**Overview**

*Purpose of critiquing & difference b/w summary & critique*

- Critique is EVALUATIVE in some way; critiquing research involves taking a skeptical point of view.
- 2 purposes in evaluating research:
  - Assessing the quality of the content
  - Commenting on the strengths/weaknesses of the study
- No research article is perfect. Need to be able to identify flaws when reading/critiquing articles, but focus on the really important flaws.
  - To be able to figure out which flaws are consequential, need to understand paper’s main purpose, conceptual questions paper trying to answer, methods used, clarity of conclusions, extent that conclusions follow from findings.
  - Conceptual problems are the most “fatal” (i.e. RQs or theoretical assumptions)
- Handout has list of questions we ask ourselves, section by section, to evaluate a research article. They don’t 100% match up with the questions in these lecture notes, but they both get at the same general idea.

*Types of Published articles*

- Empirical (what we are most concerned about in this class)
  - Reports of original research. Typically consist of distinct sections that reflect stages of research process and appear in the sequence of the stages: intro, method, results, discussion. (Will discuss these sections in more detail in a minute).
- Review (including meta-analyses)
  - Critical evaluation of material that has already been published. By organizing, integrating, and evaluating previously published material, the author of a review article considers the progress of current research toward clarifying a problem. Author:
    - Defines and clarifies the problem
    - Summarizes previous investigations in order to inform the reader of the state of current research
    - Identifies relations, contradictions, gaps, and inconsistencies in the literature
    - Suggests next step or steps in solving the problem
  - A meta-analysis basically statistically combines the results of lots of studies and reports on what kind of consensus we can come to in looking across studies.
- Theoretical (an be structured similarly to review articles)
  - Author draws on existing research literature to advance theory in any area of psychology. Presents empirical information only when it affects theoretical issues. Author often traces development of theory to expand and define theoretical constructs.
  - May present a new theory or analyze existing theory, pointing out flaws or demonstrating superiority of one theory over another.

*Hourglass form of empirical articles: 4 sections*

- Intro has 3 major sections that go from very general to very specific (top of hourglass):
  - Introduction to the problem
  - Development of the framework for the study
  - Statement of research questions/hypotheses
CRITIQUING RESEARCH ARTICLES

- Method describes how the hypotheses were tested, including how all aspects of the study were conducted. Need enough detail to determine validity of study and replicate it. Usually has 4 sections:
  - Participants
  - Measures/Variables/Instruments/Materials
  - Design
  - Procedures
- Results summarizes the data and results of statistical analyses
- Discussion discusses the results and
  - Explains whether or not the data supported the hypotheses
  - States the conclusions drawn by the authors, often relating them back to previous research
  - Indication of study’s limitations
  - Discusses implications of the study for counseling and research (This is where the hourglass widens again.)

Introductions

Purpose of Introductions
- To convey overall rationale and objectives of the study in a concise fashion.

Structure
- Structure goes from very general (lit review) to the specific. [Top part of the hourglass structure.]
- Begins by conveying what the area or topic of study is and why it is important.
- Then moves into issues that underlie this particular study by reviewing relevant literature.
  - Reviewing the literature and theory related to the area under investigation in an integrative (i.e. not study-by-study) way that sets the stage for the study.
  - It places the study in its context of what we know and what we don’t know about this area. (Note: Simply saying that the area has not been studied before is not the same as contextualizing and justifying the need for the study!)
  - It should very clearly show why this particular study is needed and the current questions, inconsistencies, or deficiencies in the literature that this study is designed to address.
  - Must convey that the limitations of prior work are central to further theory or knowledge about this topic OR must convey that the study will extend theory and constructs to a broader range of domains of performance, samples, and/or settings.
- The final paragraph(s) should include a statement of the purpose of the study and the specific hypothesis and/or research questions under study by the authors.
  - By this point, you should be convinced that what the author is studying and the predictions being made (or questions being asked) make perfect sense, are important, and address a critical issue or need in the knowledge base.
  - The RQ and/or hypotheses that the authors are studying should be testable, and variables should be operationalized.
Critiquing Introductions
Recall that our concerns in this section are:

• Objectives of the study
  o  *Ask yourself:* What is the topic of interest to the researchers? Why do they think it’s important? Have they convinced you that it is important? What are the goals? Hypotheses? Is everything clearly presented?

• Importance of research questions
  o  *Ask yourself:* Are you convinced? How does it relate to previous research? How do the authors address what’s missing or inconsistent in the literature? Are the hypotheses/RQs testable? Have the variables being studied been operationalized?

• Research idea at the conceptual level
  o  *Ask yourself:* Will this study change or add to our knowledge of human behavior? Lead others to do research in this area? Did the authors do a good job building their rationale for this study?

Method

• Is a “cookbook” or set of direction for conducting a study. Important to focus on this in a critique.

Research Designs
Research designs are a plan or structure for an investigation. Research designs are strategies that help researchers examine variables and relationships in a valid and systematic way. Each type of research design tries to isolate constructs of interest and then draw conclusions about the relationships among those constructs while trying to reduce bias and distortion (i.e., error).

• Research design involves developing a plan or structure for a study in a way that reduces as much bias/distortion/error as possible. Different research designs have different strengths and weaknesses and each one will minimize a different type of bias. (We’ll be talking about this in a lot more detail as the semester progresses.)

Research designs in counseling can be thought to exist along two dimensions: control and setting.

• **Control:** the degree to which the researcher manipulates the independent variables (variable of interest). Control can be high or low. In high control studies you can make statements about causality. In low control studies, you can make inferences about relationships, but not causality.

• **Setting:** where study takes place. Can be laboratory or field. Labs are usually simulations of reality; field is natural setting. The setting dimension, then, refers to how well the researcher can generalize the findings of the study to real life settings.

Therefore, in research designs there is always a trade-off between experimental control and generalizability.

• Experimental control allows you to say more about your constructs of interest. The more control you have of variables or conditions, the are more likely you are to be able to attribute your findings to the IV (make statement of causality).
CRITIQUING RESEARCH ARTICLES

• **Generalizability** allows you to apply your findings more readily to the real world. It refers to how “real to life” the study is.
• There is not one best type of research design. All have flaws.

Moreover, control and generalizability refers to questions of the validity of a study.

Very generally, validity asks how sure you are that you measuring what you think you’re measuring. That is,

1. Is there a relationship between the independent and dependent variables?
2. If so, is the relationship a causal one?
3. Can the relationship between the IV and DV be generalized to other people or settings?

• **Internal validity question**: Can you attribute your findings to the IV (variable of interest). Can you conclude that observed variability in your dependent variable was caused by variability in your independent variable? Can you say that the change in the DV was caused by the IV? This is referred to as experimental control. Threats to internal validity include:
  o The timing of when the IV was presented and the DV was measured (best is to measure DV, implement manipulation of IV, measure DV again)
  o Group differences that might have existed before the IV was manipulated (often control for this through random assignment).
  o Outside events that may have affected one or more participants/groups during the time the study is being conducted (often control for this by having multiple groups, and observing them at the same times and/or keeping the treatment short)
  o Normal developmental changes/improvement between pre-test and post-test
  o Regression to the mean (people who have extreme scores at one point in time tend to have less extreme scores the next time they are tested; this is especially a problem when participants are selected because of their extreme standing on some variable)
  o Attrition (people dropping out of the study; most problematic when it is not the same across groups)
  o People getting used to/gaining familiarity with the test being taken (relevant when people are taking the same test more than once)
  o Changes in the measuring devices or procedure over the course of a study (mostly applies to assessments that don’t involve tests, such as observation, interviews, coding, etc.)

• **External validity question**: Can you generalize your findings to the population or topic of interest? This is referred to as generalizability. Threats to external validity include:
  o Lack of diversity in the participant pool (can results be generalized to people/groups who don’t fit the demographics, experience, degree of dysfunction, cognitive level, etc. of the participants?)
  o Rigidity of the treatment (will the treatment have the same effects in the real world if the very strict protocols of the procedure aren’t followed?)
  o Limitation of results to the setting in which the study was conducted (would the same results happen in other settings or was there something specific about the research setting that might be responsible for the results?)
  o How well the outcomes measured in the study apply to other, related, outcomes (e.g., if behavioral changes are involved, will they generalize to larger personality changes?)
**Bubble Hypothesis (Gelso, 1979)**

- If you have a bubble on a bumper sticker and try to eliminate it, the bubble will simply pop up elsewhere.
- Illustrates idea that all research strategies have strengths and weaknesses. When you select the best strategy for answering your question, you are still making a trade-off between rigor (control/internal validity) and relevance (generalizability/external validity)
  - So, we don’t want to discount any study that has a “bubble” (because they all have them). Rather, we want to be on the lookout for “bubbles” that are problematic in terms of what types of conclusions can be made from a study.
- The goal in research is to develop a program of research where you look for “convergence of findings under methodological diversity.” That means that you try to address your questions using different types of research strategies and methods.
- Why might this be important? Because if you find similar results, you can be more certain that your findings are due to the variables of interest and not due to the similarity of methods or research designs used to examine your question.

Now let’s look at 4 types of research designs. Within each type, I will describe its central features and we’ll look at it with respect to internal and external validity. DO NOT stress about these distinctions right now… I’m just going over these types to highlight the trade-offs between internal and external validity. We will spend the second half of the semester talking about the pros and cons of different research designs. [NOTE: This stuff is all out of Ch. 4 in the textbook, so be sure to read that.]

<table>
<thead>
<tr>
<th>High External Validity</th>
<th>High Internal Validity</th>
<th>Low External Validity</th>
<th>Low Internal Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental Field</td>
<td></td>
<td>Descriptive Field</td>
<td></td>
</tr>
<tr>
<td>Experimental Laboratory</td>
<td></td>
<td>Descriptive Laboratory</td>
<td></td>
</tr>
</tbody>
</table>

**1. Experimental Lab**
- Lab setting with full control. So very high internal validity: different conditions, random assignment, manipulation of variables
- Control allows you to attribute your findings to the treatment, and not to other variables
- Can make causal inferences, and can isolate and study variables in a way that’s not possible in actual counseling settings
- Very low external validity: an artificial situation
- Analogue is a created situation, which mirrors real life, but is NOT a real life situation
- Example: studies of counseling interventions or techniques in a controlled environment, where therapist/client pairs are randomly assigned to groups

**2. Descriptive Lab**
- Also a simulation of real life because it occurs in a lab setting. But no variables are manipulated and there is no random assignment to conditions. So, the study is ultimately correlational in design.
- Less control, and thus less internal validity. Cannot make causal inferences.
- Lab, so little external validity.
- Good for describing in detail some aspect of counseling, by identifying labeling, or categorizing data as well as obtaining basic descriptive statistics about variables. Lab setting
allows some control over extraneous variables (e.g., that participants follow instructions) and makes it easier to collect data. Useful for things that would be impossible to study in real-life settings for practical or ethical reasons.

• Example: What are the precursors and effects of spontaneous use of immediacy in a counseling relationship? [Describe Immediacy study.]

3. Experimental Field

• Manipulation and random assignment, high control, thus moderately high internal validity (there’s only so much we can control)
• Done in field, or real life situation, so moderately high external validity too. (Can never really have full external validity b/c measuring things can change them.)
• Usually used to examine the effects of certain interventions—so studies the outcome of counseling
• Ethical issue: usually need control group to effectively compare. (otherwise, for example, can’t say that CBT is better than dynamic) What’s problem with that?
• Very difficult to do: in real life hard to randomly assign, get enough clients, follow up with clients you did get, receive permission from facility.
• Also, examination is very global; can’t examine specific statements or techniques
• Example: ED therapy at an ED clinic: randomly assign clients to treatment type. Measure client satisfaction with therapy, or measure change.

4. Descriptive Field

• Low internal and high external validity.
• Looks at relationships among variables as they naturally occur. Allows for study of many variables at once. Nonmanipulative and occurs in the field.
• Most commonly used (about half of published articles).
• Example: Nutt-Williams & Hill (1996) study; a study where people from all over the country fill out questionnaires about a couple different constructs (like the census)

Finally, research designs can be categorized as qualitative or quantitative

• The goal of qualitative research is to obtain a holistic description of phenomena. It utilizes a naturalistic approach and emphasizes understanding and description. The investigator’s perspective is an important element.
• The goal of quantitative research is to obtain numerical data about variables. It uses statistical procedures and emphasizes prediction, generalizability, and causality. The investigator’s perspective is minimized in the analysis and interpretation of the data.

*Measures*

Ideally, authors should report data on these things in the “Measures” section of the method. If they don’t, you don’t know how good the measures are that they are using.

*Reliability:* refers to the consistency of test scores; the extent to which the instrument measures the variable of interest without being affected by measurement error

• Internal consistency: refers to the degree of inter-item consistency, or how well the items hang together. To obtain, administer the test to one group of examinees and the correlate the items. Want some degree of correlation since items are theorized to measure a construct.
Too high of a correlation might suggest that the items measure the same aspect of the construct.

- **Test-retest**: involves administering the same measure to the same group of examinees at two different times, and then correlating the two sets of scores

**Construct Validity**: refers to the accuracy of the instrument; when it measures what it is intended to measure. Two sub-types of construct validity for measures:

- **Convergent**: correlate your measure with another instrument that is designed to assess a similar construct. Want a moderate positive correlation
- **Discriminant**: correlate your measure with another measure that is designed to assess an unrelated construct. Want a moderate negative correlation.

**Threats to construct validity include (look for these):**

- Variables not being operationalized well.
- Confounding of variables of interest with other variables
- Using only one method to measure the construct (e.g., only self-report; better to use self-report plus observations)
- Self-report problems (participants don’t report accurately b/c they want more attention, want to seem better, want to tell the researcher what s/he wants to hear)
- Participants guessing the hypothesis of the study and then complying with or rebelling against the outcome. (Placebo effects fit here too.)
- Experimenter expectancies
- Group issues (the “control” group gets extra attention/help somewhere that isn’t accounted for in the study, groups try to outperform or underperform in reaction to the condition they are in, the groups aren’t separate enough so that there is “treatment diffusion”)

**Participants**

The Participants section should discuss:

- **Who are the participants?** How were they selected? (Did they select themselves?) Who was not selected? Note that who was selected and how they were selected directly relate to the generalizability of a study (e.g., a study using only novice therapists as participants might not be generalizable to experienced therapists).
- **Participant information provided in the article should include total number, sex, race/ethnicity, age, sexual orientation, educational level.**
- **Return rate.** When conducting a survey, the return rate refers to the percentage of people who returned the survey. We want our return rate to be as high as possible. There are implications for a low return rate in terms of both internal and external validity for the study.

**Ethical Treatment**

The Ethics code describes treatment of research participants. You will likely not read about these issues in the Method section, but you want to consider the ethical implications of a study as you read about it. (And we’ll talk more explicitly about this next week.) Some concepts are described below.

**Standard 8: Research and publication**

- **Informed consent**
  - Description of study and nature of participation
CRITIQUING RESEARCH ARTICLES

- Agreement that participation is voluntary
- Free to withdraw consent at any time without penalty
- Any limits to confidentiality
- Description of risk and benefit
- Obtain permission to record voices
- If deception exists, how it is handled
- Institutional review board approvals.

Procedure
This section should provide a step-by-step description of how the research was conducted. What exactly happened in the procedure? Can you replicate the study by reading the procedure section?

Critiquing Method Sections
- Is the research design a fair test of the hypothesis or question?
- What are internal and external validity issues in this study?
- How do the internal and external validity issues relate to the concepts of control and generalizability?
- What are the threats to construct validity? Do the authors present adequate reliability and validity evidence for the measures?
- How were participants selected? Who was excluded? Impact?
- Were participants treated fairly? What ethical considerations might exist?
- Could the participants have guessed the purpose of the study, and if so, could this have affected their responses?
- What are return rate issues?
- Can you replicate the procedure?

Results
- Purpose is to summarize the data and the results of the statistical analyses. This is less of a focus in a critique than other sections, but it is still important to have a basic understanding of what they are all about. See the “What You (Minimally) Need To Know About Statistics To Understand Research Articles” handout for details.

- Generally 2 types of results reported:
  - Summary statistics & preliminary analyses
    - Usually includes means & standard deviations of DV for each level of the IV and correlation matrix for all variables.
    - By knowing s.d.’s, you can interpret findings in light of the amount of variability in the variables they studied. For example, if it seems strange that 2 variables were NOT significantly correlated, you might look at the s.d.’s of the variables and find that one or both of the variables had very little variability in the sample. The low variability might be the reason for the low correlation, as opposed to there actually being no relationship between the variables.
CRITIQUING RESEARCH ARTICLES

- Results may be related to manipulation checks, dropouts, verification of stimulus materials, etc.
  - Results related to research hypotheses
    - These are typically inferential statistics.
    - Should be organized so reader can tie specific statistical tests to research hypotheses/questions laid out in introduction
    - Should be one-to-one correspondence between hypotheses & statistical tests.

When reporting findings...
- Should report findings (i.e. XYZ was sig, supporting hypothesis B) but not discuss the results, as that goes in the discussion section.
- Purpose of each analysis should be clear.
- Often presented in tables or figures, which is useful b/c results can be understood without taking up a whole lot of space. Plus, this visual presentation of the results can often help readers understand them.
  - When tables/figures used, authors should not repeat them all in the text; but MUST refer to the appropriate table/figure.
- Some major journals such as Journal of Counseling Psychology now strongly recommend/require researchers to report effect sizes.

Statistical Conclusion Validity
- A major concern in evaluating results sections.
- The degree to which the research has come to the correct conclusion about the relationship between the variables being studied.
  - Remember, that statistics are based in probability and that the more tests one does the greater the chances of finding significant results by chance (i.e. making a Type I error) b/c the chance of your results being sig by chance are alpha level x # of tests. Authors should use a smaller alpha level when doing these extra statistical tests. Make sure authors mention this somewhere! (The bottom line is that if the authors are doing lots of statistical tests, they should say they are doing something to control for that.)
    - Statistical hypotheses are stated in terms of what is true for a population (not a sample). When we do a study it may be possible that the sample is not representative of the population, and therefore the results may not hold true for the population.
    - Statistical results are based on chance. P< .05 means that the chances of incorrectly concluding that a true relationship exists between 2 variables is less than 5 in 100. If, however, our results are part of that 5%, we are incorrectly concluding that a relationship exists (Type I error). Type I errors are bad b/c they incorrectly suggest that something is going on when it’s not. Alternatively, we can conclude that there is no relationship, when one actually exists. This is a type II error, which often results from lack of variability in the participant’s responses.
  - Conditions that create error lead to threats of statistical conclusion validity (i.e., things to look for):
    - Low statistical power (usually related to small sample sizes; more participants generally = more power)
    - Doing lots of statistical tests without controlling for error-rate problems (e.g., not using a lower alpha)
CRITIQUING RESEARCH ARTICLES

- Unreliable measures
- Restricted range of a variable (e.g., usually occurs because the instrument measuring the variable is not sensitive to the construct being measured at the upper limit or at the lower limit)
- Unreliability of treatment implementation (i.e. Did their manipulation work? Did their groups do what they were supposed to do?)
- Random irrelevancies in the experimental setting (i.e. any aspect of the experimental setting that leads to variability in responding, e.g. noise, interruptions, etc.)
- Random heterogeneity of respondents (participant characteristics the authors didn’t control for)

Critiquing Results Sections
- Could you understand what was done in the analyses?
- Did the authors control for # of tests? What were the threats to statistical conclusion validity?
- Did the authors present the results clearly (both in the text and visually via tables or graphs)?
- Did the authors report effect size along with their results? What is the practical significance of the authors’ results?

Discussion

Includes
- An explanation of whether or not the data supported the research hypotheses
  - Usually how the discussion section starts.
- A statement of the conclusions based on the results
  - Unravel major themes of the research; explains the findings
  - Authors should ensure that their conclusions do not go beyond what is supported by the results. They should consider what alternative explanations might exist for the results that they found.
  - Places findings in context of previous research and theory on this topic
- An indication of the study’s strengths & limitations
  - All studies have limitations—remember Bubble Hypotheses!
  - Should discuss how results are interpretable despite the limitations.
- A discussion of the implications for theory, research, practice
  - Often includes suggestions for future research
- The authors should bring everything together in the discussion section and it should be easy for you, the reader, to grasp how it all comes together. If you’re struggling to bring it all together as you read, then they haven’t done a good job writing it up!

Critiquing Discussion Sections
- Did the authors do a good job presenting the conclusions made from their results?
  - Did the authors interpret beyond their results?
  - Are there alternative explanations for the authors’ findings? If so, what are they?
- Did the authors do a good job putting the results of their study back into the context of previous research?
CRITIQUING RESEARCH ARTICLES

• Did the authors spell out the implications of their results for theory, research and/or practice?
• Did the authors discuss the limitations of their results adequately?
• Did the authors provide feasible ideas for future research directions?

References