Research & Interpreting Study Findings

Evidence-Based Practice in Speech-Language Therapy
(SHSC 2033)

Session 2

Thomas Klee & Elizabeth Barrett

香 港 大 學
THE UNIVERSITY OF HONG KONG

Outline

1. The structure of research papers
2. How research gets published
3. Interpreting study findings
4. Group discussion

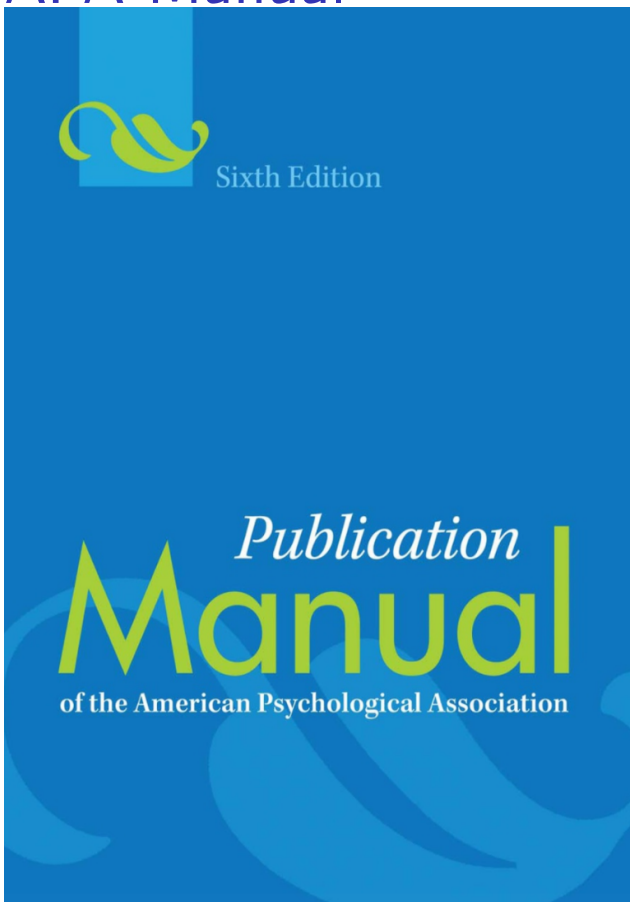
Structure of Research Papers

Structure of research papers

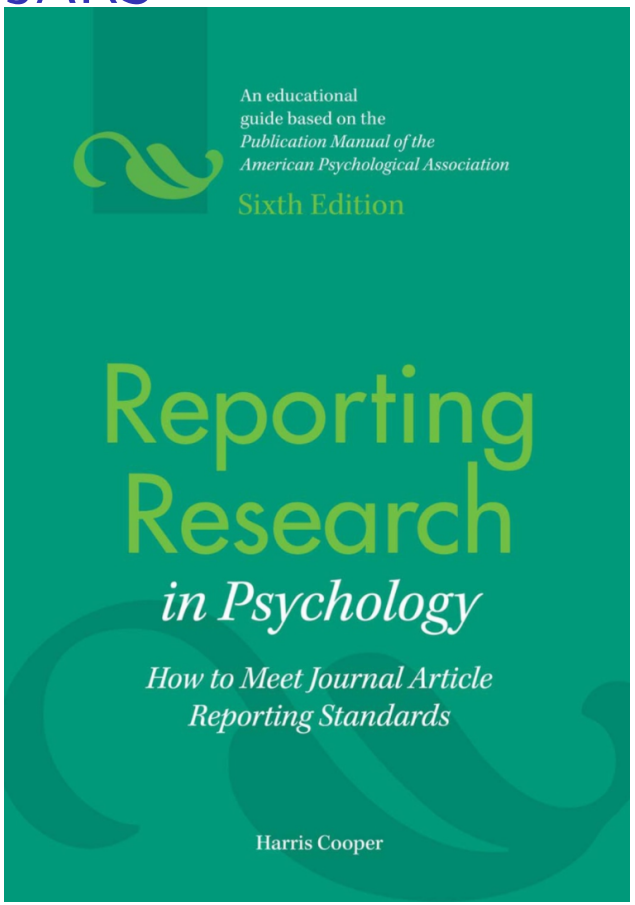
1. Title, authors, institutions
2. Abstract (or Structured Abstract)
3. Introduction
   - Statement of the problem
   - Literature review
   - Research questions; hypotheses
4. Method
   - Participants
   - Procedures
   - Data analysis
5. Results (e.g. summary statistics; statistical models)
6. Discussion
7. References

APA standards

APA Manual



JARS



APA style: good online resources

First stop for information

`https://owl.english.purdue.edu/owl/resource/560/01/`

FAQs about APA style

`http://www.apastyle.org/learn/faqs/index.aspx`

How Research Gets Published

How research gets published

- Studies are usually published in research journals.
- They can also be published online, in a book or elsewhere.
- Research journals usually have a rigorous **peer-review process**:
  1. Author submits manuscript to journal.
  2. Editor sends it out for peer review.
  3. Reviewers critique it and recommend accepting, rejecting or revising it.
  4. Editor makes final decision and informs author.
  5. If the decision was to revise, 1–4 are repeated.
- Some journals have higher standards than others. Some are predatory are should be avoided. More and more are **open access**.

Interpreting Study Findings

Alternative definition of EBP

"...the use of mathematical estimates of the risk of benefit and harm, derived from high-quality research on population samples, to inform clinical decision making in the diagnosis, investigation or management of individual patients." (Greenhalgh, 2010, p. 1)

Interpreting study findings

Important questions to ask (as a reader) or address (as a writer):

1. Was the study finding statistically significant?
2. If not, was the study's statistical power adequate?
3. Is the finding important?
4. How precise is the finding?

The last two points are much more important than the first two.

All four points are addressed on CATE, items 11–14.[1]

In next week's group discussion, you'll critically appraise a research paper using CATE.

[1] (Dollaghan, 2007, p. 67)

Statistical power

- **Power** estimates the probability of detecting a statistically significant difference in the sample *if one exists in the population*.
- Power should be calculated at the planning stage—before any data are collected. Studies not adequately powered are rarely worth doing.
- Ranges from 0 to 1.00
- Should be at least .80 *'in order to interpret findings that are not statistically significant with reasonable confidence'* [2]
- Check Methods and Results sections to see if power is reported (CATE item 12).

[2] Dollaghan (2007, p. 38)

p-values

"You should not forget that p actually stands for the conditional probability:

p(Data + $|H_0$ and all other assumptions),

which represents the likelihood of a result, or outcomes even more extreme (Data +), assuming

1. *the null hypothesis is exactly true;*
2. *the sampling method is random sampling;*
3. *all distributional requirements, such as normality and homoscedasticity, are met;*
4. *the scores are independent;*
5. *the scores are also perfectly reliable; and*
6. *there is no source of error besides sampling or measurement error." (Kline, 2013, p. 74)*

Statistical significance

- H_0 is the hypothesis being evaluated with the statistical test:
  - that there is no difference between the intervention and control groups
- $\alpha < .05$ is a typical criterion specified for deciding whether to reject H_0.
- If $p < .05$, then we reject H_0.
  - This suggests that there is reason to doubt that the true effect is zero.[3]
- Notice this doesn't tell us anything about the size of the difference between groups or whether the difference is (clinically) important.

[3] Wording suggested by Spence and Stanley (2018, p. 4).

What do small p-values actually mean?

"Statistical significance indicates that a small number of other hypothetical results (typically less than 5% from a very large number of hypothetical results) would be as extreme or more extreme than what was observed in the current study, when it is assumed that the null hypothesis is true." (Spence & Stanley, 2018, p. 3)

Why do this?

"The short hand interpretation we provide (i.e., interpreting statistical significance as "may not be zero,") can be viewed as a safety feature that may reduce science communication accidents when significance testing is used when communicating with the general public. Our short-hand interpretation also has a clear advantage of making it readily apparent how uninformative significance testing is on its own." (Spence & Stanley, 2018, p. 4)

Summary and cautions

- When a research finding is **statistically significant** (e.g., $p < .05$), it means that the true effect *may not be zero*.[4]
- Be careful to not over-interpret this. Even though researchers and statisticians frequently say so, it does not mean that was finding
  - was unlikely to have happened by chance, or
  - had a small probability (e.g. $< 5\%$) of occurring if the null hypothesis was true.
- CATE item 11

[4] Wording suggested by Spence and Stanley (2018).

Statistical significance

There may be several reasons why a finding wasn't statistically significant:

- The intervention wasn't effective.
- Measurement error was high, possibly masking any true effect. Were the measures employed valid and reliable?
- The sample size was too small to detect a difference of that magnitude.

A shift from using p-values recommended by many, including the APA

"It is past time for p-values to be retired. They do not do what is claimed, there are better alternatives, and their use has led to a pandemic of over-certainty. . . there is no justification for p-values." (Briggs, 2019, p. 22)

Effect size

- An alternative to p-values is to report an **effect size** and its **confidence interval**.
- A statistically significant finding (e.g. $p < .05$) doesn't tell us anything about how large the effect was (e.g., difference between groups; correlation).
- Just because a finding is statistically significant doesn't mean it's important; it just means that the true effect is not likely to be zero.
- Effect size measures can be calculated to help determine the importance of the research finding.
- CATE item 13

Effect size

- A measure that estimates the average size of the treatment effect across individuals (other kinds of ESs also exist)
- A measure of the average difference between intervention and control groups that occurred in a trial
- Unstandardized vs standardized ESs
- To calculate ESs, see `http://www.polyu.edu.hk/mm/effectsizefaqs/calculator/calculator.html`

Interpreting standardised ES measures [5]

1. Comparison of independent means (d, Δ, Hedges' g)
   - Small .20
   - Medium .50
   - Large .80
2. Correlation (r)
   - Small .10
   - Medium .30
   - Large .50
3. Multiple regression (R^2)
   - Small .02
   - Medium .13
   - Large .26
4. ANOVA (η^2)
   - Small .01
   - Medium .06
   - Large .14

[5] Cohen's conventions (Ellis, 2010)

Example of an ES

1. Research question: Does focussed stimulation improve the expressive vocabulary of late talkers?
2. Summary statistics [6]
   - Outcome measure: NDW at post-test
   - Intervention group ($n = 12$): $M_1 = 64.5; SD = 46.0$
   - Control group ($n = 13$): $M_2 = 25.2; SD = 22.0$
3. Effect size (Glass's Δ used in this case)
   - $\Delta = (M_1 - M_2)/SD_{control}$
   - $\Delta = (64.5 - 25.2)/22.0 = 39.3/22.0 = 1.79$
4. On average, the vocabulary size (NDW) of those receiving intervention increased by almost 2 SDs compared to those who didn't receive intervention.

[6] Girolametto, Pearce, and Weitzman (1996)

Interpreting ES

- The larger the ES, the greater the impact of the intervention on the behaviour being measured (provided the dependent variable is meaningful).
- If the ES of a new intervention exceeds that of the current intervention, it may be worth considering using the newer intervention (all else being equal).
- The ES (e.g. d) indicates the average amount of gain that can be expected from the intervention. The actual amount will vary from individual to individual.

Other important considerations

A study's importance should also be judged by looking at its **practical significance**.

- Intervention **efficacy** vs intervention **effectiveness**
  - Effectiveness studies sometimes called **pragmatic trials**
- **Social validity** of the intervention
- **Maintenance** of the intervention (short- vs long-term effects)

Precision of the findings

- An ES is calculated from sample data. It's a **point estimate** — an estimate of the ES of the population.
- If a second sample were measured, the ES calculated would be different.
- If you measured a large number of samples under identical conditions, you could calculate the range within which 95% of them would fall.
- This is called the **95% confidence interval** (95% CI).

Precision

- The 95% CI allows you to estimate an intervention's ES in the population.
- If the 95% CI excludes zero, then you know the intervention had an effect (i.e., it was effective).
- CATE item 14

Group discussion

- Break up into your assigned groups.
- Discuss items on the Study Evaluation form on Moodle for the research article you read.
- After that, we'll discuss some of the questions you posted on Moodle.

References

Briggs, W. M. (2019). Everything wrong with P-values under one roof. In V. Kreinovich, N. N. Thach, N. D. Trung, & D. V. T. Editors (Eds.), *Beyond traditional probabilistic methods in economics* (pp. 22–44). Retrieved from http://link.springer.com/10.1007/978-3-030-04200-4 doi: 10.1007/978-3-030-04200-4

Dollaghan, C. A. (2007). Appraising diagnostic evidence. In C. A. Dollaghan (Ed.), *The handbook for evidence-based practice in communication disorders* (pp. 81–104). Baltimore, MD: Paul H. Brookes.

Ellis, P. D. (2010). *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*. Cambridge: Cambridge University Press.

Girolametto, L., Pearce, P. S., & Weitzman, E. (1996). Interactive focused stimulation for toddlers with expressive vocabulary delays. *Journal of Speech and Hearing Research*, *39*, 1274–1283.

Greenhalgh, T. (2010). *How to read a paper: the basics of evidence-based medicine* (4th ed.). Chichester: Wiley-Blackwell BMJ Books.

Kline, R. B. (2013). *Beyond significance testing: statistics reform in the behavioral sciences*. Washington, D.C.: American Psychological Association.

Spence, J. R., & Stanley, D. J. (2018). Concise, simple, and not wrong: In search of a short-hand interpretation of statistical significance. *Frontiers in Psychology*, *9:2185*, 1–5. doi: 10.3389/fpsyg.2018.02185