

How to Critically Analyse Psychological Research

Table of Contents

The Theory.....	2
The Research Rationale	2
The Participants.....	2
The Design and Procedure	2
1. Research method.....	2
2. Lab vs field research.....	2
3. Demand characteristics	3
4. Experimenter bias.....	3
6. Social desirability	3
7. Validity of the experimental manipulation.....	3
8. Stimulus sampling.....	4
9. Reliability and validity of measures of the independent and/or dependent variables	4
10. Confounding variables in	4
11. Order of items/events.....	4
The Statistical Analyses	5
1. Excluded participants.....	5
2. Missing data.....	5
3. Validity and reliability of dependent variables	5
4. Sufficient statistical power.....	5
5. Statistical assumptions	6
6. Correct use of inferential statistics	6
7. Correct interpretation of analyses	6
8. Alternative analyses.....	6
The Discussion	6
1. Alternative explanations.....	6
2. Cause-effect ambiguities	6
3. Third variable	7
4. Mediators and moderators	7
5. Replication	7
6. Interaction or main effect?:	7
Place the Research in the Context of Similar Research	8
Suggestions for Future Research	8
Inappropriate Criticisms	8
1. Criticizing the article rather than the research.....	8
2. Ethical criticisms	8
3. Incomplete criticisms.....	8
4. Criticisms of the reliability or effectiveness of methodology that produced the predicted results.....	9
5. Random allocation of participants to conditions	9
How Not to Use this Document!	10
Structuring a Critical Review	10
Useful Websites	11
Background Reading	11
References.....	12

The Theory

You may wish to criticize the theory that the researchers are testing. How does it compare against competing theories in the area? What are its strengths and weaknesses? An excellent resource for thinking about theory construction is the Special Issue on “Theory construction in social personality psychology: Personal experiences and lesson learned” in the *Personality and Social Psychology Review* (Vol. 8).

The Research Rationale

Is there a fault in the logic of the theoretical rationale for the research? Have the researchers interpreted the theory that they are basing their hypotheses on correctly? Do the hypotheses follow logically from the theory? Does the research design provide a satisfactory test of the research hypotheses? Are all of the necessary experimental and control conditions included? Are all of the necessary variables measured? See McGuire (2004) for a good discussion on these points.

The Participants

About 70% percent of psychology research is conducted using young, educated, white, middle-class, Western, volunteer, psychology undergraduate students (Sherman et al., 1999; Wintre, North, & Sugar, 2001). Hence, it is possible that 70% of psychology research cannot be generalized to the rest of the world’s population. However, you should consider two points before making reference to this sample generalisation problem. First, the sample generalisation problem is unlikely to threaten the external validity of research investigating basic cognitive and perceptual processes such as vision, because there is no reason to believe that psychology undergraduates see differently to other types of people (Stanovich, 2007, p. 112). Second, the sample generalisation problem is widely recognised among psychologists (Stanovich, 2007, p. 117), and it does not need to be stated explicitly unless the characteristics of the sample pose a particular problem in relation to the specific independent and dependent variables being investigated.

The Design and Procedure

1. Research method: Every research method has its advantages and its disadvantages. Did the researchers choose the most appropriate research method for the particular research question that they were investigating? Did they deal with the disadvantages of that method? If not, how do you think that those disadvantages may have affected the results? For example, did the researchers conduct their research on the internet, and if so did they address the limitations of this particular methodology (Birnbbaum, 2004; Skitka & Sargis, 2006; Van Selm & Jankowski, 2006)?
2. Lab vs field research: Was the research conducted under artificial conditions in the laboratory or was it conducted under more naturalistic conditions in a real world setting? Lab research has the advantage of providing more control over extraneous variables. This means that it is often easier to draw firmer conclusions about the results from lab research than from field research. However, field research is often more naturalistic and realistic and can generate less suspicions in participants. Be careful to make your criticisms specific to the particular research that you are looking at. Don’t just say that the researchers used lab research and so the results may not be generalizable to real world situations. Instead, specify which results may not be generalizable to which real world situations and explain why you think they may not be generalizable (i.e., what is different about the real world situation in comparison with the lab situation).

3. Demand characteristics: Demand characteristics are “the totality of cues which convey an experimental hypothesis to the subject” (Orne, 1962, p. 779; see also the Special Issue in *Prevention and Treatment*, 2002; Strohmets, 2008). The most common sources of demand characteristics are the research setting, the implicit and explicit research instructions, and the research procedure. Demand characteristics are a problem because, if participants are able to deduce the research hypotheses, then they may respond in a manner that they think will confirm the hypothesis in order to be a “good” participant and not “ruin” the research (e.g., Norenzayan & Schwarz, 1999). What demand characteristics do you think existed in the research? As Strohmets (2008) noted, the impact of demand characteristics depends on participant’s receptivity to these characteristics and their motivation and ability to comply with them. Do you think that participants were able to guess the research hypotheses from these demand characteristics? Do you think that participants were motivated to try to confirm these hypotheses? How do you think that their attempts to confirm the hypotheses will have affected the results? Was any deception and/or concealment used in the research in order to prevent demand characteristics from having an effect, and if so, how effective do you think that this deception/concealment was?
4. Experimenter bias: The experimenter’s nonverbal behaviour may give away clues about how the participant is expected to respond (Rosenthal & Rosnow, 1969). As per demand characteristics (see above), this nonverbal behaviour may then influence participants’ responses and produce results that are caused by artificial factors that depend on the participants’ knowledge that they are taking part in an experiment, rather than by genuine psychological processes that can be generalized outside of the experimental context. Usually, experimenter bias can be avoided if the experimenter is unaware of the research hypotheses or if his/her nonverbal behaviour is unable or unlikely to influence the participants’ responses. Was the experimenter blind to the experimental conditions? If not, was there any way that his/her nonverbal behaviour could have systematically influenced participants’ responses?
5. Reactivity: Sometimes the act of measuring a thought or behaviour can change that thought or behaviour (for an overview, see French & Sutton, 2011). For example, the act of measuring the same attitude or behaviour at different times during a research study may lead participants to assume that the researchers predict the attitude or behaviour to change from one measurement to the next. If researchers make multiple measurements of the same attitude or behaviour, have they addressed the potential reactivity of this procedure?
6. Social desirability: People want to present themselves in a good light when they take part in research (Crowne & Marlowe, 1964; Paulhus & Reid, 1991). They don’t want to be seen as “bad” or “wrong”. To avoid these labels, participants will often downplay their socially undesirable attitudes or behaviours. So, for example, participants may describe themselves as being less aggressive than they actually believe that they are in order to present themselves in a more positive light. Was the research likely to have been influenced by participants’ desire to appear socially desirable? If so, how might this motivation have affected the pattern of results that the researchers found? Researchers can use safeguards against socially desirable responding such as allowing participants to make anonymous responses or measuring individual differences in social desirability and then controlling for this variable in their statistical analyses. Were any of these safeguards in place and, if so, how effective do you think that they were?
7. Validity of the experimental manipulation: Did the experimental manipulation alter the independent variable as predicted? Did it alter any other variable as well as the independent variable? For example, researchers might attempt to manipulate self-esteem by asking their participants to watch either a happy or sad video. But this procedure may manipulate mood

instead of self-esteem. Hence, the experimental manipulation is invalid because it is not manipulating self-esteem, but mood instead. Sometimes, researchers may include a *manipulation check* in their research. This is a measure that is intended to show that the experimental manipulation has had a significant effect in manipulating the correct variable. Was a manipulation check included? If a manipulation check was included, did it show that the manipulation was effective? In other words, did it indicate significant differences in the independent variable between relevant experimental conditions? Note that, even if the manipulation check is successful, it remains possible that the experimental manipulation manipulated more than just the independent variable (e.g., self-esteem) and that an additional, confounding variable (e.g., mood) was actually the one that was responsible for the significant effects that were observed.

8. Stimulus sampling: A related issue is that of stimulus sampling (Wells & Windschitl, 1999). Were the observed effects due to the independent variable or the particular stimuli that were used to represent the independent variable? For example, suppose a researcher tests the hypothesis that women like children more than men do. To test this hypothesis, the researcher presents male and female participants with a single picture of a child and asks them to rate how much they like that child. In this case, any gender effects may be more to do with the specific picture of the child that the researcher has chosen to represent the general category of "children" (e.g., perhaps the child's own gender is having an effect). In order to ensure the content validity of this variable, the researcher should sample a variety of different pictures of children (stimuli) with different gender, age, appearance, etc. in order to rule out these potentially confounding variables from the research.
9. Reliability and validity of measures of the independent and/or dependent variables: Have measures of the independent and dependent variables been shown to be a reliable in the present research and in previous research (e.g., test-retest reliability, internal reliability)? Has they been shown to be a valid measures in the present research and in previous research (e.g., face validity, content validity, criterion validity)? Have the psychometric scales been developed in an appropriate manner (e.g., Clark & Watson, 1995; Haynes, Richard, & Kubany, 1995).
10. Confounding variables in measures of the independent and/or dependent variables: Did the measures of the independent and/or dependent variables assess one or more additional variables to the one that the researchers were interested in? For example, the items in a scale measuring aggressive behaviour might also tap self-esteem to some extent. In this case, perhaps the significant effects that the researchers found represent differences in self-esteem rather than differences in aggression. Researchers can attempt to control for variation due to self-esteem by including a self-esteem scale in their research and using this as a covariate in their statistical analyses.
11. Order of items/events: The order in which researchers present items or events to participants can make a big difference to the way in which participants interpret those items or events (e.g., Bless, Strack, & Schwarz, 1993; Hilton, 1995; Schwarz, 1999). Participants will attempt to build up a picture of what the research is about from the questions they are being asked and tasks that they have to complete. Try to put yourself in the participants' position at each stage of the procedure. If you were a participant, how would you interpret the experiment based on the order of things that you are asked to do? Is your interpretation consistent with the researchers' assumptions? Another aspect of order is practice and fatigue effects. Participants who are asked to do the same sort of thing again and again may get better at it through practice effects. Do practice effects account for the research results? Alternatively, participants may get tired of completing hundreds of items and we might find significant effects for scales placed at the beginning of the research, but nonsignificant

effects for scales placed at the end of the research. The researchers might claim that these different effects are due to the content of the scales. You might argue that the different effects are simply because participants aren't really attending to the items in the last scale (i.e., a fatigue effect). One way researchers can deal with these sorts of order effects is to counterbalance the order in which the present things. Do they need to do this in the research you are looking at?

The Statistical Analyses

1. Excluded participants: Were any participants excluded from the analyses and if so why? Did the researchers justify any exclusions appropriately? For a good discussion on the reasons to exclude outliers, see Osborne and Overbay (2004).
2. Missing data: If participants leave questions or items blank, we end up with what we call missing data. There are various different methods of dealing with missing data (Schafer & Graham, 2002). Did the researchers choose the most appropriate method?
3. Validity and reliability of dependent variables: Did the researchers provide convincing evidence for the validity of each of the dependent variables that they used (including psychometric scales)? In other words, did each dependent variable show significant and appropriately sized correlations with the variables that it was supposed to be related to (*convergent validity*) and, equally importantly, weak nonsignificant relationships with the variables that it was not supposed to be related to (*discriminant validity*)? Also, was there good evidence of the internal reliability of the dependent variables? For example, did each psychometric scale have a suitable factor structure and/or acceptable Cronbach alpha coefficients ($> .70$)?
4. Sufficient statistical power: If researchers find a significant effect, then, ipso facto, they must have had sufficient statistical power to detect this effect. Consequently, it would be inappropriate to criticise the researchers for have low statistical power due to small sample size even if the researchers' sample size is smaller than that used in previous research. However, if the researchers found null findings, then this can either be interpreted as indicating that there is no effect present or that an effect is present but the researchers had insufficient statistical power to detect this effect (i.e., a Type II error; see Cohen, 1988, 1992). Hence, statistical power is a critical concern when interpreting null findings. When interpreting a null finding, consider whether the research contained enough participants to detect the effect. Look back at previous research that has found the effect in order to see how many participants were used in that research. Meta-analyses and other reviews are good sources for this information. Does the research use significantly fewer participants than previous successful research? If so, then the null findings may be due to a lack of statistical power. Faul, Erdfelder, Lang, and Buchner (2007) provide a free downloadable power analysis software that you can use to investigate whether researchers have sufficient power. It is available at: <http://www.psych.uni-duesseldorf.de/abteilungen/aap/gpower3/download-and-register> In addition, Maxwell (2004) provides some useful calculations regarding recommended sample sizes. Assume that researchers want to conduct a statistical test with Cohen's (1992) recommended power of .80 to detect a medium-sized effect using an alpha value of .05 and with equal numbers of participants in each condition. If the researchers are using a 2 x 2 between-subjects ANOVA and a single dependent variable, then, in order to detect a single, prespecified effect (e.g., a main effect), the researchers should use 30 participants in each of the four cells of the 2 x 2 design (i.e., 120 participants). In order to detect all three effects (i.e., both main effects and the interaction), the researchers should use 48 participants in each cell (i.e., 192 participants). Obviously, cell sizes will need to be

larger if (a) cell sizes are unequal, (b) the ANOVA is larger (e.g., 2 x 3 ANOVA), or (c) there is more than one dependent variable.

5. Statistical assumptions: Did the researchers meet all of the assumptions that are associated with the particular statistical tests that they used (e.g., equal cell sizes, normal distribution, homogeneity of variance).
6. Correct use of inferential statistics: All statistical techniques have their limitations. Did the researchers take these limitations into account. Have a look at some general introductions to the techniques of exploratory factor analysis (Floyd & Widaman, 1995; Russell, 2002), path analysis (Stage, Nora, & Carter, 2004), or structural equation modelling and confirmatory factor analysis (MacCallum & Austin, 2000; Schrieber, Stage, King, Nora, & Barlow, 2006) correctly? Was their dichotomization of quantitative variables appropriate (MacCallum, Zhang, Preacher, & Rucker, 2002; Maxwell & Delaney, 1993)?
7. Correct interpretation of analyses: Did the researchers interpret the results correctly? Look back at the precise predictions that the researchers made and match them against the actual pattern of results. Researchers are like politicians: They will try to place a positive spin on their results, emphasize supportive evidence, and downplay unresponsive evidence. As a critical analyst, it's your job to see through the rhetoric and spin and analyze the cold hard facts!
8. Alternative analyses: Different statistical tests can be used to address different questions. However, different statistical tests can also be used to address the *same* question. Did the researchers use the correct (i.e., most powerful, most precise) statistical test to investigate their hypotheses? Were there any alternative, more appropriate statistical analyses that could have been used to test the researchers' hypotheses?

The Discussion

1. Alternative explanations: Are any other explanations able to account for the results more parsimoniously? The authors will have attempted to rule out potential alternative explanations for their results in their paper. You may wish to highlight problems with the way in which they have dealt with these alternative explanations. For example, if the authors say that Problem X is not really a problem because of Solution Y, then you may wish to explain why Solution Y is not very effective at dealing with Problem X. Alternatively, you may propose new potential alternative explanations that the researchers did not consider in their paper. An important issue here is that of "loose ends": It is rare that researchers will find a pattern of results that perfectly fits their hypotheses. There will often be some loose end results that are not consistent with the hypotheses. These may either be null findings or significant results that contradict the researchers' predictions. The researchers will have attempted to explain away these loose ends in their article. Are their explanations satisfactory? Is there an alternative explanation that might provide a better account for the overall pattern of results, including the loose ends? Remember to be as precise, explicit, and specific as possible when discussing your own alternative explanations. Explain the processes involved.
2. Cause-effect ambiguities: Cause and effect is sometimes difficult to establish in correlational studies. The researchers may conclude that X causes Y because X is positively correlated with Y. But is it possible that the causal relationship is reversed and that Y causes X? For example, there might be a correlation between watching violent films and being an aggressive person. The researchers may conclude that violent films change people's

personalities to make them aggressive. But it is also possible that aggressive people deliberately seek out and watch violent films.

3. Third variable: The other problem with correlational designs is that a third, unspecified, variable may cause the correlation between X and Y. For example, a person's weight might correlate positively with their income, not because there is any relationship between these two variables, but because they are both correlated with a third variable: age. The older people get, the more weight they put on *and* the more income they earn (Baron & Bryne, 2003).
4. Mediators and moderators: Psychologists use the terms *mediator* and *moderator* in very particular ways (Baron & Kenny, 1986; Judd, Kenny, & McClelland, 2001; MacKinnon, Fairchild, & Fritz, 2007). If A causes B, and B causes C, then B can be said to *mediate* the effect of A on C. So, for example, seeing a lion (A) might cause you to run away (C). But being afraid of the lion (B) mediates this relationship: Seeing a lion causes you to be afraid, and it is this fear that causes you to run away (you wouldn't run away from the lion if you weren't afraid of it!). Hence, fear mediates the effect of seeing a lion on your behaviour. If A causes C, but only under B conditions, then B can be said to *moderate* the effect of A on C. So, for example, seeing a lion in the jungle might cause you to run away, but seeing a lion in a zoo might not cause you to run away. Here, the situational context (jungle vs zoo) moderates the effect of seeing a lion on your behaviour: Seeing a lion only causes you to run away when you are in jungle conditions, not when you are in zoo conditions. Mediators answer the question "*how* does the process operate?" Mediators account for the relationship between the independent variable and the dependent variable. Moderators answer the question "*when* does the process operate?" Moderators alter the direction or strength of the relationship between the independent variable and the dependent variable. Have the researchers attempted to test for mediators and/or moderators in their research? Spencer, Zanna, and Fong (2005, p. 848) pointed out several potential problems with mediational analyses. I list three key ones here: (1) Mediational analyses are essentially correlational analyses, and so are open to cause-effect and third variable interpretations (see above). (2) Mediational analyses often suffer from low power and so may yield Type II null mediation findings (i.e., null findings where, in fact, there is mediation present). (3) For mediational analyses to be theoretically meaningful, the researchers must use mediators that are theoretically distinct from the independent and dependent variables (Fiedler, 2011). Is there suitable evidence of discriminant validity between A, B, and C?
5. Replication: Have the researchers been able to replicate their effect? Finding a significant effect once at $p < .05$ means that there is a 1 in 20 chance that the effect represents a Type I error (i.e., reporting an effect to be significant when, in fact, it does not exist). However, finding the same effect twice on separate occasions at $p < .05$ means that there is a 1 in 400 chance that the effect represents a Type I error (Hays, 1994). So, replicating an effect can greatly increase our confidence in the reliability of that effect.
6. Interaction or main effect?: In some studies, independent variables can be defined relative to other independent variables. For example, in intergroup studies, the independent variable "in-group/out-group" may be defined relative to a participant's gender (male/female) and the type of target group that they are responding to (men/women). In this case, a two-way interaction involving the independent variable defined in relative terms is statistically equivalent to the main effect when the independent variable is defined in absolute terms. So, for example, the interaction between participants' gender (male/female) and target group (in-group/out-group) is statistically identical to the main effect of target group when target group is defined as "men/women" rather than as "in-group/out-group" (for further details, see

Brauer & Judd, 2001). Are the authors interpreting a two-way interaction when it is more appropriate to conceive the effect as a main effect?

Place the Research in the Context of Similar Research

What are the strengths and weakness of the present research compared with other similar studies in this area? The authors will have already addressed this point in their article. However, they may have missed something or their conclusions may be biased or incorrect. Does the research advance our understanding of the phenomena in the ways that the researchers claim? Does the research confirm or contradict previous findings? If it contradicts previous findings, is there a clear reason why?

Suggestions for Future Research

You may also be awarded marks if you make intelligent and specific suggestions for future research. So, based on your critical analysis of the research, what would your suggestions be for a more appropriate piece of research? Don't come up with 'half-baked' ideas: E.g., "Future research should look at X and Y". Follow your ideas through and be explicit: E.g., "Future research should look at X and Y. For example, future research should manipulate X using the ABC procedure and measure Y using the Blah-Blah scale. This will overcome the problems in the present research because the ABC procedure has such-and-such advantages and the Blah-Blah scale has been shown to be a more accurate measure of Y (Smith, 1982)". Make sure that you include your own predictions when discussing ideas for future research. For example, don't just say, "future research should investigate the relationship between A and B". Instead say, "future research should investigate the relationship between A and B. On the basis of the present research, I predict that A will be negatively correlated with B".

One obvious avenue for future research concerns the issue of generalization. If the effect was demonstrated under laboratory conditions, will it generalize to real world settings? Under what conditions do you think that the effect will get stronger or weaker and why? What personality variables might influence the effect and why? Will the effect generalize to other cultures?

Inappropriate Criticisms

1. Criticizing the article rather than the research: Your job is to criticize the research, not the paper reporting the research. In other words, you should critically evaluate the ideas and methods involved in the research, not the way in which these ideas and methods are presented in the paper. More specifically, you should not normally comment on (a) the clarity of the article (e.g., "the hypotheses followed in a logical manner from the research rationale", "the authors did not provide a clear discussion of the implications of their research"), (b) the writing style of the authors (e.g., "the article was too long", "the authors used too much terminology", "the authors did not conform to APA style"), or (c) any omissions in the paper (e.g., "the authors did not say how they dealt with missing data", "no information about participants' age range was provided"). You will not gain marks for making these types of criticisms. You should focus your comments on the research, not the research article.
2. Ethical criticisms: Unless you are specifically instructed to do so, it is usually not appropriate to comment on the ethical aspects of the research methodology that you are criticising. You should assume that the research has been approved by a human research ethics committee and that, therefore, ethical considerations have already been dealt with.
3. Incomplete criticisms: You need to be as explicit, specific, detailed, and comprehensive as possible when making your criticisms. In general, each critical idea that you put forward should contain: (a) a general introduction (e.g., "It is possible that social desirability

influenced the results”), (b) a specific elaboration of the criticism (e.g., “In other words, participants may not have been prejudiced because they perceived this form of behaviour to be socially undesirable”), (c) citations to theoretical and/or empirical work that supports your assertions (e.g., “Smith and Jones (1982) found that levels of prejudice were reduced when participants were aware that this form of behaviour was the subject of the researchers’ investigations”), (d) examples of your criticism that are taken from the research (e.g., “Participants were told that the current research was investigating ‘prejudice’”), (e) reference to any evidence in the target research that supports your claim (e.g., “Postexperimental feedback from participants did seem to show that they were concerned about the impression that their responses were making on others”), (f) a discussion of the implications of your criticism with respect to the research results and/or conclusions (e.g., “This problem may have reduced the level of prejudice that was found”), (g) suggestions for future research based on your criticisms (e.g., “Future research should attempt to conceal from participants the fact that prejudice is being measured”). You will get very few marks if you only include incomplete criticisms in your research. For example, you would not get many marks for simply saying “It is possible that social desirability influenced the results” and leave it at that!

4. Criticisms of the reliability or effectiveness of methodology that produced the predicted results: It is usually only appropriate to criticise the reliability or effectiveness of a study’s methodology when that methodology has failed to produce the predicted results. It is *inappropriate* to criticise the reliability or effectiveness of methodology when it has produced the predicted results. So, for example, researchers might use a self-esteem scale that previous research has found to be extremely unreliable. However, in their research, the researchers find that the self-esteem scale showed significant differences in the predicted directions. In this case, it would be inappropriate to criticise the self-esteem scale for being unreliable because the fact that it has revealed the predicted results means that it must have been reliable enough to do so. However, it would be appropriate to criticise the reliability of the researchers’ self-esteem scale if it produced unexpected null results. Note that, although you should not criticise the reliability or effectiveness of methodology when it has produced the predicted results, you may still criticise the *validity* of that methodology. This type of criticism may lead to a more general criticism of the conclusions that the researchers reached. For example, you might argue that a self-esteem scale was an invalid measure of self-esteem and that it really measured self-awareness. In this case, you would be able to challenge the researchers’ conclusions and argue that they should significant differences in self-awareness rather than self-esteem.
5. Random allocation of participants to conditions: One of the most problematic criticisms that students make concerns the random allocation of participants to conditions. In experimental studies, participants should be randomly assigned to experimental conditions. So, for example, imagine that some researchers manipulate participants’ self-esteem by giving them either positive or negative feedback about their performance on an intelligence test. They then measure participants’ aggressive behaviour in order to determine what effect differences in self-esteem have on levels of aggression. Further imagine that the researchers find that participants who received negative feedback showed significantly more aggressive behaviour than people who received positive feedback. The researchers might conclude that low self-esteem causes aggression. A student might attempt to criticize this conclusion by arguing that “if there happened to be a few extra aggressive people in the negative feedback condition, then this could also explain the result”. Admittedly, it is possible that a few extra aggressive people might have ended up in the negative feedback condition by pure chance alone. However, the statistical tests that the researchers used in order to determine whether or not there was a significant difference between the positive and negative conditions already takes this possibility into account. If the p value is less than .05, we know that there is a 1 in 20 chance that the student’s explanation (or some other

explanation) is correct. However, as scientists, we have agreed to conform to the convention of accepting this 1 in 20 risk as being low enough for us to effectively ignore it. As a critical analyst, you should also conform to this universal scientific convention and accept that, although it is possible that more aggressive people may have ended up in the negative feedback condition by chance alone, it is not acceptable to criticise the research on this basis because the chances of it having happened are relatively low given the statistical results. Note that this whole argument rests on the assumption that the researchers have randomly assigned participants to the positive and negative feedback conditions in their experiment. The random allocation of participants to conditions means that we can be relatively confident that the same types of people are equally represented within each condition. So, for example, as well as having the same proportion of aggressive and nonaggressive people in each condition, we will probably have the same proportion of men and women in each condition. This proportion may not necessarily be equal: There could only be 30% men in each condition. But this doesn't matter when it comes to interpreting differences between conditions. The crucial thing is that BOTH conditions contain 30% men and so gender cannot be used as an explanatory variable when considering any differences in aggression that are found between the two experimental conditions. Note that this argument applies to ALL personality-based variables (e.g., aggression, intelligence, conscientiousness, extraversion, etc.). So, the main point is that principle of random allocation means that you cannot use personality variables to explain differences between experimental conditions.

How Not to Use this Document!

The worst way in which you could use this document is as a pre-prepared template for criticising a piece of research. In other words, it is entirely inappropriate to simply make a list of headings as follows: (1) Lab vs field research, (2) Demand characteristics, (3) Experimenter bias, (4) Social desirability, (5) Validity of the experimental manipulation, etc. and then try to address each potential problem in the target research. Not all pieces of research will suffer from all of the potential problems listed in this document. Hence, you will not need to address all of the issues covered above for any one particular piece of research. Ideally, you should identify the most serious problems with the piece of research that you are evaluating and focus on those in your report.

Structuring a Critical Review

It is not good having a list of incisive criticisms if you don't present them in a well-structured manner. This is particularly important if you are conducting a critical review in order to build up a rationale for a research study that you aim to conduct.

To illustrate, imagine that there are three studies that are relevant to your methodological rationale: Blogs (2010), Jones (2011), and Smith (2012). Imagine that each study has a problem: Blogs' study has a small sample size, Jones's study has an invalid measure, and Smith's study suffers from demand characteristics. Further imagine that your study addresses all three of these issues: It has the right sample size, valid measures, and protection against demand characteristics. There are three ways of structuring a critical review of this literature in order to build up a methodological rationale for your own study:

- (1) Describe the work of Blogs (2010), Jones (2011), and Smith (2012), then describe the criticisms of these three studies, and then describe how your study addresses these critical issues in your own methodology. (The Goldilocks and the Three Bears approach to Critical Reviews!)

- (2) Describe the work of Blogs (2010) and the criticisms of Blogs. Describe the work of Jones (2011) and the criticisms of Jones. Describe the work of Smith (2012) and the criticisms of Smith. Then describe how your study addresses these three critical issues.
- (3) Describe the work of Blogs (2010), the criticisms of Blogs, and how you address these criticisms in your own study. Describe the work of Jones (2011), the criticisms of Jones, and how you address these criticisms in your own study. Describe the work of Smith (2012), the criticisms of Smith, and how you address these criticisms in your own study.

You need to consider which of these three approaches works best in the context of your own critical review. Possibly the least effective approach is the first approach. Effective criticisms often rely on background information about a study's methodology, and the reader may forget this information if you use the first approach. You could repeat some of this information when you get to the criticisms part, but this would be a relatively inefficient approach compared to the other two approaches. Hence, I recommend either the second or third approaches.

Useful Websites

The Critical Thinking Community: <http://www.criticalthinking.org/index.cfm>

How to Avoid Logical Fallacies in arguments:

<http://www.geocities.com/anatheist2001/subskepticismfallacies.htm>

Polson, D., Ng, C., Grant, L., & Mah, D. (1998). Athabasca University's Psychology 404 (Experimental Psychology) tutorial explaining nine sources of threat to internal validity:

<http://psych.athabascau.ca/html/Validity/index.shtml>

Jordan, C. H., & Zanna, M. P. (1999). How to read a journal article in social psychology. In R. F. Baumeister (Ed.), *The self in social psychology* (pp. 461-470). Philadelphia: Psychology Press. Retrieved on 31st October 2006 from University of Waterloo:

<http://arts.uwaterloo.ca/%7Esspencer/psych253/readart.html>

Background Reading

Aronson, E., & Carlsmith, J. M. (1985). Experimentation in social psychology. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (pp. 1- 79). New York: Random House.

Bem, D. J. (1995). Writing a review article for *Psychological Bulletin*. *Psychological Bulletin*, 118, 172-177.

Dunbar, G. (2005). *Evaluating research methods in psychology: A case study approach*. BPS Blackwell.

Halpern, D. F. (2003). *Thought and knowledge: An introduction to critical thinking*. New Jersey: Erlbaum.

Jordan, C. H., & Zanna, M. P. (1999). How to read a journal article in social psychology. In R. F. Baumeister (Ed.), *The self in social psychology* (pp. 461-470). Philadelphia: Psychology Press.

Leavitt, F. (2001). *Evaluating scientific research: Separating fact from fiction*. Upper Saddle River, NJ: Prentice Hall.

Miller, A. G. (1976). The social psychology of the research situation. In B. Seidenberg & A. Snadowsky (Eds.), *Social psychology: An introduction*. New York: Free Press.

Paul, R., & Elder, L. (2009). *The miniature guide to critical thinking: Concepts and tools*. The Foundation for Critical Thinking.

Vaughan, G. M., & Hogg, M. A. (2010). *Introduction to social psychology* (6th ed.). Upper Saddle River, NJ: Pearson Education. [Chapter 1]

References

- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.
- Birnbaum, M. H. (2004). Human research and data collection via the internet. *Annual Review of Psychology, 55*, 803-832.
- Bless, H., Strack, F., & Schwarz, N. (1993). The informative functions of research procedures: Bias and the logic of conversation. *European Journal of Social Psychology, 23*, 149-165.
- Brauer, M., & Judd, C. M. (2000). Defining variables in relationship to other variables: When interactions suddenly turn out to be main effects. *Journal of Experimental Social Psychology, 36*, 410-423.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*, 309-319.
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd Ed.). Hove and London: Lawrence Erlbaum Associates.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.
- Crowne, D. P., & Marlowe, D. *The approval motive*. New York: Wiley.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.
- Fielder, K. (2011). Voodoo correlations are everywhere – not only in neuroscience. *Perspectives on Psychological Science, 6*, 163-171.
- Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*, 286-299.
- French, D. O., & Sutton, S. (2011). Does measuring people change them? *The Psychologist, 24*, 272-274.
- Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*, 238-247.
- Hays, W. (1994). *Statistics* (5th ed.). Fort Worth, TX; Harcourt Brace.
- Hilton, D. J. (1995). The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin, 118*, 248-271.
- Judd, C. M., Kenny, D. A., & McClelland, G. H. (2001). Estimating and testing mediation and moderation in within-subject designs. *Psychological Methods, 6*, 115-134.
- MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modelling in psychological research. *Annual Review of Psychology, 51*, 201-226.
- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19-40.
- MacKinnon, D. P., Fairchild, A. J., Fritz, M. S. (2007). Mediation Analysis. *Annual Review of Psychology, 58*, 593-614.
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*, 83-104.
- Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. *Psychological Methods, 9*, 147-163.
- Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin, 113*, 181-190.
- McGuire, W. J. (2004). A perspectivist approach to theory construction. *Personality and Social Psychology Review, 8*, 173-182.
- Norenzayan, A., & Schwarz, N. (1999). Telling what they want to know: Participants tailor causal attributions to researchers' interests. *European Journal of Social Psychology, 29*, 1011-1020.

- Orne, M. (1962). On the social psychology of the psychology experiment: with particular reference to demand characteristics and their implications. *American Psychologist*, *17*, 776-783.
- Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research & Evaluation*, *9*. Retrieved 17th June 2009 from <http://PAREonline.net/getvn.asp?v=9&n=6>
- Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology*, *60*, 307-317.
- Rosenthal, R., & Rosnow, R. (Eds.) (1969). *Artifact in behavioural research*. New York: Academic Press.
- Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Personality and Social Psychology Bulletin*, *28*, 1629-1646.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147-177.
- Schreiber, J. B., Stage, F. K., King, J., Nora, A., & Barlow, E. A. (2006). Reporting structural equation modeling and confirmatory factor analysis: A review. *Journal of Educational Psychology*, *99*, 323-337.
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, *54*, 93-105.
- Sherman, R. C., Buddie, A. M., Dragan, K. L., End, C. M., Finney, L. J. (1999). Twenty years of PSPB: Trends in content, design, and analysis. *Personality & Social Psychology Bulletin*, *25*, 177-187.
- Skitka, L. J., & Sargis, E. G. (2006). The internet as psychological laboratory. *Annual Review of Psychology*, *57*, 529-555.
- Special issue. (2002). *Prevention & Treatment*, *5*.
- Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effect than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, *89*, 845-851.
- Stage, F. K., Nora, A., & Carter, H. C. (2004). Path analysis: An introduction and analysis of a decade of research. *Journal of Educational Research*, *98*, 5-12.
- Strohmetz, D. B. (2008). Research artifacts and the social psychology of psychological experiments. *Social and Personality Compass*, *2*, 861-877.
- Theory construction in social personality psychology: Personal experiences and lesson learned. (2004). Special Issue in *Personality and Social Psychology Review*, *8*.
- Van Selm, M., & Jankowski, N. W. (2006). Conducting online surveys. *Quality and Quantity*, *40*, 435-456.
- Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, *25*, 1115-1125.
- Wintre, M., North, C., & Sugar, L. (2001). Psychologists' response to criticisms about research based on undergraduate participants: A developmental perspective. *Canadian Psychology*, *42*, 216-225.