BINUS
UNIVERSITY
Faculty of
Humanities

Psychology

RELAWAN
JURNAL INDONESIA

Pusat Studi
Publikasi Ilmiah

Juneman Abraham, *http://juneman.medium.com* | Zoom, 31/08/2021

Krisis Replikasi: Apa Artinya untuk Dosen dan Peneliti Indonesia

IS THERE A REPRODUCIBILITY CRISIS?

7% Don't know

52% Yes, a significant crisis

3% No, there is no crisis

1,576 researchers surveyed

38% Yes, a slight crisis

nature

Baker, M, Penny, D (2016) Is there a reproducibility crisis? Nature 533: 452–453.

Many Psychology Findings Not as Strong as Claimed, Study Says — The New York Times

Over half of psychology studies fail reproducibility test — nature

Scientists replicated 100 recent psychology experiments. More than half of them failed. — Vox SCIENCE & HEALTH

Journal	% Findings Replicated
Journal of Personality and Social Psychology: Social	23
Journal of Experimental Psychology: Learning, Memory, and Cognition	48
Psychological Science, social articles	29
Psychological Science, cognitive articles	53
Overall	36

Psikologi

Classical results

- Smiling will make you feel happier

In 1988, researchers reported that participants found cartoons funnier when they held a pen between their teeth, forcing them to smile, as compared with when they held a pen between their lips, forcing them to pout. The finding appeared to be consistent with the facial-feedback hypothesis – the idea that our facial expression doesn't just reflect our feelings but also affects them – and according to Google Scholar it has been cited nearly 1500 times.

Classical results

- Babies are born with the power to imitate

Pick up almost any introductory psychology book and inside you'll read about research conducted in the 1970s that appeared to show that humans are born with the power to imitate.

Earlier this year, however, a methodologically rigorous investigation found no evidence to support the idea that newborn babies can imitate. Janine Oostenbroek and her colleagues tested 106 infants four times between the ages of one week and nine weeks. The researcher performed a range of facial movements, actions or sounds for 60 seconds each including tongue protrusions, mouth opening, happy face, sad face, index finger pointing and mmm and eee sounds. Each baby's behaviour during these 60-second periods was filmed and later coded according to which faces, actions or sounds, if any, he or she performed during the different researcher displays.

Cleaning your hands will wash away your guilt

"Out, damned spot!" cries a guilt-ridden Lady Macbeth, obsessively washing her hands in the hope it will clear her conscience. Many research findings have demonstrated the that the link between moral purity and physical cleanliness is more than metaphorical, and that when we're feeling guilty we're motivated to clean ourselves physically.

In one of the earliest examples of the "Macbeth Effect", Chen-Bo Zhong and Katie Liljenquist asked participants to hand-copy an account of either an ethical or an unethical deed (helping or sabotaging a work colleague, respectively), and then asked them to rate the desirability of various products. Those who'd written about an unethical deed rated hygiene-related products more highly, such as soap and toothpaste.

Hargai Replikasi yang sungguh

- Smiling will make you feel happier
- Self-control is a limited resource
- Cleaning your hands will wash away your guilt
- Being reminded of money makes us selfish

CONCLUSION 3-1: For this report, *reproducibility* is obtaining consistent results using the same input data; computational steps, methods, and code; and conditions of analysis. This definition is synonymous with "computational reproducibility," and the terms are used interchangeably in this report.

***Replicability* is obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data.**

Regardless of the specific terms used, the underlying concepts have long played essential roles in all scientific disciplines. These concepts are closely connected to the following general questions about scientific results:

- Are the data and analysis laid out with sufficient transparency and clarity that the results *can be checked*?

- If checked, do the data and analysis offered in support of the result *in fact* support that result?

- If the data and analysis are shown to support the original result, can the result reported be found again in the *specific study context* investigated?

- Finally, can the result reported or the inference drawn be found again in a *broader set of study contexts*?

https://www.ncbi.nlm.nih.gov/books/NBK547546/

False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant

Joseph P. Simmons[1], Leif D. Nelson[2], and Uri Simonsohn[1]
[1]The Wharton School, University of Pennsylvania, and [2]Haas School of Business, University of California, Berkeley

Abstract

In this article, we accomplish two things. First, we show that despite empirical psychologists' nominal endorsement of a low rate of false-positive findings ($\leq .05$), flexibility in data collection, analysis, and reporting dramatically increases actual false-positive rates. In many cases, a researcher is more likely to falsely find evidence that an effect exists than to correctly find evidence that it does not. We present computer simulations and a pair of actual experiments that demonstrate how unacceptably easy it is to accumulate (and report) statistically significant evidence for a false hypothesis. Second, we suggest a simple, low-cost, and straightforwardly effective disclosure-based solution to this problem. The solution involves six concrete requirements for authors and four guidelines for reviewers, all of which impose a minimal burden on the publication process.

Keywords

Michael Inzlicht, University of Toronto, Associate Editor, Psychological Science

- The term "**file drawer problem**" was coined by the psychologist Robert Rosenthal in 1979. Positive-results bias, a type of publication bias, occurs when authors are more likely to submit, or editors accept, positive compared to negative or inconclusive results.

- If we found out that that 9 study JPSP paper was actually the results of 40 actual studies being run, it changes how we evaluate the strength of evidence. We still *might* believe the effect, but our confidence is greatly diminished.

- The problem with this is that now the size and veracity of an entire literature comes under deep questioning; are all those results great evidence that your phenomenon of interest is real and robust or is it merely the product of hundreds of motivated researchers around the world trying, repeatedly to find an effect and occasionally finding them?

Penelitian Replikasi Tidak Menjamin Kesetaraan Budaya

1. Karena sampel kesetaraan budaya sulit untuk ditemukan

Meskipun berusaha untuk memastikan bahwa sampel diambil dari kelompok setara (Mahasiswa, ulama, anak sekolah, dokter) dalam budaya mereka sendiri, tidak adai kemungkinan bahwa label ini memiliki konotasi yang berbeda dalam pengaturan yang berbeda. Misalnya, seorang imam di Melbourne (Australia) mungkin telah melalui pengalaman yang sangat berbeda dengan imam di Cape Coast (Ghana).

2. Karena beberapa peneliti lebih teliti daripada yang lain

Dalam semua tetapi beberapa kasus yang jarang terjadi, replikasi tidak dilakukan oleh peneliti yang sama dengan yang melakukan aslinya. Jadi meskipun semua psikolog eksperimental mengikuti pedoman profesional ketika melakukan penelitian, ada kemungkinan bahwa beberapa lebih teliti daripada yang lain ketika pemantauan perbedaan antara kondisi dalam studi mereka, atau ketika mengalokasikan peserta untuk kelompok dan sebagainya (Segall et al., 1990). Dalam penelitian lintas budaya adalah wajar untuk mengasumsikan bahwa inkonsistensi antara peneliti diperbesar karena kecenderungan untuk mengontrolpenelitian sampai ke detail terakhir mungkin berbeda-beda di pengaturan budaya.

3. Karena replikasi bisa kehilangan sesuatu dalam terjemahan

Dalam setiap penelitian yang dikontrol sangat penting bahwa semua peserta menerima standar instruksi. Biasanya, para peneliti baik membaca atau mendistribusikan instruksi yang sudah disiapkan sebelumnya sehingga semua orang mengetahui sama dan jelas tentang apa yang diharapkan dari mereka. Dalam replikasi lintas-budaya petunjuk ini telah diterjemahkan ke dalam bahasa yang berbeda, jadi tentu beberapa istilah tidak diterjemahkan dengan kesetaraan yang sempurna. Seba-gai contoh, persyaratan untuk "menga-takan hal pertama yang muncul di kepala Anda"

KUALITAS, KUALITAS, KUALITAS? REPLIKABILITAS!

- *DEFINITION OF REPLICABILITY: In empirical studies with sampling error, replicability refers to the probability of a study with a significant result to produce a significant result again in an exact replication study of the first study using the same sample size and significance criterion (Schimmack, 2017).*

BIDANG LAIN

"Cancer Research Is Broken"

Posted by Andrew on 20 April 2016, 9:13 am

https://statmodeling.stat.columbia.edu/2016/04/20/cancer-research-is-broken/

Michael Oakes pointed me to this excellent news article by Daniel Engber, subtitled, "There's a replication crisis in biomedicine—and no one even knows how deep it runs."

Engber suggests that the replication problem in biomedical research is worse than the much-publicized replication problem in psychology.

One reason, which I didn't see Engber discussing, is financial incentives. Psychology researchers typically don't like criticism of their published work: no surprise, they're people like everyone else, also their careers are at stake. But in biomedicine, it's not just careers and reputations on the line, it's big money. Lots more incentives to cheat, or to use sloppy methods that can be twisted to get that magic p-value, etc.

The replication crisis has engulfed economics

November 3, 2015 7.31am WIB

No two alike? Image sourced from Shutterstock.com

- Surel
- Twitter   89
- Facebook   486
- LinkedIn
- Cetak

A sense of crisis is developing in economics after two Federal Reserve economists came to the alarming conclusion that economics research is usually not replicable.

The economists took 67 empirical papers from 13 reputable academic journals. Without assistance from the original researchers they were only able to get the same result in a third of cases.

With the original researchers' assistance, that percentage increased to about half, suggesting reporting practices and requirements are seriously deficient.

https://theconversation.com/the-replication-crisis-has-engulfed-economics-49202

Penulis

Andreas Ortmann
Professor, UNSW

Pengungkapan

Andreas Ortmann tidak bekerja, menjadi konsultan, memiliki saham, atau menerima dana dari perusahaan atau organisasi mana pun yang akan mengambil untung dari artikel ini, dan telah mengungkapkan bahwa ia tidak memiliki afiliasi selain yang telah disebut di atas.

What is an effect size?

An effect size is a quantitative measure of the difference between two groups. In systematic reviews and meta-analyses of interventions, effect sizes are calculated based on the 'standardised mean difference' (SMD) between two groups in a trial – very roughly, this is the difference between the average score of participants in the intervention group, and the average score of participants in the control group).

HOW DO YOU INTERPRET AN EFFECT SIZE?

Effect sizes are usually reported using the label 'd=', and in the form of a fraction, such as d=0.2 or d=0.5. One of the most common ways of interpreting effect sizes is based on the work of a man named Cohen, who said that: 0.2 and below = small effect size; 0.5 = medium effect size; 0.8 and above = large effect size. While these interpretations are not uncontroversial and there are other ways to calculate and interpret effect size, Cohen's suggestions are generally accepted and are a good basis for interpreting the results of trials and in reading systematic reviews and meta-analyses.

## WHAT'S THE DIFFERENCE BETWEEN AN EFFECT SIZE AND STATISTICAL SIGNIFICANCE?

You might also see studies that report 'statistical significance', which tells you if an intervention had an effect that was unlikely to have happened by chance. While it is important to know this, it is not as useful for comparing effect sizes of multiple studies, as we do in systematic reviews. This is because statistical significance does not take into account sample size (i.e. the number of participants in a study). If two studies are identical except that one has a larger sample size, we would usually consider the study with the larger sample size to be more reliable, but statistical significance does not give more weight to a study with more participants – all studies are treated equally.

Effect sizes, on the other hand, are 'weighted' according to the number of participants in a study. For instance, a study with 10 participants might have had a big effect size (such as 0.8); while another study of the same intervention may have had 1000 participants but a small effect size (such as 0.2). If all other things are equal (e.g. both studies had a low risk of bias), then both studies may have shown that the intervention had a statistically significant effect, but the overall effect size would be small, because the larger of the two studies would be given more 'weight'.

- Effect sizes are shown in a forest plot. Find out what a forest plot is and how to read it.

How to read a forest plot

A forest plot is a graphical representation of a <u>meta-analysis</u>. It is usually accompanied by a table listing references (author and date) of the studies included in the meta-analysis. In the example here, there were three studies included in the meta-analysis: Evans 2011, Jones 2010 and Smith 2013. The table also lists the mean scores and standard deviations of these scores from each of the included studies; and it lists the number of participants in each study (under 'Total').

https://www.spi.ox.ac.uk/how-to-read-a-forest-plot
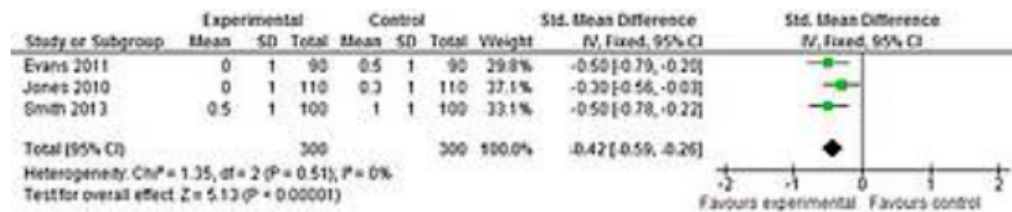
The forest plot is the graph on the right-hand side. It has one line representing each study in the meta-analysis, plotted according to the standardised mean difference (SMD – very roughly, this is the difference between the average score of participants in the intervention group, and the average score of participants in the control group). The green box on each line shows you the SMD for each study.

In the example, the first line represents the SMD of scores from the Evans 2011 study, (SMD=-0.50); the second line represents the SMD for scores from the Jones 2010 study (SMD=-0.30); and the third line represents these scores from the Smith 2013 study (SMD=-0.50).

The black diamond at the bottom of the graph shows the average effect size of the three studies.

- <u>How to interpret the sample forest plot</u>

Wiley Online Library

BAHASA

LANGUAGE LEARNING
A Journal of Research in Language Studies

SYSTEMATIC REVIEW ARTICLE | 🔓 Open Access | (cc) (i)

Replication in Second Language Research: Narrative and Systematic Reviews and Recommendations for the Field

Emma Marsden ✉, Kara Morgan-Short, Sophie Thompson, David Abugaber

A note from the Journal Editor (Pavel Trofimovich): This article is published with special permission from the board of directors of *Language Learning* , following regular peer review by four reviewers. The study emerged in conjunction with work on the multisite replication reported by Morgan-Short et al. (2018), funded by a *Language Learning* research grant to Marsden and Morgan-Short. That financial support was applied for and received before Marsden and Morgan-Short joined the editorial team of *Language Learning* .

https://onlinelibrary.wiley.com/doi/10.1111/lang.12286

What is Replication?

There are different types of replication. First, there is a type called "**exact replication**" (also called "**direct replication**"). In this form, a scientist attempts to exactly recreate the scientific methods used in conditions of an earlier study to determine whether the results come out the same. If, for instance, you wanted to exactly replicate Asch's (1956) classic findings on conformity, you would follow the original methodology: you would use only male participants, you would use groups of 8, and you would present the same stimuli (lines of differing lengths) in the same order. The second type of replication is called "**conceptual replication**." This occurs when—instead of an exact replication, which reproduces the methods of the earlier study as closely as possible—a scientist tries to confirm the previous findings using a different set of specific methods that test the same idea. The same hypothesis is tested, but using a different set of methods and measures. A conceptual replication of Asch's research might involve both male and female **confederates** purposefully misidentifying types of fruit to investigate conformity—rather than only males misidentifying line lengths.
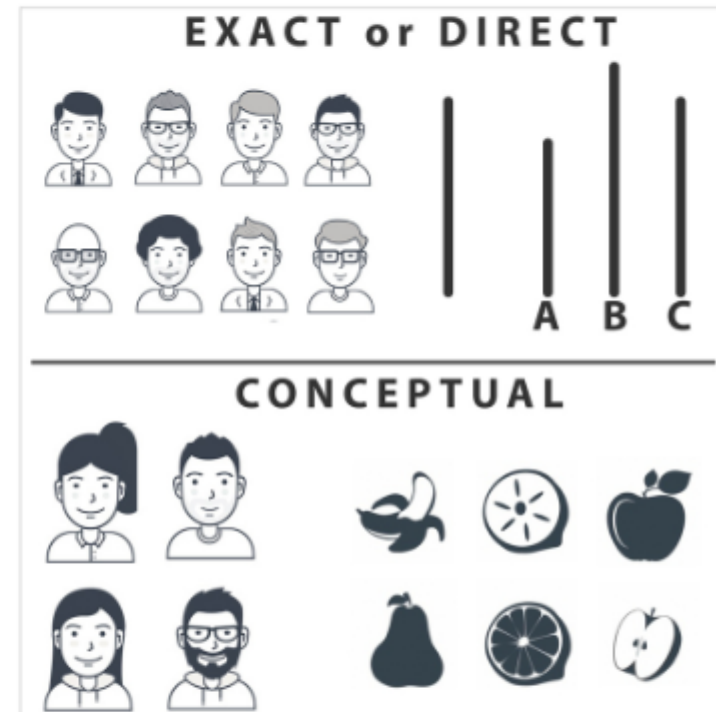

Figure 3.19: Example of direct replication and conceptual replication of Asch's conformity experiment.

https://openpress.usask.ca/introductiontopsychology/chapter/the-replication-crisis-in-psychology/

REPLIKASI "VS" KEBARUAN?

BMC Medical Research
Methodology

RESEARCH ARTICLE   **Open Access**

Attempting rigour and replicability in thematic analysis of qualitative research data; a case study of codebook development

Kate Roberts[1*], Anthony Dowell[2] and Jing-Bao Nie[3]

Abstract

Background: Navigating the world of qualitative thematic analysis can be challenging. This is compounded by the fact that detailed descriptions of methods are often omitted from qualitative discussions. While qualitative research methodologies are now mature, there often remains a lack of fine detail in their description both at submitted peer reviewed article level and in textbooks. As one of research's aims is to determine the relationship between knowledge and practice through the demonstration of rigour, more detailed descriptions of methods could prove useful. Rigour in quantitative research is often determined through detailed explanation allowing replication, but the ability to replicate is often not considered appropriate in qualitative research. However, a well described qualitative methodology could demonstrate and ensure the same effect.

Methods: This article details the codebook development which contributed to thematic analysis of qualitative data. This analysis formed part of a mixed methods multiphase design research project, with both qualitative and quantitative inquiry and involving the convergence of data and analyses. This design consisted of three distinct phases: quantitative, qualitative and implementation phases.

Results and conclusions: This article is aimed at researchers and doctoral students new to thematic analysis by describing a framework to assist their processes. The detailed description of the methods used supports attempts to utilise the thematic analysis process and to determine rigour to support the establishment of credibility. This process will assist practitioners to be confident that the knowledge and claims contained within research are transferable to their practice. The approach described within this article builds on, and enhances, current accepted models.

Keywords: Thematic analysis, Rigour, Qualitative research, Codebook, Coding

Background

Navigating the world of thematic qualitative analysis can be challenging. Thematic analysis is a straightforward way of conducting hermeneutic content analysis which is from a group of analyses that are designed for non-numerical data. It is a form of pattern recognition used in content analysis whereby themes (or codes) that emerge from the data become the categories for analysis. These forms of analysis state that the material as a whole is understood by studying the parts, but the parts cannot be understood except in relation to the whole [1]. The process involves the identification of themes with relevance specific to the research focus, the research question, the research context and the theoretical framework. This approach allows data to be both described and interpreted for meaning.

In qualitative research replication of thematic analysis methods can be challenging given that many articles omit a detailed overview of qualitative process; this makes it difficult for a novice researcher to effectively mirror analysis strategies and processes and for experienced researchers to fully understand the rigour of the

* Correspondence: kate.roberts@postgrad.otago.ac.nz
[1]Department of Primary Health Care & General Practice, University of Otago, Wellington, New Zealand
Full list of author information is available at the end of the article

https://www.researchgate.net/publication/332059787_Attempting_rigour_and_replicability_in_thematic_analysis_of_qualitative_research_data_A_case_study_of_codebook_development

Initial Code Sources
- Research Literature
- Theory
- Quanitative data
- Summarisation and identification of initial themes

Initial code development
- Application of initial codes to raw qualitative data - inductive component
- Development of further codes from the raw data - deductive component
- Intercoder review and code testing
- Code/theme saturation achieved

Codebook development
- Codes labelled, described, defined, qualifications and exclusions identified, examples from the raw text confirmed
- Intercoder testing and training (inter-rater reliability and test-retest)

Codebook application
- Application of coding template (template analytic technique) to small subset of data
- Codebook review
- Intercoder testing (inter-rater reliability and test-retest)
- Application of codebook to entire dataset utilising technology
- Confirmation of coding categories complete

Interpretation
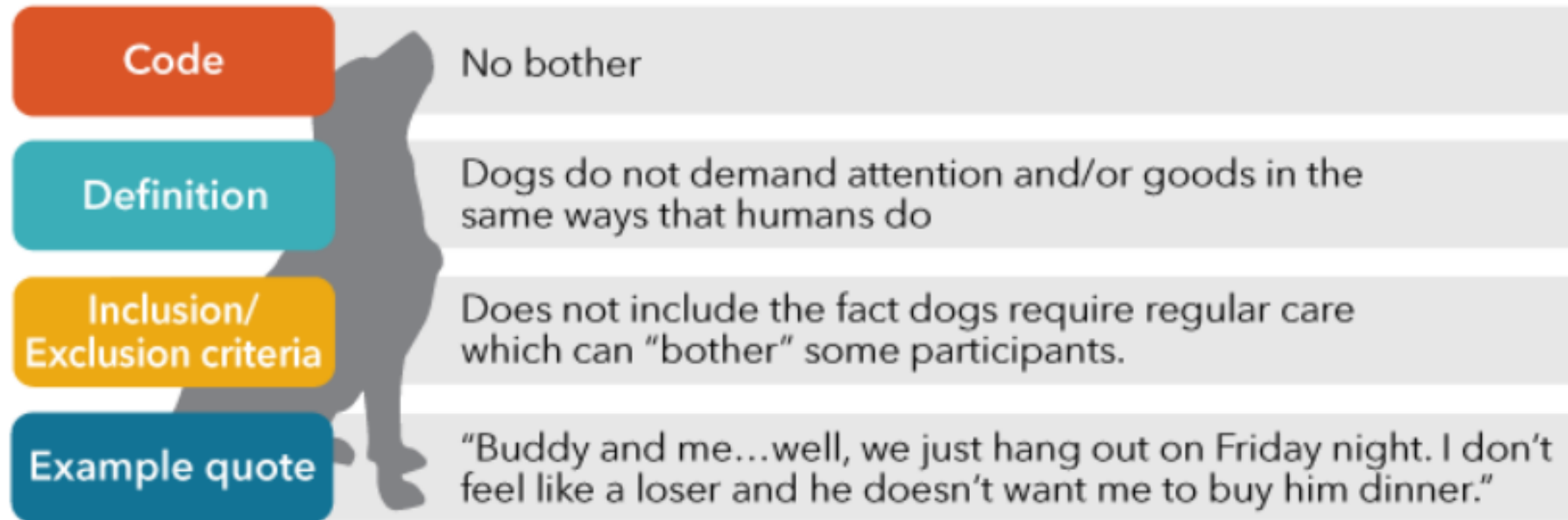- Conclusions and recomendations

Fig. 1 - Process of code creation and testing

Codes were written following the guidelines of Boyatzis [9] and were classified with the following: label, definition, description, qualifications or exclusions and examples from the raw data. An example of code labels from this research is outlined in Table 1 below.

Table 1 Example of a code definition form the code book

Code Label	Definition	Description	Qualifications or exclusions	Examples
Fear of rejection	Demonstrating anxiety about being cut-off demeaned or isolated. Fear of experiencing hurt, pain or embarrassment due to others' actions or words	Perceptions of each other based on beliefs or self-held doubts, unfounded opinions, rushing to an opinion without reason.	Can be fear of patient or practitioner being ridiculed - with or without basis	*It's like some shame based thing. The fear… they fear rejection from the practitioner…. I've had plenty of clients who have not told their GP you know, as if they're having this side relationship with another modality.* ACU013
Feelings of inferiority	Expressing a sense of division within a group of people.	Mention of power imbalance, being treated/acting differently, not feeling the 'same', differing world views	Expressed as feelings between clinicians rather than between patients.	*But I don't tend to do that with GPs, isn't that interesting? I hadn't actually thought about that. But I kind of feel like I know the physios better. We seem to be more… more on a level perhaps?* ACU012

Boyatzis RE. Transforming qualitative information : thematic analysis and code development. Thousand oaks CA: SAGE Publications; 1998.

Revisiting our SPCA study, a potential thematic code might be: "Dogs as the "perfect" companion". Underneath this code may be sub-codes such as:

Code	No bother
Definition	Dogs do not demand attention and/or goods in the same ways that humans do
Inclusion/ Exclusion criteria	Does not include the fact dogs require regular care which can "bother" some participants.
Example quote	"Buddy and me…well, we just hang out on Friday night. I don't feel like a loser and he doesn't want me to buy him dinner."

Figure 12.29. Example of a sub-code for the theme of dogs being 'no bother'.

https://tophat.com/marketplace/social-science/psychology/-/research-methods-in-psychological-science-laura-freberg/736/35315/

Testing the reliability of the code

Although inter-rater code testing and discussion occurred throughout the codebook development stage, the final codebook continued to be tested for inter-rater reliability before the data reached the interpretation stage. Reliability can be described as the consistency of judgement that protects against or lessens the contamination of projection [9]. Reliability was tested in this project in two ways:

1. Consistency of judgment over absence and presence (test-retest reliability); and
2. Consistency of judgement across various viewers (inter-rater reliability).

Both tests were checked for reliability using this formula suggested by Miles and Huberman [23]:

$$Reliability = \text{number of agreements}/\text{number of agreements} + \text{disagreements}$$

This calculation is a much cruder tool than Choen's kappa, but gives a simple measure of agreement as a

Table 2 Example of table used for absence/presence reliability

Subtheme/Code	Description
The role of research.	Mention of how research might inform decisions. Does research impact on relationships with each other? Purpose of research in practice.
Demonstrating competency	Discussion of what would display competence, what would enable development of trust between practitioners.
Science vs philosophy when defining acupuncture	Discussion of differing styles of acupuncture training and/or practice. Mention of either specifics of the western or eastern framework.

Subtheme/code	Coder 1 KR		Coder 2 TD		Coder 3 JN	
	Absent	Present	Absent	Present	Absent	Present
The role of research		Yes		Yes See Table 3		Yes See Table 3
Demonstrating competency		Yes		Yes – first Line 67		Yes p.3 lines 121–127; p. 6–7 lines 300–329
Science vs philosophy when defining acupuncture		Yes		Yes - Line 50, 78, 288		Yes p. 6 lines 287–294

Reliability = 6/6 + 0 = 1 High agreement

Testing Absence/presence of multiple sub-themes/codes within a single interview *Notes for coders- This is a reliability test for coding. Within these single interviews, please record whether the following codes/nodes are absent or present in the interview*

Table 3 Example of table used for inter-rater reliability

Participant	Coder 1 KR	Coder 2 TD	Coder 3 JN
ACU001	5 Lines 60–64 259–264 333–342 355–362 371–375	7 Lines 61–63 263/4 322–4 333–339 344–347 355–359 386–388	4 Lines 61–65 259–264 333–350 355–375
GP003	9 Lines 26–29 79–81 100–106 112–115 147–149 243–245 287–294 345–348 362–366	6 Lines 26–29 79–81 100–106 243–245 285–287 345–348	2 Lines 100–108 279–283

Reliability = 23/23 + 10 = .7 – Inadequate agreement reached

Testing frequency of observation of theme 'research versus relationships' – subtheme/code 'the role of research' in a subset of two interviews *Notes for coders - This is a reliability test for coding to assist with the demonstration of rigour within the data analysis. Within these two interviews, please record the number of times you would code text at 'the role of research' node. This relates to the category of defining current practice/research versus relationships/the role of research. Coding guide - Mention of how research might inform decisions. Does research impact on relationships with each other? Purpose of research in practice*

Psychology reproducibility 'crisis' overstated, new report claims

But authors of original paper hit back, saying latest assessment is 'very optimistic'

March 4, 2016

David Matthews

Twitter: @DavidMJourno

https://www.timeshighereducation.com/news/psychology-reproducibility-crisis-overstated-new-report-claims

Academics have continued to trade blows over the state of psychology research following the release of a paper questioning the results of a major project that cast doubt on reproducibility in the field.

In August 2015, an attempt to reproduce 100 prominent papers by the Center for Open Science found that only 36 per cent produced statistically significant results, stoking concerns about scientific reliability that have also engulfed biomedicine.

Source: iStock

But today saw a group of researchers from Harvard University and the University of Virginia respond with claims that the study contained several statistical errors and failed to repeat the experiments properly.

It lists a number of what it claims are discrepancies between the original studies and attempts at replication.

Increase the power of a hypothesis test

You can use any of the following methods to increase the power of a hypothesis test.

- Use a larger sample.
  Using a larger sample provides more information about the population and, thus, increase power. Using a larger sample is often the most practical way to increase power.

- Improve your process.
  For a hypothesis test of means (1-sample Z, 1-sample t, 2-sample t, and paired t), improving your process decreases the standard deviation. When the standard deviation is smaller, the power increases and smaller differences can be detected.

- Use a higher significance level (also called alpha or α).
  Using a higher significance level increases the probability that you reject the null hypothesis. However, be cautious, because you do not want to reject a null hypothesis that is actually true. (Rejecting a null hypothesis that is true is called type I error.)

Software is normally used to calculate the power.

- Calculate power in SAS.
- Calculate power in PASS.

Power Analysis

Power analysis is a method for finding statistical power: the probability of finding an effect, assuming that the effect is actually there. To put it another way, power is the probability of rejecting a null hypothesis when it's false. Note that power is different from a Type II error, which happens when you fail to reject a false null hypothesis. So you *could* say that power is your probability of not making a type II error.

A Simple Example of Power Analysis

Let's say you were conducting a drug trial and that the drug works. You run a series of trials with the effective drug and a placebo. If you had a power of .9, that means 90% of the time you would get a statistically significant result. In 10% of the cases, your results would not be statistically significant. The power in this case tells you the probability of finding a difference between the two means, which is 90%. But 10% of the time, you wouldn't find a difference.

https://clincalc.com/stats/samplesize.aspx

Statistically powerful

A within-subjects design is more statistically powerful than a between-subjects design, because individual variation is removed. To achieve the same level of power, a between-subjects design often requires double the number of participants (or more) that a within-subjects design does.

Problems with within-subjects designs

The biggest downsides of within-subjects designs are the potential threats to **internal validity**. Because of repeated testing over long time periods, time-related and carryover effects can confound the results of a study by presenting alternative explanations.

https://www.scribbr.com/methodology/within-subjects-design/

Preregistration Challenge - The Center for Open Science

https://cos.io/prereg/ ▼

What is **Preregistration**? When you **preregister** your research, you're simply committing to your plan in advance, before you gather data. **Preregistration** separates hypothesis-generating (exploratory) from hypothesis-testing (confirmatory) research.

Seven **easy steps** to publishing a **Registered Report**

DEVELOP IDEA
1 Submit your research question and study design
This includes your introduction, method, proposed analyses, and pilot data

EDITOR EVALUATION
2 The Journal's editorial team sends suitable submissions for peer review

STAGE 1 PEER REVIEW
3 Peer review of your research plan focusing on:
- The importance of the research questions
- The ability of the proposed methods to answer those questions

IN PRINCIPLE ACCEPTANCE
4 Your article will be accepted in principle for publication or rejection
This is a commitment from the journal to publish your article, if you follow the methodology outlined in the study design

WRITE REPORT
5 Conduct your study
Undertake your research according to the methodology outlined in the study design

STAGE 2 PEER REVIEW
6 Submit your article for re-review
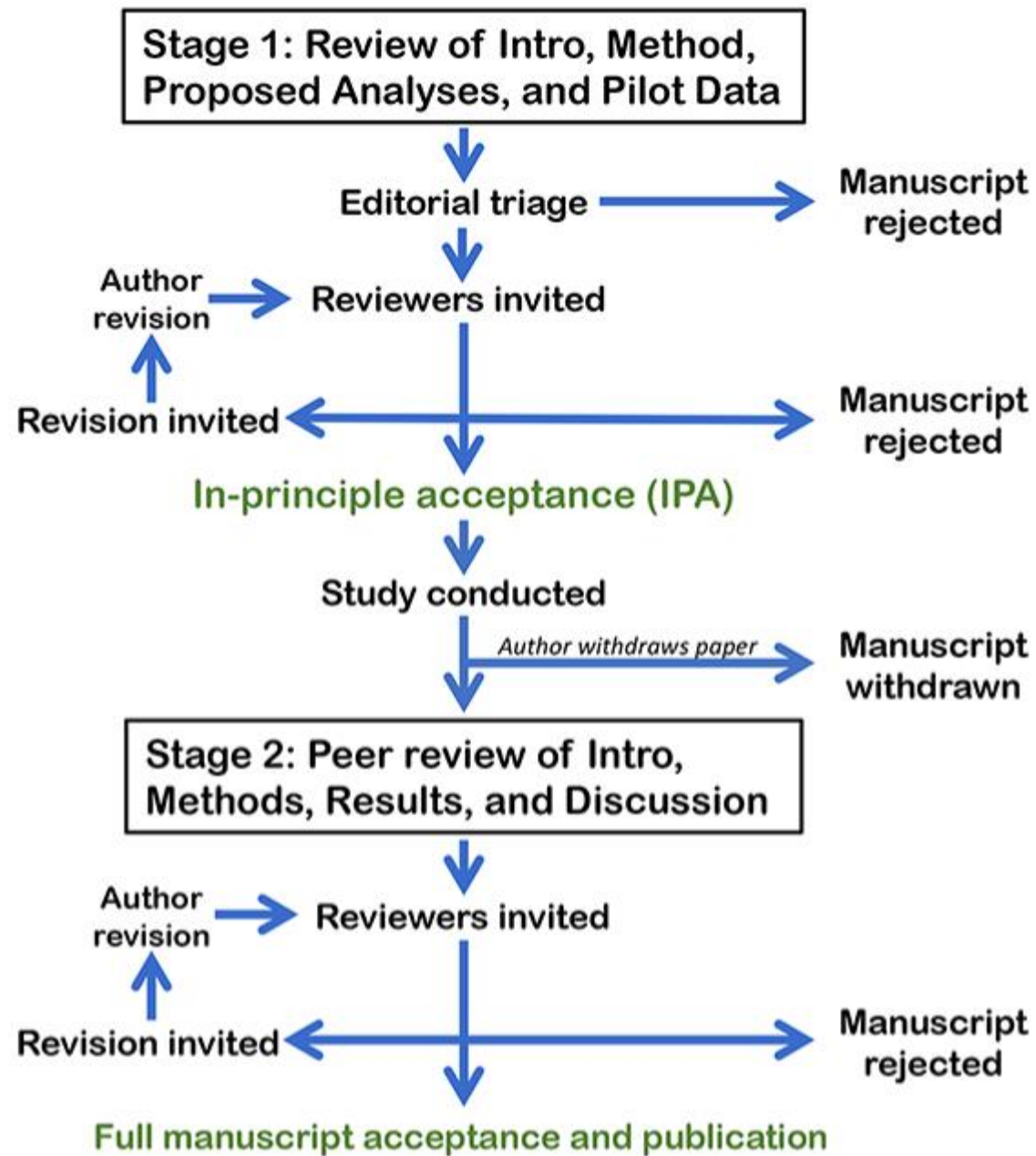Peer review assesses whether the study matches the approved research plan, and draws appropriate conclusions

PUBLISH REPORT
7 Successful publication of your final manuscript

WILEY

https://authorservices.wiley.com/author-resources/Journal-Authors/submission-peer-review/registered-reports.html

Home » Browse » National culture as a correlate of research output and impact

Check for updates

BRIEF REPORT

REVISED National culture as a correlate of research output and impact [version 3; peer review: 2 approved]

Previously titled: National culture as a correlate of research performance and impact

✉ Juneman Abraham (iD)

➕ Author details

This article is included in the Research on Research, Policy & Culture gateway.

ALL METRICS

2025
👁 VIEWS

168
⬇ DOWNLOADS

Get PDF

Get XML

Cite

Export

Track

Email

Share

Abstract

National culture has been overlooked in discussions related to research output and impact owing to individual, socio-political structure, and economic factors. This study shows the relationships between the dimensions of cultural value orientation of the nation and research output & impact. More than 60 countries were included, and Spearman correlation analysis was employed. The

Open Peer Review

Reviewer Status ✓ ✓   ⓘ

Reviewer Reports

Invited Reviewers

	1	2
Version 3 (revision) 07 Feb 20	✓ read	✓ read
Version 2 (revision) 24 Sep 19	? read ↑	? read ↑
Version 1 28 Feb 19	? read ↑	? read ↑

https://f1000research.com/articles/8-237