

Quality Guidelines for Authors and Reviewers

Manuscript Quality – *PRiMER* seeks to publish papers that rigorously evaluate educational interventions and learning outcomes or test for behavioral changes resulting from an educational intervention. While we will consider manuscripts that the *Annals of Family Medicine* or *Family Medicine* might reject, this is not because we lower the level of rigor expected in the study design or execution. *PRiMER* will consider small studies, pilot projects, single-institution studies, work that seeks to replicate or confirm findings that are already known, or research that explores a broadly known construct in a new context. Print journals may be less likely to accept these manuscripts due to printing costs and concerns that they are not sufficiently impactful in the field. Published papers in *PRiMER* must still contribute knowledge that incrementally adds to what is known about a topic or phenomenon. We will consider quantitative, qualitative, and mixed methods submissions that are sufficiently rigorous.

We generally will not publish “reactions” to new curricula or interventions (eg, surveys of learners about satisfaction with a course) unless the learning modality or instructional content is new, there is generalizable knowledge to be gained by instructors at other institutions, and key portions of the instructional materials are made available (typically by references to a stable format for publication of online content, such as the STFM Resource Library).

Manuscripts outside the scope of medical education may also be considered for *PRiMER* publication if the first author is a medical student or a family medicine resident, and the content is pertinent to family medicine as a broad discipline. Resident and student manuscripts will be held to the same quality standards outlined below.

For a general overview of research methods, consider this free online methods guide: The Research Methods Knowledge Base (<https://www.socialresearchmethods.net/kb/>)

The following elements should be included in all submissions when applicable:

- **Data sources** must be fully identified
- **Human subject recruitment** procedures must be described, if applicable, including:
 - Institutional Review Board (IRB) interactions (see statement below)
 - Consent processes
 - Participation Incentives
 - Recruitment procedures, description of outreach, etc.
- Description of the **sample characteristics**, and as appropriate, a comparison of the sample to the population the sample should represent, or from which it is drawn
- **Limitations:** Given the nature and mission of *PRiMER*, we expect that submissions will generally have extensive limitations. Rather than aiming for flawless studies, we prefer to have a comprehensive description of all limitations, to allow complete assessment of quality by reviewers, and fully informed interpretation by readers. The statement of limitations is included as part of the Discussion section, as it is in most scientific journals.

- A **statement about an interaction with an IRB** should be present in most studies, except as noted below. As a general rule, studies that utilized publically available data (eg, public CDC data sets, reviews of published literature, observation of public activity, opinion or theory pieces, etc) do not require any IRB interaction. All other studies require IRB interaction. All manuscripts that require an interaction with an IRB should stipulate which type of interaction occurred:
 - IRB determination that a project is “not research” (eg, quality improvement or assurance) or “not human subjects research” (eg, biological samples of deceased individuals).
Example: “This project was determined by ##### IRB to constitute a quality improvement activity, and not human subjects research.”
 - IRB determination that a project is research, but exempt from review¹
Example: “As an anonymous survey, this project was determined to be exempt from review by ### IRB, citing exemption #2.”
 - IRB review (expedited or full review). Example: “This project was reviewed and approved by ### IRB.”

Manuscripts are expected to consist of the following structural elements, in order:

1. An **Introduction** and literature review should lead to the study question(s). The literature review should include and refer to appropriate literature. There should be a description of the research question(s) and hypotheses to be tested or problems to be solved by the project
2. A **Methods** section should be appropriate to answer the study questions. The methods should predict all reported results (eg, if the authors say they conducted a t-test, the results of the t-test should appear in the results).
3. A **Results** section should be directly tied to all steps in the methods section.
4. A **Discussion** section should explain the meaning of the results and help readers place the research findings in appropriate context. The discussion section should not take “flights of fancy.” More information on writing an effective discussion section is available at: <http://www.rcjournal.com/contents/10.04/10.04.1238.pdf>. The Discussion section should also contain a clear and complete description of study limitations.
5. **References** should be listed in standard format, and follow AMA style. (consistent with *Family Medicine*).
6. **Appendices (optional)**: Material that might go into an appendix should be handled by having the author submit the item (a curricular description, survey form, etc) to the STFM Resource Library, and then cite the resource (along with all other citations). Reference #7 in this article is a good example:
<https://journals.stfm.org/primer/2017/prunuske-0002/>.

Use of the Kirkpatrick Model of Assessment

Although manuscripts do not need to refer directly to the Kirkpatrick Model (KM) of Assessment,² it is helpful for reviewers, associate editors, and authors to have a basic understanding of the four levels at which outcomes might be assessed, according to KM. Briefly, the four KM levels are:

¹ <https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/index.html#46.101>

² <http://www.kirkpatrickpartners.com/Our-Philosophy/The-Kirkpatrick-Model>

1. Level 1: Reaction—*The degree to which participants find the training favorable, engaging and relevant to their jobs.*
 - This level might be thought of as the typical conference feedback form, eg, a “post-only” assessment of how the learner “liked” or experienced a training or educational event.
 - For PRiMER, we will only rarely publish “reactions”, and then only if the learning modality or instructional content is new, there is generalizable knowledge to be gained by instructors at other institutions, and key portions of the instructional materials are made available (typically by references to a stable format for publication of online content, such as the STFM Resource Library). Level 1 (reaction) studies are more likely to be considered if done as a rigorous qualitative process.
2. Level 2: Learning—*The degree to which participants acquire the intended knowledge, skills, attitude, confidence and commitment based on their participation in the training.*
 - This is often measured by a “pre/post” design, where students complete a process (such as a test or survey), are exposed to an intervention, and then resubmit the same test or survey immediately after. This model is often affected by maturation, by quick reinforcement or “teaching to the test,” and other threats to validity.
 - For PRiMER, learning MUST be assessed beyond simple pre-test followed by nearly immediate post-test. Examples of acceptable forms of learning assessment may include:
 - Performance on a standardized process (eg, United States Medical Licensing Examination [USMLE], Family Medicine Computer-Assisted Simulations for Educating Students [fmCASES], a regularly administered institutional examination, a periodic and regular survey, etc), and comparison with a nonexposed or differently exposed cohort (eg, comparison of students exposed to an intervention, vs unexposed, on board scores or an Objective Structured Clinical Examination [OSCE])
 - Posttest at an extended interval, on a test or instrument that measures a construct that was intended to be affected by the intervention (ie, assessment of long-term retention of knowledge).
3. Level 3: Behavior—*The degree to which participants apply what they learned during training when they are back on the job.* For example, if residents and faculty were given comprehensive education and feedback about opioid prescribing in a family medicine office, a decreased number of opioid prescriptions would be a change in behavior.
4. Level 4: Results—*The degree to which targeted outcomes occur as a result of the training and the support and accountability package.* For example, if residents and faculty were given comprehensive education and feedback about opioid prescribing in a family medicine office, a decrease in patient emergency room visits for opioid overdose would be considered a result.

Although researchers should use a rigorous approach to educational assessment, some educational interventions will not achieve their goals, or will have unintended consequences. Performance, behavior, and results should be measured, but they may not improve. The journal welcomes submissions with unexpected or negative results.

For quantitative studies, consider:

- Is the number of units used in the sample study (N) very small? Manuscripts that try to do “too much” with a very small sample may be problematic. We do accept small studies, but they should be handled appropriately for their sample size. Examples of good practices are:
 - The reliance on purely descriptive statistics if inferential (ie *P*-value generating) procedures would be underpowered
 - The use of statistical procedures that are intended for small samples (eg Fisher’s Exact Test)
 - Emphasizing qualitative analyses and results for studies that have them
- Is the statistical test appropriate?
- Have the authors utilized the best procedures (not just appropriate), and have they controlled, either experimentally or statistically, for confounding factors?
- Is a comparison group appropriate, and if so, was a comparison group (of any sort) used? A comparison with self (eg, a time series), or a simple descriptive study, may not require a separate comparison group.
- Are the instruments employed valid and reliable? How do the authors know?

For qualitative studies, consider:

- Methods across qualitative studies can vary extensively. This is appropriate. Ideally, a qualitative study will describe a broad type of study (eg, “grounded theory”, “phenomenological”, “content analysis”, etc)
- There are basic components that should be present in all qualitative studies:
 - Data collection procedures or sources
 - Sampling methods
 - Analytic processes
 - Transcription procedures
 - Some statements about control or consideration of biases
 - A general sense that the authors have not selectively identified only quotes that fit their research question
- Additionally, the issues that apply to quantitative studies (described above) may be applied to qualitative study assessment, with appropriate consideration for how those requirements fit the qualitative model described.