and appraise each multiple choice item and determine its usefulness in their assessments. The authors used reliability, validity and item analyses to assess the success of this approach. They were successful in developing an EBM/CDM multiple choice question database to assess different knowledge domains and different levels of learners.
- *Interprofessional competency*. Lingard et al. (2012) assessed the dynamics associated with leadership on interprofessional teams based on the framework from the Canadian Interprofessional Health Collaborative. The interprofessional framework identifies shared leadership in response to situational needs rather than based on a hierarchical system of responsibility. Lingard et al. (2012) used case study methodology (Stake 1995) combining interviews and observations of five interprofessional health care teams to assess several questions: (1) Do interprofessional health teams agree about the importance of collaborative leadership? What evidence is there that indicates their views? (2) Using case study observation of teams at work, what leadership dynamics are identified? Are the dynamics consistent with the interprofessional leadership competency? (3) What discrepancies exist between the way collaborative leadership is described and the findings from case studies conducted by the authors? How do the authors characterize them? The study identified tensions in the framework of interprofessional competency and shared leadership and identified engrained structures. The authors found that

identifying the tensions allowed discussion and reflection about its meaning to occur and may enable the teams to address some issues and establish more collaborative leadership.

If the reader is interested in learning more about conceptual frameworks, the papers by Bordage (2009) and Reeves et al. (2008) are very informative. Bordage (2009) provides a rationale for conceptual frameworks and theories and includes multiple examples. Reeves et al. (2008) focus on theories that are commonly used in qualitative research and provide useful tables, examples and references. In addition, the three AMEE guides that we identify have detailed information about the specific conceptual frameworks that are the focus of the guide – assessment theories (Schuwirth & van der Vleuten 2011), experiential learning theories (Yardley et al. 2012) and self-regulation theory (Sandars & Cleary 2011).

Appropriate methods for teaching scholarship

A scholar reaching this point has likely identified mentors/help, goals and a scholarly plan, but still has to choose an appropriate methodology for the inquiry. As stated earlier, a scholar preparing inquiry into an educational phenomenon may need to utilize both traditions of discovery scholarship and teaching scholarship (Palladino et al. 2013). The work of teachers can be classified within five domains: teaching, learner assessment, curriculum development, advising and mentoring and educational leadership/administration (Simpson et al. 2007a, b)

A scholarly approach in any teaching domain can lead to scholarship (Shulman 1993; Hutchings & Shulman 1999). In this section, while keeping in mind the typical activities of the reader, the authors will focus on defining how a scholarly approach for teaching activities and learner assessment can lead to the creation of "products" that can be used by others to judge excellence in teaching and use its recommendation alterations in practices to promote learning.

A scholarly approach to teaching (Fincher et al. 2000; Glassick 2000) requires a teacher to: define clear goals for teaching interactions; prepare using best practices from the literature; select suitable instructional methods and materials using sound educational theories; measure the outcomes and impact of teaching and demonstrate continual improvement of one's work as a teacher. The primary goal of teaching is to assure that learning has occurred by measuring changes in learners' knowledge, skills, attitudes and behaviors and to assure that this happens in a fair and effective manner; in order to accomplish this, teachers need to engage an effective scholarly approach to teaching. Teachers must create assessments that are designed to measure well-defined learning outcomes, are grounded in established educational theory/evidence from the literature and matched to the stated learning objectives for the teaching activity (Baldwin et al. 2011). A scholarly study of assessment tools and practices can inform teachers and educational administrators about the quality of the teaching interaction and also provides data for the learner for his/her own self-assessment (Gusic et al. 2013). By thoughtfully reflecting upon the success and failures of their assessment activities, teachers can use the same processes for self-improvement as part of a scholarly plan (Baldwin et al. 2011; Gusic et al. 2013). Importantly, an educator must document evidence that each of these steps has been taken for it to become scholarship (Simpson et al. 2007a; Palladino et al. 2013). Curriculum vitae do not usually provide space for this type of information and so, an educator should collect and document their activities within a portfolio (Gusic et al. 2007; Simpson et al. 2007c). A developmental portfolio (i.e. teaching portfolio) provides a venue within which an educator can reflect on accomplishments and develop goals for continued professional development as a teacher (Baldwin et al. 2008). A mentor can assist in this critical activity, helping the mentee to enhance their work and also advance in their career.

Table 5 provides an example of how an educator can demonstrate a scholarly approach to teaching and learner assessment activities.

Although teaching scholarship can take various forms, the results of teaching scholarship must be shared and judged as valuable by peers, thereby creating a platform upon which others can continue to build (Shulman 1993; Hutchings & Shulman 1999). In line with Glassick's (2000) criteria for effective presentation, an educator can make her/his scholarship available for peer review and adoption by others and share the strategies and tools they use in teaching and assessment of learners (Chandran et al. 2009a, b; McGaghie 2009). Assessment strategies, tools and the results of assessment activities should be shared with stakeholder groups (teachers, learners, educational administrators) and also disseminated to the larger educational community to advance what is known about assessment. An educator can deliver workshops or didactic sessions about teaching or assessment methods at meetings through a peer-reviewed process. Instructional materials and assessment tools can be published in peer-reviewed repositories or shared warehouses. Educators may contribute to books, edited by peers, about teaching or assessment strategies; there are also examples of the work of educators published in peer-reviewed journals (Kamel et al. 2011; Kelly et al. 2012; Adamas-Rappaport et al. 2013; Jurjus et al. 2013; Pourshanazari et al. 2013; Sawatsky et al. 2013). Importantly, teaching activities and assessment development can also be used as educational research to generate new knowledge about teaching and assessment or to understand more about the interaction between teachers and learners.

A teaching/assessment scholar is often invited to provide professional development for colleagues or to participate in the peer review of other educators (Chandran et al. 2009a; Gusic et al. 2013). While these pursuits do not lead to tangible products per se, these activities do contribute to the community of educators and demonstrate recognition of the expertise and experience of an educational scholar.

Appropriate methods for educational discovery

The reader, by this point, may have identified a project to enhance his/her teaching and/or assessment strategies, but she/he may also want to confirm or understand the results of

Table 5. Some examples of teaching scholarship (based on the toolbox for evaluating educators) (Gusic et al. 2013).

Glassick's criteria (2000)	Criteria applied to a teaching *activity*	Evidence of a scholarly approach	Criteria applied to a learner assessment activity	Evidence of a scholarly approach
Clear goals	Learning objectives for the teaching session(s) are: • Stated clearly • Specified to measure learners' performance • At appropriate level for targeted learners	Educator has reviewed the content of courses delivered earlier in the curriculum to ensure that the objectives for this interaction build on what was previously learned and advance the learner's skills	Learner assessments: • Are appropriate for the content and level of learning objectives/competencies • Define expectations for learner's performance	Educator creates an assessment that requires the learner to apply knowledge in creating a list of possible diagnoses to explain a patient's clinical presentation; learner must be able to justify the selected diagnoses in order to pass the exam
Adequate preparation	• Congruence with institutional/program goals and integration with other components of curriculum • Use of best practices from the literature, professional development activities and personal experience • Resource planning	Educator attends professional development conference to learn about teaching strategies that promote active learning; meets with curriculum committee to gain approval to incorporate small group learning sessions in the course	• Congruence with institutional/program goals and integration with institution's system of assessment • Use of best practices from the literature, professional development activities and personal experience • Resource planning (facilities, faculty and schedules	Educator: creates the exam based on existing practice and in collaboration with others who are responsible for courses being delivered simultaneously in the curriculum; assessment is a component used in decision-making for progression to next stage in the curriculum; meets with curriculum committee and Dean's office to gain approval for time needed for new assessment format
Appropriate methods	• Teaching methods aligned with learning objectives • Methods are feasible, practical, ethical • Innovative teaching methods used to achieve objectives	Educator develops clinical cases with a clinician colleague for use in small group, teacher facilitated sessions in which learners apply their knowledge of pathology and pathophysiology to explain clinical findings and create a differential diagnosis for the case	• Assessment format aligned with learning objectives • Assessment process is consistent and uses accurate scoring methods • Assessment occurs in setting suitable for demonstration of relevant learning • Sufficient sample of the learner's performance collected to assure accurate capture of real ability/competency • Methods are useful, feasible, practical, ethical • Use of innovative assessment methods to measure performance	Educator works with assessment expert in the Dean's office to establish how many cases to use in the assessment and to create a scoring rubric for the assessment; develops a guide for all faculty who will score the exam
Significant results (Kirkpatrick & Kirkpatrick 2006)	• Satisfaction/reaction • Learning: Measures of knowledge, skills, attitudes and/or behaviors • Application: Desired performance demonstrated in other settings • Impact: On educational programs and processes within and/or outside institution	Educator: develops a survey to measure learners' perceptions about the effectiveness of the small group sessions in meeting the learning objectives and; develops a form for teachers to provide feedback to the learners after each small group session about how well they are articulating their thought processes and about the accuracy of their explanations of underlying biomedical principles during the discussions	• Satisfaction/reaction: Assessment evidence provides meaningful feedback about quality and implementation of assessment • Learning: Measures knowledge, skills, attitudes and/or behaviors • Application: Desired performance demonstrated in other settings • Impact: On progress decisions about learners and on educational programs and/or programs of assessment within and/or outside institution	The assessment provides quantitative and normative (relative to peers) data as well as narrative data that describes the learner's strengths and areas for improvement
Reflective critique	• Reflection and results of evaluations used for ongoing improvement	Educator collects data from learners and from small group teachers and uses information from the end of course assessment to modify the order of the cases used in the small courses to align with content being presented in other courses being presented simultaneously (using tools such as a teaching log)	• Reflection and results used for ongoing improvement of the assessment itself and/or the program of assessment	Educator analyzing aggregate performance data to evaluate the quality of each case used in the exam, reviewing the case content as well as the scoring rubric (using tools such as historical cohort comparisons)

the activity through the discovery paradigm. Going forward in this section, the discussion will be limited applied educational research rather than theoretical research.

In order to assist the reader with further classifying her/his educational discovery (research) questions, the authors suggest utilizing the following schema:

- Is the question about classifying and/or measuring variables that represent the current or past state of the phenomenon? If yes, then the question is likely a descriptive or historical one (this is a distinct category from the generic term "descriptive" scholarship applied to scholarship that may result from describing teaching activities) (Postlethwaite 2005; Ringsted et al. 2011).
- Is the question about relating one phenomenological variable to another or making predictions about one from another? If yes, then it is likely a correlative (associative) one (Postlethwaite 2005; Ringsted et al. 2011).
- Is the question about proving causation of the phenomenon? If yes, then it is likely a causal one (Postlethwaite 2005).
- Is the question one that tries to make meaning (i.e. explain) a phenomenon? If yes, then it is likely a qualitative one (Postlethwaite 2005; Ringsted et al. 2011).

Table 6 provides detailed examples for the first three categories. The descriptions of each category that follow it are intentionally conceptual and may need to be supplemented with other resources to build a fuller understanding. Also, to ensure the use of appropriate methods for any question, the authors recommend the reader consult or collaborate with experts or mentors with specific experience and/or expertise with the reader' specific program of study.

Descriptive studies

The aim of descriptive research is to establish the current or past state of an educational phenomenon by defining it, organizing it and measuring its variables (Postlethwaite 2005; Ringsted et al. 2011).

Descriptive studies attempt to describe and measure variables from a sample derived from a larger population; therefore, a researcher should always be interested in assuring that the sample is fully representative of the larger population and thus reducing sampling error (Umbach 2005). There are a number of ways to assure a representative sample (consecutive and block), but they all must assure that the sample reflects the population as a whole (including relevant subgroups) (Rea & Parker 2005). Often, researchers are at the mercy of a "convenience sample," thus placing the sampling error on the hopes of equitable data return for all representative subgroups in the population (Rea & Parker 2005).

Another troubling aspect of sampling with some descriptive studies is rate of return on self-reporting measures; because most subjects are voluntary (and often anonymous), it is not unheard of to have a 30% or less return rate requests for data (Livingston & Wislar 2012). Therefore, it is critical to capture all the subject characteristics (demographics and grades) whenever possible to be able to establish comparisons of the characteristics of the subjects in the sample to the target population. Another related risk is subject inconsistency in self reporting (Stone et al. 2000). If this risk is a concern, a researcher may build redundant items in his/her measurement tools or consider multiple measures of the same subjects (Stone et al. 2000).

The merit of descriptive research studies sometimes hinges upon the quality of their measurement tools; therefore, researchers must assure that the tools they use measure what they intend (valid) and are consistent (reliable) (Harden & Shumway 2003;Thorndike & Thorndike-Christ 2010; Ringsted et al. 2011). When the tools themselves are not always the focus of a correlative study (as in psychometric research, described below), researchers must assure that measurement tools and processes themselves have evidence of reliability and validity or that the tool development process is made transparent enough and systematic to allow peer judgment of these issues (Postlethwaite 2005; Umbach 2005).

The analytics of descriptive studies are usually not challenging as they are descriptive measures such as central tendency (mean, median mode) and range (e.g. SD). An important consideration of reporting descriptive research is how the data should be visually displayed; examples include tabular, box and whisker plots, pie charts and histograms (Rea & Parker 2005). The choice depends on what type of conclusions one is trying to illustrate.

Correlative studies

The aim of correlative research, also known as associative research, is to build quantitative associations between two or more variables and understand how these variables relate (Postlethwaite 2005; Umbach 2005). A confusing overlap exists between the terms correlational studies and correlational analysis (such as bivariate analysis). Correlational studies very often use correlational analyses, but they can also use other methods. For example, while attempting to establish validity for a particular test, one author compared test scores for the same test for three different learner groups using group mean differences (ANOVA), and then used a bivariate correlative analysis (Pearson) to compare one of the groups test sore to their scores on other tests (Crites et al. 2012).

The two main types of correlative studies are relational and predictive (Diem 2002). Relational analyses attempt to build associations between variables to provide evidence of immediate association (Lempp & Seale 2004). An example of this is a survey study of medical students' beliefs about a hidden curriculum currently in existence in their educational program (Lempp & Seale 2004). Predictive studies attempt to provide association through a temporal sequence; one variable to some degree can predict the presence of other(s) in the future. An example of a predictive study is predicting osteopathic student performance on a license exam based upon prior academic achievement with medical school classwork (Evans et al. 2003). One major educational sub-discipline that uses correlative studies is psychometrics, which establishes multiple lines of evidence to establish properties such as reliability and validity for assessment tools (Stone et al. 2000; Harden & Shumway 2003; Postlethwaite 2005).

In addition to errors inherent in sampling, data collection and measurement tools, correlative studies are often at risk of

Table 6. Three examples of educational discovery.

Glassick's criteria (2000)	Criteria applied to a *descriptive* inquiry	Criteria applied to a *correlative* inquiry	Criteria applied to a *causal* inquiry	Evidence of a scholarly approach
Clear goals	Example questions: *What are the mean and ranges of USMLE 1 scores by region?* *What do Program Directors believe the most important criteria for residency selection?*	*Example Questions:* *Which learning variables are associated with success on a genetics exam?* *What is the validity and reliability of a communication assessment tool in the context of OSCE exercises?*	*Example Questions:* *Is Team-based Learning a Superior to Lecture Formats for Teaching Evidence-based Medicine?* *Will students be able to understand better cellular transport mechanisms if an animated video are added to the existing teaching format?*	Educator has reviewed the FINER criteria and assured that questions address all requisite criteria for effective scholarly questions
Adequate preparation	Examples of one preparation step: Gain access to the restricted database Build survey; validate it	Examples of one preparation step: Recruit collaborators in the Genetics Department Apply for funding for tool development and field testing	Examples of one preparation step: Seek approval to change course Ask for coaching and support on video development	Investigator has built a comprehensive plan, which includes a comprehensive literature review
Appropriate methods	Example methods: Comparisons of means and ranges across regions Sampling check to assure representation of each type of medical school and specialty	Example methods: Capture learner specific variables (Table 7) and correlate them Use several methods to measure reliability and validity	Example methods: Compare means between two courses at two different sites Pre–post achievement score comparisons using same group	The investigator matched the appropriate design and methodology to the question
Significant results and effective presentation	Example results: There is little variation in scores across regions; Results appropriate for oral abstract Program directors use an unexpected mix of criteria to weight residency judgments; Results appropriate for peer-reviewed journal	Example results: Genetic scores correlate to specific underlying learner traits; Present paper during a basic science educator conference Tool has some evidence for reliability and validity; Submit tool to a peer-reviewed assessment repository	Example results: Group assigned to TBL instruction had slightly higher attainment; Submit for peer review journal Pre–post measures show overall group improvement; Submit to poster abstract during educator conference	The important results from the inquiry were summarized well and the dissemination process was well selected appropriate for audience and impact
Reflective critique	Example of critique: Discussion of research implications (from original paper and abstract) is added to section of Teaching Portfolio	Example of critique: Discussion of research implications (from original paper and abstract) is added to section of Teaching Portfolio	Example of critique: Discussion of research implications (from original paper and abstract) is added to section of Teaching Portfolio	Critique was balanced between how the scholarship adding to existing knowledge and its own inherent limitations

Table 7. Common variables that may be sources of biases and confounding in educational discovery scholarship.

Variables related to setting:

1. Educational mission and sub-missions of school: (Lewkonia 2001; Cohen 2009)
   a. Availability and quality of teaching facilities and educational support technologies
   b. Availability and quality of materials for teaching
2. Institutional ''culture'' and ''climate'' of learning (Genn 2001a, b)
3. Easily accessible, extracurricular ''alternate learning resources'' available to students during down time (co-interventions), including electronic and social media (Centre for Learning & Performance Technologies 2014)

Variables related to participants:

1. Preferred learning styles (Grasha 2002)
2. Stage of cognitive development (Merriam et al. 2007)
3. Prior experience with content domain and preparatory knowledge attainment (i.e. were prerequisites adequately attained) (Merriam et al. 2007)
4. Cultural-linguistic background (Adams & Strother-Adams 2008)
5. Metacognitive control of learning and learner dependency (Bruning et al. 2011)
6. Baseline attitudes toward learning domain (Ruggiero 1998)
7. Variation in cognitive abilities (Merriam et al. 2007)
8. Learning disabilities (Steinert 2008)

Variables related to intervention:

1. Instructor:
   a. Expertise with specific teaching methods and content (Grasha 2002)
   b. Preferred teaching style (Grasha 2002)
2. Teaching formats/methods and setting of learning (Grasha 2002)
3. Teaching tools, technology and task engagement (Smith & Ragan 1999)
4. Sequencing of learning tasks (Smith & Ragan 1999)
5. Explicitness of stated learning targets and their match to instruction (Kern et al. 2009)

confounding. Confounding is defined as a real mathematical relationship between two variables that may lead to an incorrect conclusion of an associative or causal relationship between the variables (Sonis 1998). The reason is that the two variables vary with a third variable (called a confounding variable) which actually has the true relationship with the first two variables (Sonis 1998). An analysis of educational level and heart disease in Swedish men found that the lower levels of education was linked to heart disease, but a subsequent analysis found that the most predictive variables for heart disease was cognitive, socioeconomic and behavioral factors, which were heavily associative (they confounded) the variable of low educational achievement (Falkstedt & Hemmingsson 2011). The best way to reduce risk from confounding in correlative studies is to partition (intentionally avoid sampling subjects with the confounder or censure their data) or collect data on several potentially related variables and use one of the multiple variance models (such as regression) to explore their interdependent relationships after data collection is complete (Sonis 1998). Some potentially confounding variables to consider when designing an educational discovery study are in Table 7.

Additional risks in correlative studies are chance, or random, error and observer variation. When comparing two or more groups or variables, correlative studies often generate probabilities (p values and confidence intervals) as a quantification of chance error (Sweet & Grace-Martin 2012). Because chance error can be impacted by sample size, researchers have another reason to be mindful of an adequate sample size (i.e. not just sampling for representation) (Leedy & Ormrod 2001). Rater observation measures, when utilized, are other sources of error; when using multiple raters, one should consider if some analysis of rater agreement is needed (e.g. interclass correlation and kappa) (Banerjee et al. 1999).

Causal analyses

Conclusions inferred from multiple correlative studies may provide sufficient evidence of a causative relationship between two or more variables and, therefore, imply more rigorous studies are unnecessary. More often, these methods lack the ability to fully account for the systematic errors of confounding and bias and thus fail to establish proof of causation. Although methods to account for known confounders were discussed earlier, many confounders are unknown or cannot be easily identified or measured and, therefore, may unintentionally impact any study design (Haynes et al. 2006). Bias is any conscious or unconscious action on behalf of subjects, researchers or outsiders to impact a study result which violates the study assumptions (Haynes et al. 2006). Examples of bias are subject selection/allocation errors that may impart more likelihood of an outcome for one group over another, allowing outside interventions (co-intervention) or cross-interventions between study groups (contamination), or watching comparison groups differently so an investigator is more likely identify outcomes in one group compared to another (Haynes et al. 2006). For these occasions, causal study designs are needed to definitively establish causality (Shadish et al. 2002).

The key concept of causal designs is that one variable is hypothesized to ''cause'' another (Shadish et al. 2002). If the first variable leads to a positive educational outcome (second variable) and first variable is structured in a way that can be replicated, then it could be recommended as an ''educational intervention'' (Hutchinson 1999). Educational interventions are complicated affairs, often made up several sequenced and structured activities (Hutchinson 1999). Thus, in order to devise a ''collective variable'' as an educational intervention, all the facets of the intervention need to be explicitly described with sufficient clarity to be replicated elsewhere (Hutchinson 1999). Some of the features to consider while devising educational interventions are in Table 7 when describing the generalizability of the intervention.

Beyond addressing sample size/representation, measurement/tool issues and chance to establish causation, researchers often need to address confounders and biases with more rigorous study designs (Shadish et al. 2002). Some educational researchers often quote the randomized controlled trial (RCT) as the best design to reduce risks from bias and confounding (Leedy & Ormrod 2001; Shadish et al. 2002; Ringsted et al. 2011). By randomizing subjects to different study groups, all the potential systematic errors at the beginning of a study (some biases and most confounders) present during initiation of a study are assumed to be distributed equally across groups at the beginning of the study and therefore will not impact the results (Leedy & Ormrod 2001; Shadish et al. 2002; Ringsted et al. 2011). Additionally, by careful monitoring of

learners to disallow co-interventions and contamination (the "control" in RCTs) during study delivery, biases are assumed to be further reduced (Leedy & Ormrod 2001; Shadish et al. 2002; Ringsted et al. 2011). However, there are limitations to an RCT design's feasibility in educational research (Sullivan 2011). First, it is often impossible randomize learners within one educational program due to ethical, logistical or educational policy limitations (Sullivan 2011). For example, it is often hard to justify to educational oversight bodies that, during a sponsored or accredited learning program, two or more learner groups receive different educational experiences (Sullivan 2011). Second, controlling human cognition and learning is nearly impossible in open educational settings. Some studies have shown that students often seek learning from alternative sources in between structured learning sessions (co-interventions) and cross-learning between cohorts when comparing different teaching interventions (contamination) (Howe et al. 2007). Therefore, it is more likely that one will use a clustered randomized model (if multiple sites used) or one of the available quasi-experimental designs if a single site is used (non-randomized comparative groups, pre–post-test designs, time series and their variants) (Leedy & Ormrod 2001; Shadish et al. 2002; Howe et al. 2007; Ringsted et al. 2011). Whether a randomized (a.k.a. experimental) design or a quasi-experimental design is used, the reader may want to measure any variables that have been shown to impact human cognition and learning and may impact her/his study design and adjust for them during the analysis phase (Table 7).

Sometimes, researchers want to suggest causation with smaller sample sizes, less-intense interventions or smaller units of study. These are called pilots or preliminary studies, and they can be a good first choice before committing significant time and resources to a larger study (Leedy & Ormrod 2001). The methods for these studies are generally less rigorous and conclusions are usually treated as potential hypotheses for future studies.

Qualitative approaches

Once they establish the causes of educational outcomes through quantitative analysis, researchers often still need to know how and why they happened (Patton 2001). Qualitative approaches have a broad array of traditions that use methods to dive further into a phenomenon to better understand why variables relate the way they do (beyond simply proving they have a mathematical relationship) (Patton 2001). The common theme to the qualitative approaches is that the investigator has to have proximity and engagement with the phenomenon (either close observation or participatory) over time to perform an adequate inquiry (Patton 2001).

Qualitative inquiry has many philosophical and methodological traditions, making it impossible to categorize and review them all. Common qualitative designs in education include case studies, phenomenology, grounded theory, ethnography and historical analyses (Patton 2001). These fully developed qualitative study designs are appropriate when the main purpose of a research project is deriving broad or deep and broad meaning from events, rather than inferring mathematical properties of their relationships.

A good qualitative inquiry is as rigorous as a quantitative one and skilled qualitative researchers use thoughtful (and defensible) sampling methods, engage thorough collection methods such as triangulation and data saturation, allow for reproducible analysis through methods such as recursive coding and theme building, and incorporate cross checks for consistency and confirmability (Otero & Harlow 2009). The nature of the inquiry and the emerging data determine the intensity and thoroughness of these steps. Since qualitative research is diverse and usually requires some level of expertise, consultation prior to beginning one of these designs is strongly advised.

Mixed methods

Educational researchers often need qualitative methods to supplement findings from one of the quantitative approaches described earlier. For example, a teacher may find his/her educational unit improved knowledge retention in a particular domain by causative design, but he/she is unable to find the underlying reason (which of all the events that happened inside or outside the learning environment account for the outcome?). In this event, he/she may want to use qualitative methods to derive a fuller understanding of the results found through quantitative analysis. Examples of common qualitative data collection methods used in mixed method studies include surveys, focus groups, interviews and participant observation.

The value of adding qualitative methods to a quantitative study is to capture all the significant data while the events occur. Depending upon the scope of the qualitative question(s), it may deserve its own study with one of the more rigorous and comprehensive approaches described above. Often, qualitative research questions in mixed methods studies are narrower in scope and can be handled with a less-resource intensive qualitative approach (e.g. surveys and focus groups).

Sometimes it is advisable to begin a qualitative inquiry prior to beginning a quantitative study. Reasons for such an approach may be to help clarify research questions or help better define variables prior to establishing research goals.

Significant results and effective presentation

Once a scholar develops a unit of scholarship, how does he/she best describe and share it? The decision of where and how to share scholarship can surface by identifying a few characteristics of the project:

- *Profundity of results*. Do the results challenge a deeply held educational belief, theory or practice? Does it address an accepted practices but in a novel way? Is the proposed project designed to describe or share an educational project that other educators may adopt or adapt?
- *Breadth of appeal*. How big is the audience for this result? Where and how do they typically share their scholarship?

Once the reader has considered these issues, she/he may want to explore all of the available options for dissemination, preferably with the help of a mentor. Many authors begin by submitting their work at educational conferences. There are several of these opportunities to submit your work, such as the

Association for Medical Education in Europe (AMEE, www.amee.org), American Association of Medical Colleges (AAMC, www.aamc.org), the International Association of Medical Science Educators (IAMSE, www.iamse.org) meetings and many discipline-based societies. Many of these conferences require peer reviews of submissions and this adds to scholarly merit and provides valuable feedback. It is advisable to collect observations and criticisms of your work and use them to your advantage by strengthening your work prior to submission for publication. There a variety of presentation types that you may consider for educational conferences. For example, an educational project in its early stages may be appropriate for a roundtable discussion or small group discussion. Indeed, discussing an educational topic and considering a potential project with colleagues may lead to a collaborative effort. Smaller scope projects may be appropriate for other presentation types, such as an oral paper presentation or abstracts (oral or poster).

A scholar may find the decision to submit his/her work for publication a daunting task. The Group on Education Affairs (GEA) of the Association of American Medical Colleges compiled a list of several journals and online resources that target medical education topics (AAMC-GEA-MESRE Section 2013). Some journals, such as *Obstetrics and Gynecology* and *Cell Biology Education* focus on both research and education specific to a specialty or discipline. Other journals, such as *The Clinical Teacher* and *Teaching and Learning in Medicine*, invite projects that lean toward a much broader educational audience. Other sources, like MedEdPORTAL, are designed to exchange peer-reviewed educational resources, such as tutorial, simulations and assessment tools. Submitting authors should read the journal's mission and review some of the recent publications of the journal to ascertain whether it is aligned with the purpose of the author's work. Judging both the purpose of a project and the extent the results are relevant to a focused or broad audience will help one ascertain potential options for publication.

Once a publication source is decided upon, one should make sure to follow closely the directions and submit using the proper format and procedures. While the review process varies, the editor typically reads the submission and decides whether to forward it for further review. The editor considers general characteristics of the paper, such as whether it matches the mission of the journal and the audience, and whether there are any fatal flaws in the work (Bordage et al. 2001). If the decision is to forward for further review, the reader will usually receive valuable feedback that may improve the quality of the manuscript.

Some journals or repositories, such as MedEdPortal, outline the guidelines for submission review; however, not all do. If the criteria are not clear, the authors recommend using Glassick's six criteria to judge adequacy (Glassick et al. 1997). Many of these issues have been addressed throughout the Guide, but it is appropriate to punctuate a few issues that can make the difference between an accepted and rejected submission.

First, authors should describe the goals of their work and make a convincing argument about how their work contributes to the existing literature. Early in the introduction of the paper, emphasize the problem statement that articulates the issues and context that gave rise to the study. McGaghie and colleagues have provided examples that a scholar can model, helping the reader to quickly recognize the importance of a scholarly work and how it can help coordinate a line of study about a phenomenon (Bordage et al. 2001; McGaghie 2009).

Second, make sure that the listed citations are deliberate. Too often, authors provide citations that seem to have little relevance with how a submission coordinates with existing literature. This is not to suggest that cited literature should always confirm results or support conclusions. Rather, one should ask how citations inform the reader about how your work aligns with what is known about a phenomenon or how your work will help provide new directions for investigation.

Third, some authors have difficulty connecting results with conclusions. A common pitfall, e.g. is misinterpreting a probability or "*p*" value. It is essential for authors to avoid conflating a 0.05 or below *p* value with conclusions about the "significance" of the results. Some authors refer to a statistical significance and make unsubstantiated claims about the importance of the findings. For example, a statistically significant correlation coefficient may actually explain little about how one variable explains the variation in another variable and do not correct mistakes made in studies that fail to address the several systematic errors described in the preceding section. In this sense, statistical tests are at risk for the same "garbage in, garbage out" phenomenon as seen in computing science. Many editors insist that authors qualify *p* values and confidence intervals with effect sizes, which can help provide vital information about the practical, educational and clinical significance of results (Colliver 2002). There are many types of effect sizes and some books that describe how to conduct statistical analyses often include guidelines on the interpretation and the reporting of effect sizes and confidence intervals (Thompson 2006; Field 2009). Helpful resources are also available to help figure effect sizes into tables (Nicol & Pexman 2010). It is wise, then, for authors to consider carefully how the results are reported and interpreted, which are the basis for a project's conclusions.

Finally, in order to be successful in sharing scholarship, the authors recommend that scholars have two traits: a thick skin and persistence. Many journal editors reject a high proportion of submissions and, therefore, reflecting upon the reasons for rejection can become an important part of the scholarly process. Authors should carefully consider feedback from reviewers or editors to improve the quality of their scholarship and use this feedback for further submissions. Not giving up is another important trait. The authors have seen people (including themselves) submit the same pieces five times and only get it on the sixth. The important thing to remember is that the peer-review process can ultimately help a project become more refined.

Getting credit and reflective critique

Generally, scholars create scholarship from their work because it reflects what they enjoy doing; however, it would be

wonderful if others appreciated it too. Educational scholarship is being held as an equivalent achievement with other scholarly activities at most institutions and many medical schools have separate tracks for educational scholars that value multiple forms of scholarship in decision-making about academic advancement (Bunton & Mallon 2007).

Each academic institution has its own idiosyncrasies about how different types of scholarship count towards promotion or tenure and how the different types are judged in the review process (Bunton & Mallon 2007). The reader will most likely still find promotion committee members talking about traditional metrics such as impact factors and journal prestige, but newer forms of peer-reviewed publication sources (e.g. MedEdPortal) and newer metrics (e.g. hits on an educational resource website) are challenging some promotions committees to expand their thinking. For these reasons, it is advisable that each reader understand what documentation is required and how the respective committee for his/her promotion pathway judges and weighs each type of scholarship. Sometimes the reader will find it necessary to educate promotion committee members on some of these more novel scholarly valuation metrics in the portfolio so his/her accomplishments can be fairly measured.

Promotion packets must be succinctly structured to effectively demonstrate attainment of the requirements for promotion, and soliciting the help of a mentor very early in the process is highly recommended. Understanding how committee members prefer formatting and organization of educational promotion packets is also important; therefore, opening individual dialogues with leaders of these committees or departmental chairs is a good way to accomplish this goal. Having a well-organized promotion packet that is well-documented and easily understood is critical to having a successful outcome.

Promotion boards often set the expectation that promotion packets include documentation and/or statements which show reflection by candidates. Examples of how these reflections are provided earlier in this Guide but, for the developmental portfolio (i.e. teaching portfolio), they should consist of brief statements about how the different pieces of scholarship referenced in the promotion packet (individual pieces or a line of inquiry with multiple pieces) has added to what is known, how it can be used by others, and what limitations it holds. For example, reflectivity on teaching scholarship might include statements about choices made during instructional design, which learner outcomes resulted from these choices, and how this new knowledge impacts teaching practices going forward. As an example for reflectivity on discovery (research), a scholar might mention choices made for research study design, how the quantitative and/or qualitative results linked to these choices and how these results add to what is known about the generalizability of educational phenomena. Developmental portfolios, when well-constructed and demonstrating high-quality reflection, can themselves be considered a piece of scholarship by promotion committees. Since instruction about portfolio construction is beyond the scope of this Guide, the authors suggest reviewing the references provided (Baldwin et al. 2008; Gusic et al. 2013).

Research ethics

The reader might be surprised to learn that educational scholarship often needs oversight by research committees. Over the last 70 years, there has been an expanded focus on the rights of human research subjects. Several summary proclamations, including the Nuremberg Code, the Belmont Report and the Helsinki Declaration, have set the guiding ethical principles for human subject protections (The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research 1979; HHS – Office for Human Research Protections 2005; World Medical Association 2008). Many countries have established laws, regulations or policies that reflect these principles and require the establishment of human research boards (also known as institutional review boards, or IRBs), whose task it is to review biomedical research protocols to assure compliance with the regulatory criteria. In some countries, the oversight has been extended to social and behavioral science research protocols, including educational research protocols.

Educational research subjects face the potential risks of psychological stress, loss of autonomy and loss of anonymity. Since most educational research procedures mimic educational procedures, the risk of psychological stress is usually minimal. Learners are often in an unequal power relationship with the institutions in which they study, so there is the potential risk that they may experience some loss of autonomy (coercion). Since educational data is captured and recorded, subjects and their personal information (demographics, grades and survey responses) are at risk of being exposed, even if precautions are taken (e.g. keeping research database with codes that do not identify individual subjects). For these reasons, several countries require that all researchers, including educational researchers, complete training related to research ethics and submit their research protocols for review by an oversight committee/review board.

Because of the variability of oversight requirements between countries and regions, it can be challenging for educational researchers to know when and to whom to submit a protocol review. Through a recent literature search, the authors found very few references collectively archiving the educational research oversight activities by region or country. Furthermore, the number of international oversight policies and bodies for biomedical research are growing rapidly, especially in developing countries. UNESCO has housed the Global Ethics Observatory (GEObs), a searchable database for finding ethical experts, institutes, ethics training and legislation by country or region, can provide a starting to point for the reader to find out more about the regulatory oversight obligations for her/his country (UNESCO 2013). Likewise, the International Bioethics Database has a searchable database of bioethics centers that may be able to provide guidance (Bioethics Research Library – Georgetown University 2013).

Human research committees across the globe may have different procedures for reviews, but generally they have similar options when determining whether to waive review ("exempt"), to perform a rapid type of review by one or few members (often called "expedited" for less than minimal risk protocols), and full board review for more than minimal risk

protocols (World Health Organization 2011). The US department of HHS describes minimal risk as "probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests" (U.S. Department of Health & Human Services 2009). The definition of minimal risk is intentionally vague to allow contextual interpretation of risk by IRB members; therefore, IRBs, not researchers, determine the level of risk for any proposal.

Another difficult issue is whether certain scholarly teaching products (e.g. assessment results or using specific learner feedback to evaluate teaching) require IRB review. A piece of scholarship derived from teaching does not technically meet the definition of research, but to many IRBs, it can look like quite like it. Teaching scholarly activities involve inquiry with humans, record human data in some form, may including sharing data with outsiders, and can produce generalizable knowledge (the sine qua non of "research" by some institutions). Each of these steps has potential risks for learners similar to risks for research subjects described earlier; therefore, some IRBs hold the prerogative to judge whether any activity meets their definition of research (requiring review) or not (exempted from review). For these reasons, it is wise to contact an IRB representative to get clarification whenever any doubt exists, especially when using activities with learners that are not described in an approved syllabus, sharing individual or group academic data with individuals not listed a faculty members for the unit of study, or prior to presenting or publishing data outside the home institution. Some peer reviewed journals now require IRB determination for every submission, regardless of scholarship type (Kanter 2009).

If the reader has no formally established IRBs or the local boards decline to review educational research protocols, she/he is still obligated to perform the research in an ethical manner. There are guides to performing ethically based research (U.S. Department of Health & Human Services 2009; World Health Organization 2011), and many of the sites described previously can provide options for more in depth training on research ethics. If the reader is doing assessment research, he/she can also refer to the Joint *Standards for Educational and Psychological Testing* (American Educational Research Association et al. 1999).

Conclusion

The authors created this Guide with the intention that early career teachers use it as a starting point for building a scholarly career. Through their engagement with the developmental literature and a dash of their own successes and failures, the authors synthesized a comprehensive and yet (hopefully) practical set of instructions for building educational scholarship. The recurring traits running through the Guide – respecting a tradition and its methods, staying organized, fitting scholarship into personal development, effectively utilizing peers and mentors, clearly communicating accomplishments, being critically reflective and understanding the rights of learners – all should contribute to a successful scholarly career and lead to excellent scholarship for readers. As readers progress through their scholarly careers, this Guide should become increasingly irrelevant to them; this, at least, is the hope of its authors.

Notes on contributors

GERALD E. CRITES, MD, Med, is an Associate Professor of Medicine and is the Director of Program Evaluation and Educational Research, Georgia Regents University/University of Georgia Medical Partnership in Georgia, U.S.A. Dr. Crites supports faculty and medical student educational scholarship and has research interests in assessment and evaluation.

JULIE K. GAINES, MLIS, is the Head of the Georgia Regents University/University of Georgia Medical Partnership Campus Library in Georgia, U.S.A. As a core educator, she is fully integrated in the medical school curriculum by teaching Evidence-Based Medicine topics and supports the faculty and students' research interests.

SCOTT COTTRELL, Ed.D., serves as the Associate Dean for Student Services and Curriculum in the Department of Medical Education at the West Virginia University School of Medicine in West Virginia, U.S.A. His research interests include curricular development and assessment. He enjoys teaching statistics and measurement.

SUMMERS KALISHMAN, PhD, Associate Professor in Family and Community Medicine, works in program evaluation, faculty development and workplace learning at the University of New Mexico School of Medicine.

MARYELLEN GUSIC, MD, as the Executive Associate Dean for Educational Affairs and is the Dolores and John Read Professor of Medical Education and Professor of Pediatrics at Indiana University School of Medicine in Indiana, U.S.A. Dr. Gusic's scholarly interests include evaluating the performance of educators and determining the impact of faculty development and mentoring.

BRIAN MAVIS, PhD, joined the Office of Medical Education Research and Development (OMERAD) in the College of Human Medicine at Michigan State University in 1992 and has been the director of OMERAD since 2004. He developed a student performance and outcomes database to support program evaluation and facilitate educational scholarship.

STEVEN J. DURNING, MD, PhD, is professor of medicine and pathology at the Uniformed Services University (USU). He directs the Introduction to Clinical Reasoning Course and is a general internist.

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

References

AAMC-GEA-MESRE Section. 2013. Annotated bibliography of journals for educational scholarship [Online]. Washington, DC: Association of American Medical Colleges. [Accessed 15 July 2013] Available from https://www.aamc.org/download/184694/data/annotated_bibliography_of_journals.pdf.

AAMC-Group on Educational Affairs. 2013. Group on educational affairs (GEA) [Online]. Washington, DC: Association of American Medical Colleges. [Accessed 5 September 2013] Available from https://www.aamc.org/members/gea.

Adamas-Rappaport WJ, Waer AL, Teeple MK, Benjamin MA, Glazer ES, Sozanski J, Poskus D, Ong E. 2013. A comparison of unguided vs guided case-based instruction on the surgery clerkship. J Surg Educ 70(6):821–825.

Adams JQ, Strother-Adams P. 2008. Dealing with diversity: The anthology. Dubuque, IA: Kendall Hunt Publishing.

Albert Einstein College of Medicine of Yeshiva University. 2013. Office of academic appointments – Clinician-educator track [Online]. Bronx, NY:

Yeshiva University. [Accessed 5 September 2013] Available from http://www.einstein.yu.edu/administration/academic-appointments/academic-tracks/clinician-educator.aspx.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education & Joint Committee on Standards for Educational, Psychological Testing (U.S.). 1999. Standards for educational and psychological testing. Washington, DC: American Educational Research Association.

Awaya A, McEwan H, Heyler D, Linsky S, Lum D, Wakukawa P. 2003. Mentoring as a journey. Teach Teach Educ 19(1):45–56.

Baldwin C, Chandran L, Gusic M. 2011. Guidelines for evaluating the educational performance of medical school faculty: Priming a national conversation. Teach Learn Med 23(3):285–297.

Baldwin CD, Gusic M, Chandran L. 2008. Leadership lesson: The educator portfolio: A tool for career development. Faculty vitae [Online]. [Accessed 2 January 2014] Available from https://www.aamc.org/members/gfa/faculty_vitae/148574/educator_portfolio.html.

Banerjee M, Capozzoli M, McSweeney L, Sinha D. 1999. Beyond Kappa: A review of interrater agreement measures. Can J Stat 27(1):3–23.

Beattie DS. 2000. Expanding the view of scholarship: Introduction. Acad Med 75(9):871–876.

Bioethics Research Library – Georgetown University. 2013. International Bioethics Organizations Database [Online]. Washington, DC: Bioethics Research Library – Georgetown University. [Accessed 27 September 2013] Available from http://bioethics.georgetown.edu/databases/organizations/.

Bordage G. 2009. Conceptual frameworks to illuminate and magnify. Med Educ 43(4):312–319.

Bordage G. 2010. Teaching: Scholarly teaching and scholarship of teaching making it count twice [Online]. Boston, MA: Partners Healthcare. [Accessed 5 September 2013] Available from http://www.partners.org/about/media-center/videos/grand-rounds-scholarly-teaching.aspx.

Bordage G, Caelleigh AS, Steinecke A, Bland CJ, Crandall SJ, McGaghie WC, Pangaro LN, Penn G, Regehr G, Shea JS. 2001. Review criteria for research manuscripts. Acad Med 76(9):897–978.

Bou Akl I, Ghaddar F, Sabra R, Parmelee D, Simaan JA, Kanafani ZA, Zgheib NK. 2012. Teaching clinical pharmacology using team-based learning: A comparison between third- and fourth-year medical students. J Clin Pharmacol 52(12):1806–1814.

Boyer EL. 1990. Scholarship reconsidered: Priorities of the professoriate. Princeton, NJ: Carnegie Foundation for the Advancement of Teaching.

Bruning RH, Schraw GJ, Norby MM. 2011. Cognitive psychology and instruction. Boston, MA: Allyn & Bacon/Pearson.

Bunton SA, Mallon WT. 2007. The continued evolution of faculty appointment and tenure policies at U.S. medical schools. Acad Med 82(3):281–289.

Centre for Learning & Performance Technologies. 2014. 100+ examples of use of social media for learning [Online]. Centre for Learning & Performance Technologies. [Accessed 3 October 2013] Available from http://c4lpt.co.uk/social-learning-handbook/100-examples-of-use-of-social-media-for-learning/.

Challis M. 2000. AMEE Medical Education Guide No. 19: Personal learning plans. Med Teach 22(3):225–236.

Chandran L, Gusic M, Baldwin C, Turner T, Zenni E, Lane JL, Balmer D, Bar-On M, Rauch DA, Indyk D. 2009a. APA Educator Portfolio Analysis Tool [Online]. MedEdPORTAL. [Accessed 15 October 2013] Available from https://www.mededportal.org/publication/1659.

Chandran L, Gusic M, Baldwin C, Turner T, Zenni E, Lane JL, Balmer D, Bar-On M, Rauch DA, Indyk D, Gruppen LD. 2009b. Evaluating the performance of medical educators: A novel analysis tool to demonstrate the quality and impact of educational activities. Acad Med 84(1):58–66.

Cohen JJ. 2009. Revisiting the medical school educational mission at a time of expansion. In: Hager M, Russell S, editors. Revisiting the medical school educational mission at a time of expansion October 2008 Charleston, SC. New York: Josiah Macy, Jr. Foundation. pp 14–21.

Colliver JA. 2002. Call for greater emphasis on effect-size measures in published articles in Teaching and Learning in Medicine. Teach Learn Med 14(4):206–210.

Creswell JW. 2009. Research design: Qualitative, quantitative, and mixed methods approaches. Los Angeles: Sage.

Crites GE, Markert RJ, Goggans DS, Richardson WS. 2012. Local development of MCQ tests for evidence-based medicine and clinical decision making can be successful. Teach Learn Med 24(4):341–347.

Detsky AS, Baerlocher MO. 2007. Academic mentoring – How to give it and how to get it. JAMA 297(19):2134–2136.

Diem KG. 2002. Using research methods to evaluate your extension program. J Extension [Online], 40. Available from http://www.joe.org/joe/2002december/a1.shtml.

Evans P, Goodson LB, Schoffman SI. 2003. Relationship between academic achievement and student performance on the Comprehensive Osteopathic Medical Licensing Examination-USA level 2. J Am Osteopath Assoc 103(7):331–336.

Falkstedt D, Hemmingsson T. 2011. Educational level and coronary heart disease: A study of potential confounding from factors in childhood and adolescence based on the Swedish 1969 conscription cohort. Ann Epidemiol 21(5):336–342.

Field A. 2009. Discovering statistics using SPSS: (and sex and drugs and rock 'n' roll'). Thousand Oaks, CA: SAGE Publications.

Fincher RM, Simpson DE, Mennin SP, Rosenfeld GC, Rothman A, McGrew MC, Hansen PA, Mazmanian PE, Turnbull JM. 2000. Scholarship in teaching: An imperative for the 21st century. Acad Med 75(9):887–894.

Garvey B, Stokes PK, Megginson D. 2009. Coaching and mentoring: Theory and practice. London: SAGE.

Genn JM. 2001a. AMEE Medical Education Guide No. 23 (Part 1): Curriculum, environment, climate, quality and change in medical education – A unifying perspective. Med Teach 23(4):337–344.

Genn JM. 2001b. AMEE Medical Education Guide No. 23 (Part 2): Curriculum, environment, climate, quality and change in medical education – A unifying perspective. Med Teach 23(5):445–454.

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(9):877–880.

Glassick CE, Huber MT, Maeroff GI. 1997. Scholarship assessed: Evaluation of the professoriate: An Ernest L. Boyer project of the Carnegie Foundation for the advancement of teaching. San Francisco: Jossey-Bass.

Grasha AF. 2002. Teaching with style: A practical guide to enhancing learning by understanding teaching and learning styles. Pittsburgh: Alliance Publishers.

Gusic M, Amiel J, Baldwin C, Chandran L, Fincher R, Mavis B, O'Sullivan P, Padmore J, Rose S, Simpson D, et al. 2013. Using the AAMC toolbox for evaluating educators: You be the judge! [Online]. MedEdPORTAL. [Accessed 15 November 2013] Available from https://www.mededportal.org/publication/9313.

Gusic M, Chandran L, Balmer D, D'Alessandro D, Baldwin C. 2007. Educator portfolio template of the academic pediatric association's educational scholars program [Online]. MedEdPORTAL. [Accessed 15 October 2013] Available from https://www.mededportal.org/publication/626.

Hammick M, Olckers L, Campion-Smith C. 2009. Learning in interprofessional teams: AMEE Guide no 38. Med Teach 31(1):1–12.

Harden RM, Shumway JM. 2003. AMEE Guide No. 25: The assessment of learning outcomes for the competent and reflective physician. Med Teach 25(6):569–584.

Haynes RB, Sackett DL, Guyatt GH, Tugwell P. 2006. Clinical epidemiology: How to do clinical practice research. Philadelphia: Lippincott Williams & Wilkins.

Health Resources and Services Administration. 2013. Grants – open opportunities [Online]. Washington, DC: U.S. Department of Health and Human Services. [Accessed 5 September 2013] Available from http://hrsa.gov/grants/index.html.

HHS – Office for Human Research Protections. 2005. The Nuremberg code [Online]. Washington, DC: U.S. Department of Health & Human Services. [Accessed 27 September 2013] Available from http://www.hhs.gov/ohrp/archive/nurcode.html.

Hicks R, McCracken J. 2009. Mentoring vs. coaching – Do you know the difference? Phys Exec 35(4):71–73.

Howe A, Keogh-Brown M, Miles S, Bachmann M. 2007. Expert consensus on contamination in educational trials elicited by a Delphi exercise. Med Educ 41(2):196–204.

Hutchings P, Shulman LS. 1999. The scholarship of teaching. Change 31(5):10.

Hutchinson L. 1999. Evaluating and researching the effectiveness of educational interventions. BMJ 318(7193):1267–1269.

Jurjus RA, Dimorier K, Brown K, Slaby F, Shokoohi H, Boniface K, Teresa Liu Y. 2013. Can anatomists teach living anatomy using ultrasound as a teaching tool? Anat Sci Educ [Online]. [Accessed 10 December 2013] Available from http://onlinelibrary.wiley.com/doi/10.1002/ase.1417/abstract.

Kamel H, Dhaliwal G, Navi BB, Pease AR, Shah M, Dhand A, Johnston SC, Josephson SA. 2011. A randomized trial of hypothesis-driven vs screening neurologic examination. Neurology 77(14):1395–1400.

Kanter SL. 2009. Ethical approval for studies involving human participants: Academic Medicine's new policy. Acad Med 84(2):149–150.

Kelly W, Durning S, Denton G. 2012. Comparing a script concordance examination to a multiple-choice examination on a core internal medicine clerkship. Teach Learn Med 24(3):187–193.

Kern DE, Thomas PA, Hughes MT. 2009. Curriculum development for medical education: A six-step approach. Baltimore, MD: Johns Hopkins University Press.

Kirkpatrick DL, Kirkpatrick JD. 2006. Evaluating training programs: The four levels. San Francisco, CA: Berrett-Koehler.

Leedy PD, Ormrod JE. 2001. Practical research: Planning and design. Upper Saddle River, NJ: Merrill Prentice Hall.

Lempp H, Seale C. 2004. The hidden curriculum in undergraduate medical education: Qualitative study of medical students' perceptions of teaching. BMJ 329(7469):770–773.

Lewkonia RM. 2001. The missions of medical schools: The pursuit of health in the service of society. BMC Med Educ 1(4):1–5.

Lingard L, Vanstone M, Durrant M, Fleming-Carroll B, Lowe M, Rashotte J, Sinclair L, Tallett S. 2012. Conflicting messages: Examining the dynamics of leadership on interprofessional teams. Acad Med 87(12):1762–1767.

Livingston EH, Wislar JS. 2012. Minimum response rates for survey research. Arch Surg 147(2):110. doi: 10.1001/archsurg.2011.2169.

Macaffee D, Garvey B. 2010. Mentoring and coaching: What's the difference? BMJ careers [Online]. [Accessed 3 October 2013] Available from http://careers.bmj.com/careers/advice/view-article.html?id=20001204.

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

McLean M, Cilliers F, Van Wyk JM. 2008. Faculty development: Yesterday, today and tomorrow. Med Teach 30(6):555–584.

Merriam SB, Caffarella RS, Baumgartner L. 2007. Learning in adulthood: A comprehensive guide. San Francisco: Jossey-Bass.

Michaelsen LK, Sweet M. 2008a. The essential elements of team-based learning. New Dir Teach Learn 2008(116):7–27.

Michaelsen LK, Sweet M. 2008b. Fundamental principles and practices of team-based learning. In: Michaelsen LK, Parmelee DX, Mcmahon KK, Levine RE, editors. Team-based learning for health professions education. Sterling, VA: Stylus. pp 9–31.

National Board of Medical Examiners. 2013. Stemmler Medical Education Research Fund [Online]. Philadelphia, PA: National Board of Medical Examiners. [Accessed 5 September 2013] Available from http://www.nbme.org/research/stemmler.html.

Nicol AAM, Pexman PM. 2010. Presenting your findings a practical guide for creating tables. Washington, DC: American Psycholog. Assoc.

Novak JD, Gowin DB. 1984. Learning how to learn. Cambridge: Cambridge University Press.

Novak JD, Musonda D. 1991. A twelve-year longitudinal study of science concept learning. Am Educ Res J 28(1):117–153.

Otero VK, Harlow DB. 2009. Getting started in qualitative physics education research. In: Henderson C, Harper KA, editors. Reviews in PER Volume 2: Getting started in physics education research. College Park, MD: American Association of Physics Teachers. pp 59–60.

Palladino C, Gusic M, Fincher R, Hafler J. 2013. Educational scholarship in medical education. In: Walsh K, editor. Oxford textbook of medical education. Oxford: Oxford University Press. pp 658–688.

Patton MQ. 2001. Qualitative research & evaluation methods. Thousand Oaks, CA: Sage Publications.

Postlethwaite TN. 2005. Educational research: Some basic concepts and terminology [Online]. Paris: UNESCO International Institute for Educational Planning. [Accessed 15 November 2013] Available from http://www.iiep.unesco.org/fileadmin/user_upload/Cap_Dev_Training/Training_Materials/Quality/Qu_Mod1.pdf.

Pourshanazari AA, Roohbakhsh A, Khazaei M, Tajadini H. 2013. Comparing the long-term retention of a physiology course for medical students with the traditional and problem-based learning. Adv Health Sci Educ Theory Pract 18(1):91–97.

Rea LM, Parker RA. 2005. Designing and conducting survey research: A comprehensive guide. San Francisco: Jossey-Bass.

Reeves S, Albert M, Kuper A, Hodges BD. 2008. Why use theories in qualitative research? BMJ 337:a949.

Ringsted C, Hodges B, Scherpbier A. 2011. 'The research compass': An introduction to research in medical education: AMEE Guide no. 56. Med Teach 33(9):695–709.

Ross MT, Cameron HS. 2007. Peer assisted learning: A planning and implementation framework: AMEE Guide no. 30. Med Teach 29(6):527–545.

Ruggiero VR. 1998. Changing attitudes: A strategy for motivating students to learn. Boston: Allyn and Bacon.

Sambunjak D, Straus SE, Marusic A. 2006. Mentoring in academic medicine: A systematic review. JAMA 296(9):1103–1115.

Sandars J, Cleary TJ. 2011. Self-regulation theory: Applications to medical education: AMEE Guide No. 58. Med Teach 33(11):875–886.

Sawatsky AP, Mikhael JR, Punatar AD, Nassar AA, Agrwal N. 2013. The effects of deliberate practice and feedback to teach standardized handoff communication on the knowledge, attitudes, and practices of first-year residents. Teach Learn Med 25(4):279–284.

Schuwirth LW, van der Vleuten CP. 2011. General overview of the theories used in assessment: AMEE Guide No. 57. Med Teach 33(10):783–797.

Seehusen DA, Weaver SP. 2009. Resident research in family medicine: Where are we now? Fam Med 41(9):663–668.

Shadish WR, Chelimsky E, Cook TD, Campbell DT. 2002. Experimental and quasi-experimental designs for generalized causal inference. Boston, MA: Houghton Mifflin.

Shulman LS. 1993. Teaching as community property. Change 25(6):6.

Simpson D, Fincher R, Hafler J, Irby D, Richards B, Rosenfeld G, Viggiano T. 2007a. Advancing educators and education: Defining the components and evidence of educational scholarship. American Association of Medical Colleges Group on Educational Affairs Consensus Conference on Educational Scholarship, 9–10 February 2006 Charlotte, NC. Washington, DC: Association of American Medical Colleges.

Simpson D, Fincher RM, Hafler JP, Irby DM, Richards BF, Rosenfeld GC, Viggiano TR. 2007b. Advancing educators and education by defining the components and evidence associated with educational scholarship. Med Educ 41(10):1002–1009.

Simpson D, Marcdante K, Fenzel J. 2007c. The educator's portfolio & curriculum vitae – Workshop & resource guide [Online]. MedEdPORTAL. [Accessed 15 October 2013] Available from https://www.mededportal.org/publication/677.

Smith PL, Ragan TJ. 1999. Instructional design. New York: Wiley.

Sonis J. 1998. A closer look at confounding. Fam Med 30(8):584–588.

Stake RE. 1995. The art of case study research. Thousand Oaks: Sage Publications.

Steinert Y. 2008. The "problem" junior: Whose problem is it? BMJ 336(7636):150–153.

Stone AA, Turkkan JS, Bachrach CA, Jobe JB, Kurtzman HS, Cain VS, editors. 2000. The science of self-report: Implications for research and practice. Mahwah, NJ: Lawrence Erlbaum.

Sullivan GM. 2011. Getting off the "gold standard": Randomized controlled trials and education research. J Grad Med Educ 3(3):285–289.

Sweet SA, Grace-Martin K. 2012. Data analysis with SPSS: A first course in applied statistics. Boston: Allyn & Bacon.

The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. 1979. The Belmont Report

[Online]. Washington, DC: U.S. Department of Health & Human Services. [Accessed 27 September 2013] Available from http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html.

Thompson B. 2006. Foundations of behavioral statistics: An insight-based approach. New York: Guilford Press.

Thorndike RM, Thorndike-Christ T. 2010. Measurement and evaluation in psychology and education. Upper Saddle River, NJ: Merrill.

U.S. Department of Health & Human Services. 2009. Code of federal regulations – Title 45 Part 46 [Online]. [Accessed 5 September 2013] Available from http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html.

Umbach PD. 2005. Getting back to the basics of survey research. New Dir Inst Res 2005(127):91–100.

UNESCO. 2013. Global Ethics Observatory (GEObs) [Online]. Paris: UNESCO. [Accessed 27 September 2013] Available from http://www.unesco.org/new/en/social-and-human-sciences/themes/global-ethics-observatory.

University of Ottawa. 2013. Office of Professional Affairs [Online]. Ottawa, ON: University of Ottawa. [Accessed September 5 2013] Available from http://www.med.uottawa.ca/ProfessionalAffairs/eng/faculty_career_paths.html.

West DC, Pomeroy JR, Park JK, Gerstenberger EA, Sandoval J. 2000. Critical thinking in graduate medical education: A role for concept mapping assessment? JAMA 284(9):1105–1110.

World Health Organization. 2011. Standards and operational guidance for ethics review of health-related research with human participants. Geneva: World Health Organization.

World Medical Association. 2008. WMA Declaration of Helsinki – Ethical principles for medical research involving human subjects [Online]. Ferney-Voltaire, France: World Medical Association, Inc. [Accessed 27 September 2013] Available from http://www.wma.net/en/30publications/10policies/b3/.

Yardley S, Teunissen PW, Dornan T. 2012. Experiential learning: AMEE Guide No. 63. Med Teach 34(2):e102–e115.