
Reading Empirical Research

In traditional arts and humanities scholarship, the principal publication is the monograph or book. Essays and scholarly articles are common, but not as important as books. In the sciences and social sciences, the principal publication is the journal article. Journals may contain letters to the editor, book reviews, and other items, but they mainly contain research articles.

Three types of research articles can be distinguished: the *research report*, the *meta-analysis* study, and the *review article*. The *research report* is used to communicate the results of some original research. Perhaps 95% of all articles in the science and social sciences are research articles. The *review article* is a more theoretical work that summarizes many studies related to a particular topic and attempts to present a coherent interpretation about what all of the research implies. An example of a review article is a paper by Patrik Juslin and Petri Laukka (2004), reviewing research on music and emotion. Sometimes a review article does the reverse: it may draw attention to the incoherent and contradictory state of research pertaining to some topic, and may offer advice for future research that might help to resolve the mess. Most review articles are well-written and highly informative. They provide a good place to start when you begin working on a new project. Unfortunately, review articles are rarely labelled “I’m a review article.” Some journals, like *Psychological Review* specialize in publishing review articles. Because review articles are popular with readers, commercial publishers often aim to produce “Handbooks,” where each chapter reviews research on a specific topic written by a different author.

A rarer type of article is the meta-analysis study. The meta-analysis is a “study of studies.” It is typically done when a large number of studies have been carried out related some problem. For example, many studies have been carried out related to whether television violence promotes violent behavior in viewers. Over a hundred studies have been carried out on this topic. Some of the studies seem to show a link, whereas other studies seem to show no link. In a meta-analysis, the researchers identify all of the pertinent studies. They then evaluate the quality of each study, including the quality of the samples used, the number of participants, the quality of the stimuli, the extensiveness of the controls, and other factors. Poor studies are simply discarded if they fail to achieve the minimum quality criteria established by the researchers. Then the researchers combine together all of the good studies, and do a statistical analysis on the aggregate data. The aim is to see if all of the studies ultimately tell a coherent story. In critically reading any meta-analysis, the main concern is how the author(s) set(s) the criteria for including or excluding a study in the final data analysis. Does the exclusionary criteria seem sensible? In general, meta-analyses are rare in music research. However, a fine example is found in the work of Friedrich Platz and Reinhard Kopiez (2012) who examined the influence of visual information in assessing the quality of a musical performance.

If it is a research report, determine what kind of study it is. Is it an exploratory study, descriptive study, correlational study, experimental study, modeling study, or some combination?

Assessing Tests of Hypotheses

For correlational and experimental studies here are some useful guiding questions:

1. What is the hypothesis (or hypotheses) being tested? Is the statement of hypothesis clear?
2. Identify the key terms in each hypothesis and ask how each term is operationally defined. How do the operational definitions deviate from the main sense of the corresponding concept?
3. What is the population? It is actually rare for research reports to explicitly identify the population. Much research simply assumes that the population is all people.
4. Is the sample reasonably representative of the presumed population?
5. How many different stimuli are used? (The fewer the number of stimuli used, the less the data independence: the conclusions may pertain only to those specific stimuli rather than to a general class of stimuli.)
6. What variables are controlled?
7. What variables are left uncontrolled? (Remember that not all variables can be controlled: The main question is whether the research controlled those variables mostly likely to produce confounding effects.)
8. What are the dependent variables?
9. Do the dependent measures rely solely on self-report? (Self-report is more susceptible to demand characteristics.)
10. Are the dependent measures *implicit* (less susceptible to demand characteristics and confabulation), or *explicit* (less good).
11. Given the design, what demand characteristics might be reasonably expected? Might these interfere with the experiment?
12. Are the conclusions over-generalized?
13. Was more than one experiment carried out? Is there converging evidence arising from different operationalizations, different stimuli, or different subjects?

If you are well-versed in statistics, then you can also address the quality of the statistical analyses. Are the results statistically significant? At what confidence level? Which variables are nominal, ordinal, interval or ratio? Are parametric or non-parametric tests used? Were tests of normalcy carried out? Are any statistical model assumptions violated? Did the author(s) perform multiple tests? If so, did they correct for multiple tests?

Effect Size

A common mistake by novices is to criticize an experiment for having too few participants. Experiments may have as few as five or six participants. The

importance of the results depends largely on the *effect size* (also known as the *magnitude of the effect*). Recall our earlier discussion about testing the toxicity of cyanide. We don't need many subjects if the effect size is large.

Even in the case of music experiments, we will often see large effect sizes. The effect may be clearly evident in every one we test with only a few stimuli. For Western-enculturated listeners, does playing a slow passage in the minor mode sound "sadder" than playing the same passage in the major mode? Five or six participants listening to two or three stimuli may be all that is necessary in order to find statistically significant results consistent with the hypothesis.

Effect sizes can be reported in percentages or using statistics such as *d-primes*, *Cohen's d*, r^2 , R^2 , or *phi* (ϕ).

In most research, we are first interested in whether the results are statistically significant. If not, we know that the results can simply be attributed to chance. If the results are significant, then we can pay attention to the effect size. There are many results that are significant, but with small effect sizes. Small effect sizes suggest that the relationship is not strong, and that we should consider looking at other factors that may play a bigger role. On the other hand, an effect size might appear weak simply because there are some confounding elements that haven't been properly controlled. It is possible that, with the right controls, a small effect will turn out to be a big effect that was diluted.

Take-Home Messages

When we finish reading a research article, our natural tendency is to make a mental note of the "conclusion." We want to distill what we've read into some sort of "take-home message." For example, after reading an article by Smith and Jones, we may try to remember that they showed (say) that lower pitch is heard as sounding sadder. Good researchers try to avoid such simple distillations. The sign of a careful researcher is that they remember a few important details about the study. Instead of simply recalling "low pitch is heard as sadder," they will recall something like: "in a two-alternative forced-choice experiment, American non-musician listeners report that the melody with two or three pitches lowered by a semitone sounds sadder."

Research is difficult enough without being burdened by ideas that are wrong. It's not what you don't know that is most problematic: it's what you think is right (but is wrong) that causes the greatest havoc. You never know when it will be appropriate to reinterpret a past result. Instead of showing "low pitch sounds sadder," you may later realize that a better interpretation is "lower-than-normal pitch sounds sadder," or "higher-than-normal pitch sounds happier (less sad)." Research results may depend critical on the specific question. For example, a researcher may get different results if they ask participants to judge "unpleasantness" rather than "dissonance." So it may be important to recall that experiment X measured "unpleasantness" whereas experiment Y measured "dissonance." Conversely, many judgments are highly correlated. Researchers may find they get the same results whether they ask

listeners to judge “sadness” or “sleepiness.” There are always other ways of interpreting the results of an experiment.

For important experiments, good researchers will recall the dependent and independent variables used and the way that the terms of the hypothesis were operationalized. They’ll also remember whether the effect size was large or small. Finally, they’ll pay attention to whether other researchers have replicated the result.

Further Information

Professional journals typically want authors to be brief and to-the-point. This is especially true in printed (rather than online) journals. Journal editors sometimes ask authors to shorten or summarize an otherwise verbose description. Consequently, journal articles may omit various details about the research. If you have a question about some published article, don’t be afraid to write to the author. Most researcher’s are happy to answer questions about their work.

References

Friedrich Platz and Reinhard Kopiez (2012). When the eye listens: A meta-analysis of how audio-visual presentation enhances the appreciation of music performance. *Music Perception*, Vol. 30, No. 1, pp.71-83.