Evaluating Empirical Research

It is important to maintain an objective and respectful tone as you evaluate others' empirical studies. Keep in mind that study limitations are often a result of the epistemological limitations of real life research situations rather than the laziness or ignorance of the researchers. It's normal and expected for empirical studies to have limitations. Having said that, it's part of your job as a critical reader and a smart researcher to provide a fair assessment of research done on your topic.

B.A. Maher's guidelines (as cited in Cone and Foster, 2008, pp 104-106) are useful in thinking of others' studies, as well as for writing up your own study.

Here are the recommended questions to consider as you read each section of an article. As you think of other questions that are particularly important to your topic, add them to the list.

Introduction

- Does the introduction provide a strong rationale for why the study is needed?
- Are research questions and hypotheses clearly articulated? (Note that research questions are often presented implicitly within a description of the purpose of the study.)

Method

- Is the method described so that replication is possible without further information
- **Participants**
    - Are subject recruitment and selection methods described?

- Were participants randomly selected? Are there any probable biases in sampling?
- Is the sample appropriate in terms of the population to which the researchers wished to generalize?
- Are characteristics of the sample described adequately?
- If two or more groups are being compared, are they shown to be comparable on potentially confounding variables (e.g. demographics)? If they are not comparable, is this handled appropriately?
- Was informed consent obtained?
- Was the size of the sample large enough for the number of measures and for the effect being sought?

- **Design**
  - If appropriate, was a control group used?
  - Was the control appropriate?
  - What was being controlled for?
  - If an experimental study, were participants randomly assigned to groups?

- **Measures**
  - For all measures (measures used to classify, dependent variables, etc.), did the authors provide evidence of reliability and validity, either by summarizing data or by referring the reader to an available source that provides the information?

- Do the reliability and validity data justify the use of the measures? Specific evidence is particularly important if a measure is created just for this study.
- Do the measures match the research questions and hypotheses being addressed?
- If different tasks or measures are used, was their order counterbalanced? Do the authors analyze for potential order effects?
- Are multiple measures used, particularly those that sample the same domains or constructs but with different methods (e.g., self-report, rating by others, self monitoring, or direct observation)?
- If human observers, judges, or raters were involved, was interobserver or interrater agreement (reliability) assessed? Was it obtained for a representative sample of data? Did the two raters do their ratings independently? Was their reliability satisfactory?

♦ **Bias and Artifacts**

- Was administration and scoring of the measures done blindly (i.e., by someone who was unaware of experimental hypotheses)?
- If a quasi-experimental study, do authors include appropriate steps to rule out competing explanations of the findings?
- Were procedures constant across participants in all groups? Were any confounds introduced as the result of using different procedures? How troublesome are these?

♦ **Independent variables**

- If an experimental study, was there a check that the independent variables were manipulated as described?

- If an intervention study, did a sufficient sample of therapists or change agents implemented the intervention (i.e., to enhance generalizability)?
- If more than one treatment or condition are being compared, did the authors document that these conditions differ in ways they are supposed to differ? Are
  they the same in every other way (e.g., length, qualifications of therapists or change agents)? If not, is this confound likely to influence the conclusions seriously?
- What aspects of the procedures and independent variables limit the external validity of the study?

Results

- Do the data fulfill the assumptions and requirements of the statistics (e.g., homogeneity of variance for repeated-measures analyses of variance)?
- Were tests of significance used and reported appropriately (i.e., with sufficient detail to understand what analysis was being conducted)?
- In correlational studies, did the authors interpret low but significant correlations as though they indicated a great deal of shared variance between the measures?
- Are the correlations limited by restricted ranges on one or more measures? Do the authors provide means and standard deviations so that you can determine this?
- If there were a large number of statistical tests performed, do the authors adjust the alpha level or use appropriate multivariate techniques to reduce the probability of Type I error that could be due to the large number of tests performed?

- Do the authors report means and standard deviations (if relevant) so that the reader can examine whether statistically significant differences are large enough to be meaningful
- For multivariate statistics, is there an appropriately large ratio of participants to variables (at least seven for every dependent variable used in an analysis)?

Discussion

- Do the authors discuss marginally significant or nonsignificant results as though they were significant?
- Do the authors over interpret the data (e.g., use causal language to integrate correlational findings or interpret self-report of behavior as equivalent to direct observation)?
- Do the authors consider alternative explanations for the findings?
- Do the authors have a "humility" section that mentions the limitations of the research (including methodological problems)? Do the authors point out aspects of subject selection, procedures, and dependent variables that limit the generalizability of the findings?
- Do the authors "accept" the null hypothesis?

For more information on and examples of evaluating studies and thinking critically about research, I highly recommend consulting Meltzoff, J. (1998). *Critical thinking about research: Psychology and related fields*. Washington, DC: American Psychological Association.

This handout contains the guidelines presented by Maher, B. A. (1978), and cited in Cone, J. and Foster, S. (2008). *Dissertations and thesis from start-to-finish: Psychology and related fields (2nd. ed.).* APA: Washington DC.

References:

Cone, J. and Foster, S. (2008). *Dissertations and thesis from start-to-finish: Psychology and related fields (2nd. ed.).* APA: Washington DC.

Meltzoff, J. (1998). *Critical thinking about research: Psychology and related fields*. Washington, DC: American Psychological Association.

Developed by Hidy Basta, Ph.D.
Director of the Center for Teaching and Learning, Antioch Seattle