

Experiment Replication: A Proposed Solution for Developing Psychological Research in Indonesia

Muhammad Abdan Shadiqi

Faculty of Psychology, Universitas Indonesia
Department of Psychology, Universitas Lambung Mangkurat

Hamdi Muluk and Mirra Noor Milla

Faculty of Psychology
Universitas Indonesia

Is it possible that psychology can be as strong as natural science? Having replication studies could be the answer to this question. Philosophically, a replication is ‘the heart of any science,’ however it receives a little attention from social science. In Indonesia, there are three major problems: (1) only few number of researchers implement replication studies; (2) only few replication studies present strong evidence; and (3) only a small number of replication studies have been published. This might occur because the knowledge on how to conduct a replication study is inaccessible to most psychology researchers in Indonesia. This article explains a definition of a replication study, types of replications, and strategies to conduct replication experiments. We will explain how to conduct a replication study, starting from determining and reviewing reference articles to designing a replication study.

Keywords: replication study, experimental, methodology

Apakah mungkin ilmu Psikologi dapat sekuat ilmu alam? Salah satu cara untuk menjawab hal ini adalah dengan melakukan studi replikasi. Pada perspektif filosofi, replikasi adalah ‘*the heart of any science*’ tetapi bagi perspektif ilmu sosial kurang dihargai. Tiga permasalahan yang terjadi di Indonesia adalah: (1) sedikit jumlah peneliti yang melakukan studi replikasi; (2) sedikit hasil replikasi yang terbukti secara baik; dan (3) sedikit jumlah naskah studi replikasi yang dipublikasikan. Hal ini mungkin terjadi karena pengetahuan mengenai cara melakukan studi replikasi masih belum terjangkau ke banyak peneliti Psikologi Indonesia. Artikel ini menyajikan penjelasan mengenai definisi, jenis, dan strategi untuk melakukan studi replikasi eksperimen. Kami memaparkan langkah praktis dimulai dengan menentukan artikel acuan, mereview artikel, hingga merancang studi replikasi.

Kata kunci: studi replikasi, eksperimental, metodologi

More than four decades ago, Smith (1970) conveyed that psychologists tend to neglect replication studies. Many psychological studies have been questioned as they are difficult to replicate (Yong, 2012). For example, Yong explained that three different research team failed to replicate Bem’s experimental study (2011). Bem’s studies consisted of nine experiments about psi or psychic effects with precognition and premonition to memorize words. Bem asked college student volunteers to memorize 48 words in two different ways: asking participants to immediately write down the words that were memorized versus asking them to type the words beforehand (in the practice session) and then to recall them. The result

showed that participants could recall more words when they had time to practice them through typing. Bem’s study, according to LeBel and Peters (2011), emphasized heavily on conceptual replication, and did not pay attention to the importance of measurement and experiment procedures, and had mistakenly applied hypothesis significance testing.

Another example is ego-depletion research that was first conducted by Roy Baumeister and then continued by Michael Inzlicht. Winerman (2016) has specifically termed Inzlicht’s work on ego-depletion as *chasing puffs of smoke*. According to Winerman, Inzlicht had spent decades studying the effect of ego-depletion and proposing that self-control was an unlimited resource. However, Hagger et al. (2016) who conducted multi-lab experiments involving 23 laboratories and 2,141 participants found that the effect of ego-depletion was very small and even none. The two

Correspondence concerning this article should be addressed to Muhammad Abdan Shadiqi, Faculty of Psychology, Universitas Indonesia, Depok, Jawa Barat, Indonesia. E-mail: m.abdan_shadiqi@yahoo.co.id

examples of replication studies on psi effect and ego-depletion have indicated that building empirical evidence takes a lot of effort.

Another fascinating finding in the world of science is Fanelli's finding (2011). He found that there were too many "positive" results (confirmed hypotheses) reported in journals from 1990 to 2007. Fanelli's finding is based on his review on 4,600 studies reported in various scientific journals. The "positive" results were commonly reported in social science journals. It should be noted that Fanelli's finding is only limited to scientific articles. The number might increase if the review also included all other scientific products, such as thesis, dissertation, and research project reports.

Meanwhile, "negative" results (unconfirmed hypothesis) from replication studies are likely to be concealed, in other words not published. In fact, according to Matosin, Frank, Engel, Lum, and Newell (2014), "negative" results are important components for evaluation and critical validation towards scientific thinking. Therefore, scientists should report all data without worrying about the results because "negative results" are also important and scientists should report their findings based on sound hypotheses (Matosin et al., 2014). Further, after scientists or writers have the urge to report the "negative" results, journal editors or managers need to provide an opportunity for scientists or writers to publish the "negative" results.

The support from journal managers gives a significant impact to the publication of replication studies. From 465 journals in neuroscience, only 6% (28 journals) explicitly stated that they received the manuscripts of replication studies (Yeung, 2017). Notably, studies in neuroscience usually have a robust methodology with physiological and nerve measurement. Philosophically, neuroscience has a connection with psychology, particularly the cognitive approach (Machamer, 2002; Proctor & Capaldi, 2006). Meanwhile, Martin and Clarke (2017) reported that only 33 out of 1,151 psychological journals (3%) clearly indicated their acceptance to replication studies in the aims of journal or the instruction for authors. Due to the lack of support, replication studies in psychology are not yet frequently conducted, reported, and developed.

To fill in the gap, we will explain the concepts, procedures, and practical steps in conducting replication studies. It is expected that this paper will open an opportunity for researchers to strengthen and develop psychological theories in Indonesia through replication studies.

Definitions and Philosophical Foundations

A replication study can be defined as a means to reiterate an experimental procedure with the purpose to verify research findings (Kline, 2013). A replication study has the main function of increasing the generality of the findings (Allen & Preiss, 1993). This implies that successful replication requires some researchers generate the same findings and reach a universal conclusion. A replication study is a paradigm in natural science which is then adopted and used in social science. In natural science, studies are often rigorously conducted to search for evidence. Generally, replication studies are conducted using experimental methods. For the purpose of theoretical development and empirical findings, Allen and Preiss stated that replication studies are the stepping stones that provide information about the results of studies for a meta-analysis (a study featuring a quantitative synthesis of experiments and surveys). Simply put, according to Allen and Preiss, replications have a relationship with meta-analysis in terms of expanding theoretical issues and generalizing findings.

There are four important aspects of knowledge that direct the need for a replication study (Allen & Preiss, 1993), which are:

Stability. The need for a replication study is triggered by the needs to have a consistent result at different times and in different circumstances. For each experiment replication, the results should not change

Bias reduction. A replication study is needed to avoid inferring findings with misleading interpretations and biases. Ideally, there should be no individual biases influencing scientific findings.

Predictability. The need for a replication study is based on predictability within tolerance limits and accuracy when testing an effect. In psychology and social science in general, the prediction is not based on an individual score, but an average of several individual scores (mean scores of sample).

Contextual irrelevance. A replication study aims for generality that is supposed to be free from theoretical and contextual issues. Simply put, a replication study is needed to resolve relevance issues within the context of a theory.

Referring to stability aspect explained by Allen and Preiss (1993), new finding can refute the old finding; it means that a philosophical cornerstone namely "falsifiable" will play a role. Falsification is a concept proposed by a philosopher, Karl Popper, in relation to the structure of science (Ackermann,

2009). A theory can be falsified. However, a theory is considered good if during falsification, it is failed to be refuted or it is free from falsification (Bem & de Jong, 2006). According to Yusainy (2015) and LeBel, Berger, Campbell, and Loving (2017), a replication is an empirical mechanism to test, falsify or refute a theory.

Replication studies are used to enhance generalizability and reveal the strengths or weaknesses of a theory. From a philosophical perspective, Schmidt (2009) has placed replication studies as the heart of any science because they contribute enormously to science. However, it should be underlined that conducting a replication study is not a plagiarism because it follows systematic steps and research procedures. Colella-Sandercock (2017) explained a replication study is an alternative to avoid a plagiarism issue because researchers may determine parts of the research design in the original study that could be revised.

Literature Review of Replication Studies in Indonesia

Scientific articles written by Indonesian researchers or reporting studies with Indonesian samples, particularly experimental replication studies, are difficult to obtain either through Google Scholar or PsycINFO. We used PsycINFO database to search because it was considered as the most effective and efficient database for psychology that could track down replication studies reported by Indonesian researchers or conducted with Indonesian samples. Finding that there are few replication studies conducted by Indonesian researchers in PsycINFO implies that replication studies have been reported properly in international journals. However, there might be other replication studies that have not yet reported or published in PsycINFO indexed journals. Therefore, in this article we will use several examples of replication studies in psychology conducted by postgraduate students in Universitas Indonesia. Meanwhile, the search with Google Scholar was deemed ineffective (see Nordhays & Moffat, 2017) because the results was too broad and not specific to psychological studies. Furthermore, Google Scholar did not only show scientific journals, but also general references (e.g., books, reports, presentation slides). When searching with PsycINFO, we used a basic search (i.e., any field, not limited to titles and abstracts, but also contents and references) up to December 2017. The keywords “replicate AND Indonesia” generated 10 articles

and the keywords “replication AND Indonesia” generated 20 articles. The keywords used was based on Makel, Plucker, and Hegarty (2012) who also searched scientific journal articles with the word replication. It should be noted that PsycINFO is an online database that contains abstracts and expansive indexation with more than 13 million scientific writings on behavioral and mental health sciences since 1800s (APA, 2017).

Table 1 (see Appendix) shows 10 articles out of 30 articles revealed in PsycINFO search. These 10 articles were chosen using criteria as follows: (1) a replication study as confirmed by a statement in the method or other section in the article. For example, the introduction and result sections indicated a replication of a certain theoretical model or a conceptual replication from previous studies; (2) an experimental study or a survey testing a model or measure; (3) the study was conducted by Indonesian researchers or the sample was from Indonesia’s regions; and (4) a psychological or behavior-related study. There were articles indicated that the studies replicated only the findings of previous studies (particularly, a survey), but they did not replicate the methods, for example, Paez et al. (2008). There were also articles reported studies that replicated a theoretical model and used a correlational survey design, such as the studies by Astuti and Dharmmesta (2011) and Tjiptono et al. (2014). The two studies tested whether a previous theoretical model was relevant or not to be used with different samples. Meanwhile, there were other studies not related to replications that appeared in the search result because of the word replication (replicate or replication) in the articles. When this search included other criteria; that is, an experimental method and one of the authors was an Indonesian, only three articles were identified: Suhoyo et al. (2014), Hagger et al. (2016) and Wiradhany and Nieuwenstein (2017). Based on the search results, it is concluded that the number of replication studies conducted by Indonesian researchers that are published in qualified international journals are still limited. We presume that there are not many psychology researchers in Indonesia that apprehend the importance of replication studies and methodologically, they probably have not yet understood the steps to conduct a replication study.

However, if we discuss the issues of replication studies not in terms of international publications but academic related, there is an effort to increase replication studies by Indonesian scholars. For example, in the Faculty of Psychology, Universitas Indonesia

(UI), several postgraduate theses were designed as replication and replication-plus-extension studies which were basically replication studies with slightly changing in particular sections of the study (for example, by adding a new variable). In this article, the types of replication studies will be discussed in the replication type section. For now, we will explain three recent theses published in 2017. These are the examples of theses with experimental replication studies in the Faculty of Psychology of UI. Notably, the theses were obtained directly from each researcher (the three master's graduates) because searching through the online library catalogue with the keyword "replication" did not provide satisfying results. This might be due to the limitation in the search function as there is no such keyword in the text or the content of the theses.

One of the examples of replication studies is master's thesis written by Langit (2017) who conducted a pure experimental replication study adapted from Kross et al. (2014) about the use of name as a self-talk strategy during self-reflection which is considered a part of self-regulation mechanism. The result of study 1 indicated that self-talk could change the perception towards stressors in the future. In study 2, Langit (2017) conducted a replication-plus-extension study with participants who were used to use pronoun in daily activities. The result of study 2 indicated that self-talk strategy using personal pronoun or first-person pronoun ("I") in the group that were used to use personal pronoun did not generate differences in the perception of stressors. The finding in study 1 was consistent with Kross et al.'s (2014) finding, while the finding in study 2 was different because a habituation effect of using personal pronoun in daily life occurred only among participants from a certain cultural group (in this case, Minangnese culture, Padang; Langit, 2017).

The other examples are Nurifana's and Mantara's theses (2017) that reported replication-plus-extension studies (i.e., method replications) that were adopted from Hornsey and Imani (2004). Nurifana investigated *intergroup sensitivity effect* based on the source of criticism and stereotype of the target of criticism using an experimental design. Meanwhile, Mantra conducted an experimental study on *intergroup sensitivity effect* based on the source of criticism and duration of stay in a foreign country. Nurifana found that criticism from outgroup was more sensitive than ingroup. Specifically, she also found that weak group bonds determined whether criticism from outgroup addressed directly toward one's group had generated

a high level of sensitivity or not. Meanwhile, Mantara found that criticism from ex-outgroup was considered more positive than outgroup and duration of stay had no significant effect on *intergroup sensitivity*. Generally, replication studies by Langit (2017), Nurifana (2017) and Mantara (2017) had through the process, started with determining a study to be replicated and ended with designing a replication study and implementing the experiment. We will explain this in the practical step section.

Replication studies that were found in search database and those that were reported in master's theses in Universities Indonesia should be viewed positively. The scientific climate in psychology needs to be further increased and having a replication study would be beneficial. Students, lecturers, and researchers should understand that conducting a replication study can be considered as an effort to strengthen psychology with a positivistic approach. However, we should remember that in the future, we will need to develop our own theory relevant to the Indonesian context that could be different from the Western theories. To build this new theory, the foundation of science (for example, methodology) needs to be enhanced and the existing theories should be falsified, one of them using replication studies.

Challenges of Replication in Psychology

The problem with behavioral science is that as a "soft" science, it lacks a true paradigm. Studies in behavioral science are less replicative than those in natural science. This could be because participants in behavioral studies are human. Different from behavioral science, for example, natural science researchers who studied animals or other natural objects are easier to find stable patterns in their experimental condition. Human beings are equipped with mind, feelings, and other dispositional factors, and therefore, many factors might 'interfere' the findings in behavioral science studies. Individual differences and social context could interfere research in behavioral science (Kline, 2013). Since the position of psychological science determines the implementation of replication studies, it is necessary to control factors that could influence research participants in psychological studies.

Another problem is related to statistics, where Gelman and Geurts (2017) explained three examples of findings in social psychology indicating "a replication crisis". One of them is related to a statistical error in removing data, for instance, few par-

ticipants are removed from the analysis with a certain purpose, either to confirm a hypothesis or to remove outliers in order to have normally distributed data. Then, the next errors are related to mistakes in choosing the type of regression analysis, miscalculating p-values, and having a small sample size. In fact, the statistical errors could be avoided if the research is well designed. Gelman and Geurts stated that researchers should learn from their mistakes and understand that replication (either a successful or failed replication) is an important aspect of research.

Meanwhile, Stroebe and Strack (2014) argued that a replication crisis is not caused by a methodological issue or an evidentiary mechanism, but it is more related to an epistemological misunderstanding of the phenomenon of the original study. According to Stroebe and Strack, we need to first understand the differences between applied research (such as an intervention study) and basic research (such as a study for testing a theory). Stroebe and Strack asserted that when testing a theory in a basic research, we focus more on a phenomenon, but in fact, the phenomenon might not lead to a consistent finding because the operationalization of variables is not strictly done. Moreover, the study could be conducted in a different time with a different population or even using a different theoretical construct (Stroebe & Strack, 2014). Lilienfeld (2017) stated that replication crisis have made us at least more humble and avoid reporting positive results excessively.

Another issue has been explained by Makel et al. (2012) who reviewed various articles based on his search using keywords “replicat*” (this generated articles containing the word replicate). Of 100 journals with good impact factors, there were only 1.57% of publications containing the word replicate. Meanwhile, Nosek conducted a megaproject research with 269 authors. They replicated studies reported in 98 articles in three psychological journals (Baker, 2015). Only 39 out of 100 replication studies in 98 articles had similar results to the original studies or replicated well. This means that there are challenges for future psychology researchers to design and conduct replication studies that strictly follow scientific principles. The small number of replication studies that were well executed has raised another question; that is, whether the findings in previous studies are indeed consistent or they need to have additional explanations from time to time.

Meanwhile, Kline (2013) clarified that the percentage of replication studies in behavioral science is less than 1%. Bohannon (2016) revealed a surprising

finding that 60% of 100 published experiments conducted by psychology researchers in 2015 were failed to be replicated. (see *Open Science Collaboration*, 2015). A similar failure also occurred in experimental economics where more than 60% of 18 studies were failed to be replicated (Bohannon). According to Kline (2013), this condition occurs because of several reasons as follows:

Misinterpreting statistical significance. Many researchers believe that the results of studies should be concluded from the significance values. In fact, it is more important to see the results from the effect sizes. Having excessive confidence on the significance values when interpreting the results could make us assume that a replication is not important.

Many journal editors and reviewers emphasize more on novelty. Numerous journal editors and reviewers have emphasized more on the originality of studies, such as developing new theories and methodologies that substantially contribute to science. Thus, a replication is commonly viewed as repeating old ideas and those who conducted replication studies are considered not creative, dull scientists who can only imitate but have no innovation.

Other disincentives for conducting a replication. Many postgraduate programs require a thesis or dissertation to provide a contribution to science based on an original and independent research. Doctoral students are prevented to do a replication study to fulfill this requirement. This regulation has been implemented to doctoral students in Universitas Indonesia. However, in recent years, master’s students have been directed to undertake a replication study and/or a replication-plus-extension study with additional variables.

In addition to the three factors that impede the increase of replication studies as explained above by Kline (2013), there are other factors that might explain why replication studies have not run well. *Open Science Collaboration* (2012) explained that failures in conducting replication studies can occur if: (1) the effect found in the original study is false; (2) the *effect size* obtained is lower than what is reported in the original study; (3) there are faults in the design, implementation, or analysis of the original study and replication studies; and (4) the methodology of replication studies is different from the original study. Because of the four problems explained by *Open Science Collaboration*, only few researchers have conducted replication studies. The same four problems might contribute to the small number of replication studies in Indonesia as discussed in

the previous section, Literature Review of Replication Studies in Indonesia.

Aims in Conducting Replication Studies

Thomas S. Kuhn explained that science might operate in two conditions, which are (1) *normal science*, a high level of paradigm development, occurred when a paradigm develops another paradigm by extending previous research findings. This is known as *theoretical cumulativeness*; and (2) *anomalies*, a condition when a current paradigm could not resolve existing theoretical or empirical problems. This condition has challenged young researchers or researchers from different scientific backgrounds to reveal and defend the existing paradigm (Kline, 2013). Replication studies are considered to be a part of normal science as explained by Thomas S. Kuhn. This means that replication studies attempt to test or extend the findings of previous studies in order to develop normal science.

Hence, as part of a positivistic paradigm, replication studies often use a quantitative approach, particularly an experimental method, which is a common approach in natural sciences. Psychology as social science has attempted to adopt a positivistic paradigm. Therefore, many replication studies in psychology incorporate an experimental design, although this condition does not rule out the possibility that there are replication studies using a correlational survey to test theoretical concepts and measures (psychometric testing). However, it should be noted that the spirit of replication is an experimental method which is generally used in natural science but rarely used in social or behavioral science (Kline, 2013). A further explanation on the use of experimental methods and replication surveys can be seen in the replication type section.

There are five purposes to conduct replication studies according to Schmidt (2009), which are as follows.

To control sampling error (related to result changes). For example, in a population, there are four classes of samples and the original study use only three classes of samples. Using a replication, we can repeat the process by taking samples randomly from each class.

To control artifacts or initial findings (weak internal validity). For example, from the four existing classes of samples, the original study has found that a specific hypothesis is confirmed in the first class of sample. However, there is a presumption

that the result is influenced by variables in other classes. Therefore, a replication is conducted to test whether there is an error or not in the internal validity as indicated in the initial finding (the first class) by comparing the findings between the classes of samples.

To control fraud. This is similar to the second purpose concerning a specific hypothesis. This shows that a replication aims to control fraud during re-search implementation.

To generalize the result of study to a broader or different population. A replication aims to repeat the findings with a certain population; thus, it reveals the strengths (or weaknesses) of empirical findings.

To verify the hypotheses underlying previous experiments. This needs other experiments to be conducted and even with different materials. The fifth purpose is related to a conceptual replication. Various types of replications will be discussed in the next section.

Replication Types

There is no single nomenclature to classify replication studies, however, Thompson (1997) distinguishes it into internal replication and external replication. Internal Replication is a replication conducted by researchers in the original study through statistical resampling (*bootstrapping*) and cross-sample validation (Thompson). The example of internal replication is Ekman's and Heider's (1988) study replicating Ekman's study about the expression of contempt that was conducted with different samples two years after the original study. External replication is a replication conducted by other researchers (not researchers in the original study) who collected data from new samples in a different time and place (Thompson). In this article, we will only focus on external replication because in Indonesia, external replication has been used more frequently than internal replication (see the Literature Review of Replication Studies in Indonesia section). Moreover, external replication has been suggested as a starting point for Indonesian psychology researchers who want to understand, learn, and adopt a precise and standardized experimental method in order to generalize the findings.

External replication studies can be explained in two different contexts (Kline, 2013). First, replications based on experimental research. Second, replications for psychometric purposes, such as instrument evalu-

ation and adaptation. Usually, replication studies use an experimental method, but some researchers replicated a theoretical model using a correlational survey design. Kline did not state explicitly that a survey should be put under external replication. However, considering that the spirit of replication is an experimental method (particularly, it used in natural science studies), replication studies using a survey method might be a part of external replication that is related to measure and theoretical model development. In Kline's category, a survey might be placed under external replication for psychometric purposes if the survey aims to test psychometric properties of a measure.

The examples of replication studies using a survey method in psychology or behavioral science can be seen in Table 1 in the Literature Review of Replication Studies in Indonesia section, where some of them replicated a theoretical model and hypothesis, as well as tested inter-variable relationships in the original study (see Appendix). La Sorte (1972) stated that replications using a survey method in sociology are considered a series of procedures that cumulatively results in a verification and generalization of a theory. There is also a survey method that is used as a data collection method in an experimental research which is known as a survey experiment. For instance, Coppock (2016) conducted 12 replication studies based on a survey experimental design in political sciences, where data were collected through an online survey in Mechanical Turk, Amazon. In economics, surveys are commonly used for testing hypothesis in replication studies. For example, a systematic review by Nordhaus and Moffat (2017) reported 27 survey studies investigating the impacts of climate changes on economic condition. It seems that the methodology used in social sciences in general (psychology, sociology, politics, and economics) that involve human being as participants allows the use of a survey method in replication studies aiming to test a theoretical model, hypothesis, and inter-variable relationships in the original study. Both survey and experimental methods, if they were conducted precisely and responsibly following the principle of "good science", could be included in the design of replication studies to attain the purpose of the study.

Specifically, this article will discuss replication studies based on an experimental method. Therefore, Kline (2013) explained that in the first context (experiments), replication studies are divided into:

Exact replication (direct replication, literal replication, or precise replication). All the main

aspects of the original study (a sampling method, design, and measurement) are exactly repeated. The replication is difficult to implement, particularly because human factors (participants or researchers), time, settings, and samples to some extent can influence the process.

Operational replication (partial replication or improvisational replication). Only sampling method and methods (but not measurement) in the original study are repeated. To obtain the results, researchers follow the basic "recipe" in the method section of the original study. The replication provides more informative results compared to the first type of replication because variations in procedures, settings, or samples have generated a strong effect.

Balanced replication. Operational replication is used as a control condition, whereas the other condition represents a manipulation of an additional important variable to test a new hypothesis. In an experiment, a control condition is related to the condition in the original study, whereas an additional condition is the condition that is slightly different from the condition in the original study. This is what is called as Balanced Replication.

Construct replication (conceptual replication). The study avoids an exact imitation with the specific method used in the original study. Thus, the original method is not imitated. Researchers might specify design, measurement, and data analysis which ever considered appropriate in order to compare the findings with the original study. To obtain those specifications, a systematic review is needed in order that researchers could find different characteristics of replications.

Different from Kline's category, other researchers distinguish two types of replications, which are: direct replication and conceptual replication (Schmidt, 2009; Makel et al., 2012). Direct replication is a repetition of experimental procedures, whilst conceptual replication is a repetition of hypothesis testing or retesting results with different methods (Schmidt, 2009). It was found that conceptual replication was more successful than direct replication although the rate of success was not statistically significant (Makel et al. According to Makel et al., the result seems to be in contrast to the intuition, why is conceptual replication more successful than direct replication? There is a possibility that failed results of conceptual replication studies tend not to be submitted to or not accepted by scientific publications. Makel et al. emphasize that conceptual replication does not necessarily meet all the purposes of replication because it highly

depends on experimenter limitations and measurement errors.

Aside from the types of replications above, there is also *replication-extension* or replication-plus-extension, that is, a combination and comparison of the results between one or more original studies and the new study (Bonett, 2012). In practice, a replication-plus-extension study includes other variables that presumably influence the findings of the new study. Replication-plus-extension studies with additional variables have been commonly conducted by postgraduate students in the Faculty of Psychology, Universitas Indonesia as a part of their research thesis. Few studies are explained in the previous section, Literature Review of Replication Studies in Indonesia. As we have understood the urgency and conceptual explanation of replication studies, We will discuss the practical steps to conduct a replication study. In short, the steps in conducting a replication study is started with determining an original study, understanding the original study, writing a summary, and is ended with designing a replication study according to certain principles. The next section illustrates the practical steps to determine articles reporting the original study that will be replicated.

Determining Journal Articles as References for a Replication Study

Researchers should be able to accurately and strictly select journal articles reporting the original study that will be replicated. This implies that certain criteria should be used to select appropriate journal articles. *Open Science Collaboration* (2012) suggested three reputable journals to find the appropriate study for replication: *Journal of Experimental Psychology: Learning, Memory, and Cognition* (JEP: LMC), *Journal of Personality and Social Psychology* (JPSP), and *Psychological Science* (PSCI). The three widely-known journals present experimental studies. JEP and LMC focus on cognitive studies, while JPSP focuses on social studies and PSCI contains research on general psychology topics (*Open Science Collaboration*, 2015). To determine the most appropriate articles, a research team may start with choosing 30 articles published in a certain year. *Open Science Collaboration* (2012, 2015) has suggested starting the search from the year of 2008, for example, by listing 30 articles published in *Psychological Science* in 2017. It is important to note that a sample of 30 articles might be considered inadequate. Kar and Ramalingam (2013) explained that 30 is considered

as the magic number for sample size estimation. The same issue might occur with replication studies, therefore, we encourage researchers to select and read more than 30 articles in order to widen their knowledge and create opportunities to further develop their study design. The next step is to select articles with topics that are similar to researchers' interest. Another important step is to identify a key finding from one study to be replicated. Notably, an article can have more than one study and not all articles have studies that can be replicated since a study might include a specific sample, depend on historical events, and have limited information on measurement (*Open Science Collaboration*, 2012).

A reference article containing a study that will be replicated should be an article from a reputable journal. Why should the article come from a reputable international journal? It is because an article from a reputable journal has passed a rigorous selection process, for example, double-blind peer view or a review process where the identity of reviewers and author(s) are mutually disclosed. Studies reported in an Indonesian journal are also possible to be replicated, provided that the article comes from a journal that implements a rigorous peer review process, such as a national accredited journal. After the initial step to determine the original study that will be replicated, the next important step is to review, understand, and summarize the original study in order to design a good replication study.

Writing a Review and Summary of the Original Study

In addition to have a good understanding on experimental and statistical methods, it is important that researchers fully understand the article that contains the original study. To facilitate this, there are several points that need to be identified, which are as follows.

Referencing articles. Writing citation in APA format is important. By doing this, researchers will not only be familiar with reference writing, but they will also know the quality of the cited journal, whether it is from a reputable journal **or not**.

Research assumptions. Researchers should fully understand the assumptions behind their research that lead to research hypotheses. Researchers need to review key theories underlying research assumptions and thus, enable them to be identified an easily understood.

Variables. Identifying the variables used in the original study will facilitate researchers to understand

the psychological constructs, materials, and measurement. Generally, in an experimental study, there is a dependent variable that can be measured by one (a series of) questionnaire or behavior observation. Moreover, the most important feature in an experiment is the manipulation or treatment of an independent variable.

Methodology. This part explains the criteria of participants, types of methods (laboratory experiments, field study, or online), measures, materials (experimental manipulation), research procedures, and statistical tests.

Results and findings. Researchers understand the results and key findings of the original study. The results are presented in statistical notations, implying that the basic understanding of statistics is needed to identify research findings.

Study Limitations. Study limitations can be used as a consideration to conduct a replication study or a further study with an addition of other variables, using different methods, etc.

Researchers usually implement the six points above. Some of the points have been applied by doctoral students in the Universitas Indonesia for writing a literature review. However, it should be noted that the six points above are actually optional. The researchers can also add the relevance of each reviewing manuscript to her/his research design. Despite this, we think that the steps to write a review and a summary of the original study are important for designing a replication study. The important aspect in designing a replication study is to clearly understand the original study. Researchers are also suggested summarizing not only the original study, but also other articles related to the original study. If researchers have a “bank” (supply) of the summary of articles related to the original study, this will broaden the framework of research topic and therefore, facilitate researchers to discuss the findings and discover a theoretical gap in the findings. In the next section, we will discuss tips to design a replication study as a continuation of reviewing and summarizing the original study.

Designing a Replication Study

Brandt et al. (2014) have written in details “a replication recipe” to conduct a pure replication study which is known as a *close/direct/exact replication*, a study that closely adopts the methods and procedures of the original study. I argue that the five “ingredients” below could be used with different

types of replication study. Therefore, in designing a replication study, one should consider the points below. Brandt et al. suggested the five ingredients, we re-write the explanation of Brandt et al., which are as follows.

Being careful in defining the effects (influences) and methods that will be replicated. For instance, when conducting an experiment using a 2x2 within subject design (i.e., one group is given a series of treatments), it is possible that a positive effect will be found in one condition and a negative effect will be found in another condition. Not all study can be replicated, there is theoretical and practical considerations to determine if a replication study is needed.

Following the methods as closely as possible to the original study (including participant recruitment, instructions, materials, measures, procedures, and analysis). Researchers should start with contacting the authors of the original study. If it is difficult, then the researchers should create methods and/or measures that resemble those in the original study. In some cases, it is necessary to consider if the measurement in the original study is relevant to be used in the replication study. It should also be assessed if the measurement in the original study could be used in a different time and place and did not depend on the historical and cultural contexts. For example, if the original study focuses on the 2012 General Election and the study is replicated for the 2017 General Election, it would be necessary to adjust the measures of the original study. It is also possible to collaborate with the authors of the original study. In either case, a replication study should be conducted as closely as possible with the original study, reducing any differences with the original study.

Having high statistical power. Statistical power is the power of testing, that is, the probability to accept a false null hypothesis. For example, when a hypothesis suggests that there is no difference in X, in reality, there is no different effect on X. On the contrary, when a hypothesis suggests that there is a difference in X, the conclusion will be accurate if in reality, there is a statistical difference in X. There are various ways to calculate statistical power, such as using G-Power analysis. *Open Science Collaboration* (2012) explained that the effect size for a replication based on statistical power analysis should be more than .80 (a large effect size of Cohen’s D; $d= 0.2$ for small effect size, $d= 0.5$ for medium effect size, $d= 0.8$ for large effect size; Gravetter & Wallnau, 2012). A large effect size is useful to strengthen previous findings. If the effect size is small or not proven to

be influential, it can still contribute to the theoretical explanation providing that the study has been done rigorously, particularly if it implements the first and second point above. Determining statistical power is related to sample size or the number of samples that will be involved in a study.

Providing details of a replication study to enable interested researchers to evaluate the replication efforts (or to implement the replication). Researchers of a replication study should provide details of the methods, analysis, and results of their study to enable reviewers, editors, and readers to review the study. Moreover, researchers need to make a comparison between the original and replication study, and plan to analyze and evaluate the replication efforts. For the purpose of evaluation and explanation of the studied effects, it is necessary to present data, syntax of statistical analyses (for example, syntax in SPSS), and all other analyses.

Evaluating the replication results and comparing them critically with the results of the original study.

It is necessary to evaluate the effect sizes and Confidence Intervals (CIs) in the replication study, and compare them with those in the original study. Evaluation of a replication study is related to two points, which are (a) the size, direction, and confidence interval of the effect, particularly whether the replication effect is significantly different or not from the hypothesis; (b) an additional testing evaluating if there is a significant difference between the results of the replication study and those of the original study. Afterwards, researchers might explain if the result of a replication is (a) successful, implying that it is different from the null hypothesis (confirming that there is a difference), and if the result is similar to or even larger than the result of the original study and shares the same direction; (b) showing an informative replication failure, indicating that the result is similar to the null hypothesis (confirming that there is no difference) or it is in the opposite direction with and significantly different from the result of the original study; (c) showing a practical replication failure, the result is significantly different from the null hypothesis (confirming that there is a difference) and from the result of the original study; (d) inconclusive, it is not significantly different from the null hypothesis (confirming that there is no difference) and from the result of the original study (theoretically or in reality, there should be a difference). Researchers should realize that there are many factors that could influence the final result of a replication study, such as theoretical/conceptual and methodological aspects.

Based on the five ingredients, Brandt et al. (2014) presented 36 questions related to the design of a replication study. The first one, which is formulated by question 1 to 9, is related to being careful in defining the influence (effect) and methods used in a study. The other ingredients are scattered in and explained by other questions (Brandt et al.). In general, the questions can be categorized into questions related to the preparation stage (question 1 to 29) and result or post replication stage (question 30 to 36). Researchers can use this as a guideline to design their study and disseminate the findings. Further, the questions can also assist readers and reviewers to evaluate a replication study. The 36 questions, which are originally from Brandt et al (open access), can be seen in Table 2 (see Appendix). Brandt et al. suggested reporting the findings of a replication study online and uploading the report in psychfiledrawer.org, or open-science-framework.org. This online report will facilitate other researchers to evaluate the results of a replication study.

Proposed Solutions and Directions to Develop Psychological Replication Studies in Indonesia

The efforts to strengthen social science like natural science have encountered various issues and challenges. In the early section of this article, the conditions of psychological research in Indonesia and in the world have been explained. Several issues in psychological research are as follows: a lack of popularity to report negative results (unconfirmed hypotheses; see Lilienfeld, 2017); difficulties or failures to find evidence in several replication studies, meaning that the validity and accuracy of the results of studies need to be questioned (see Yong, 2012); inadequate reports of the results of replication studies (see Makel et al., 2012), and a lack of replication studies in Indonesia as evidenced by a limited number studies found in the world's psychological research database.

As there have been various issues and challenges in developing psychological research, several recommendations are suggested in order to make the research easier to be replicated. Asendorpt et al. (2013) suggested that in order to increase the replicability of psychological studies, researchers need to reduce the source of error in their study design and data analysis. The source of error can be minimized in several ways (Asendorpt et al., 2013): increasing a sample size; increasing the reliability of measures;

increasing the sensitivity of study design by reducing systematic errors in the instructions, questionnaire administration, and manipulation used; increasing the use of statistical analysis according to the study design; conducting a number of study effectively and avoiding the use of many underpowered studies to achieve significant results; and working with more than one statistical analysis particularly for a study that involves several variables, for instance, not only using a *post hoc* Benferroni, but also using other type of analysis, such as a non-statistical (multiple test) solution.

Asendorpt et al. (2013) also suggested several recommendations for authors, journal managers, and lecturers in Research Methods and Statistics in order to resolve the issues of replicability related to the publication process. For authors, it is important to increase research transparency which can be done by presenting a comprehensive literature review and reporting a justification for sample size (Asendorpt et al.). The authors are also expected to accelerate scientific progress, for instance by publishing their studies, replicating their own studies, and taking part in online scientific discussion forums. For journal managers, including reviewers and editors, they are expected to implement a good research practice, for example, by not only accepting articles with positive results (Asendorpt et al.). Journal managers are also expected to proactively maintain the journal quality by allowing reviewers to discuss the manuscript openly with the authors, including access their raw data if necessary. For lecturers in Research Methods and Statistics (Asendorpt et al.), they are expected to introduce and consolidate their statistical material to increase the understanding of replication concepts, to encourage a critical thinking and to allow a hypothesis that rejects the findings, and to establish a scientific culture that is more to "getting it right" than "getting it published".

Gelman and Geurts (2017) explained briefly three solutions to resolve the issues and crisis of replication studies in psychology, which are as follows.

Scientific communication. This solution is related to reporting and publishing the result of replication studies. It is best not to limit the publication only to replication studies with "statistically significant results", but also to those with positive and negative findings. Next, establishing a scientific communication can be done by having a collaborative research with other researchers who obtain conflicting or inconsistent results with our study. To increase the number of replication studies, we need to facilitate the understanding of other researchers to our study

by providing them with a detailed description of the methods used in the study to make it easier to be replicated.

Designing and collecting data. When designing a study, it is best to preregister the study in the website or scientific institutions that organize replication studies. It is also important to design a study by estimating the effect size (not only based on the significance or p value) and paying attention to the accuracy of measurement when collecting data. Researchers need to ensure that their study is well designed and resemble the original study.

Data analysis. Researchers should carefully consider the statistical analysis used in their study. Researchers can use Bayesian inferential analysis and multiple hierarchical modeling, and explore the potential factors that may influence the findings. Researcher can also conduct a meta-analysis and control *error* rates.

Based on the explanation of replication study issues in Indonesia and the recommendations to resolve the issues as suggested by Asendorpt et al. (2013) and Gelman and Geurts (2017), it can be concluded that the essential thing to support the increase of replication studies and to strengthen the findings of replication studies in Indonesia is by increasing the understanding of research methods and basic statistical methods. Psychology lecturers in Indonesia should be open to information and new understanding of replication studies as an effort to strengthen the results of study and theoretical concepts. Replication studies are expected to have clear procedures and guidelines for implementation. Similarly, during data analysis, it is expected that a replication study select an accurate statistical analysis and perform replications effectively and rigorously with sufficient number of studies or trials. Undergraduate students should be encouraged to use replication studies for their thesis, whereas postgraduate students (master's degree students) are suggested to conduct replication-plus-extension studies. The important point is to enhance the foundation of science and increase the number of replication studies.

Notably, it is still important to appreciate the efforts to develop findings and theories with a strong cultural base, particularly because of Indonesia's demographic condition that consists of diverse ethnicities and individual characteristics. The research efforts might generate "unique" findings that are suitable to our culture. In our opinion, it is legitimate to conduct studies using a different approach, such as an indigenous approach, a cross-cultural approach, and pro-

bably a qualitative approach. However, we believe that in the future, the findings from a certain theory (including the theory with a strong cultural base) will clash with those from a different theory and each theory will seek for evidence to confirm their findings. During this process, studies using a quantitative method, particularly replication studies, might be needed.

Limitations

Notably, the source of data used in this article deserves a serious attention. There is only a limited number of international publications, particularly in psychology, that report replication studies in Indonesia (particularly, those with an experimental design). This article presents the result of search findings using PsycINFO, the only database used to search scientific articles written by Indonesian researchers. It might be valuable to include different databases, or for a comparison, to use the same database to search replication studies conducted by other researchers from different countries. This could be an opportunity, as well as a direction for future researchers who are interested in investigating the trend of replication studies in Indonesia.

Conclusion and Suggestions

There is no necessity to follow the path of positivist paradigm, but the effort to construct new theories which are different from Western theories requires a robust methodological foundation. Particularly in Indonesia, cultural diversity might explain differences or inconsistencies in the findings of previous studies. Therefore, one way to confirm and extend these findings is by conducting a replication study. A replication study focuses on repeating the process and findings in the previous study using an experimental method with specific procedures. It is important to develop replication studies because there are several issues in psychology that need to resolve, for example, a failure to repeat previous research findings; this issue has made the psychological theory behind the findings is questionable. In short, an initial step to increase the number of replication studies in Indonesia is to include replication study as a part of lecture in Research Methods and Statistics. The next steps are to encourage undergraduate students to use a replication study in their thesis and to encourage psychology researchers in Indonesia to conduct replication studies. After the culture of “good science” has been esta-

blished, the results of replication studies should be written in a standardized format of writing for scientific journals. Thus, the support from journal editors and reviewers are needed to increase the number of publications of replication studies.

Despite the above limitations, it is expected that the editors of Indonesian journals are open to, or even explicitly invite the paper using replication methods. This can be started by having an explicit statement concerning the results of study using replication methods in the aim and scope section of the journal. In addition, in the future, it might be necessary to hold scientific conferences and deliberately invite scientists to replicate the studies published in Indonesian journals. It is expected that these efforts will increase the number of scientific articles reporting replication studies conducted by Indonesian authors and/or replication studies conducted in Indonesia.

References

- Ackermann, R. (2009). Popper. In J. Kim, E. Sosa, G. R. Rosenkrantz (Eds.), *A companion to metaphysics, second edition*. Blackwell Publishing, Singapore.
- Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior* (Vol. 1980). Englewood Cliffs, NJ: Prentice-Hall.
- American Psychological Association (APA). (2017). *APA databases: Creating value for libraries*. Retrieved from <http://www.apa.org/pubs/marketing/value.aspx>
- Archer, D., & Gartner, R. (1984). *Violence and crime in cross-national perspective*. New Haven, CT: Yale University Press.
- Arli, D., Dietrich, T., Tkaczynski, A., & Rundle-Thiele, S. (2017). Indonesian healthy living intentions: Segmentation study insights. *International Journal of Nonprofit and Voluntary Sector Marketing*, 22(2), 1-9. <https://doi.org/10.1002/nvsm.1574>
- Asendorp, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J. A., Fiedler, K., ... Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, 27, 108-119. <https://doi.org/10.1002/per.1919>
- Astuti, P. H., Nusantara, J., & Dharmmesta, B. S. (2011). The evaluation of customer complaint handling with justice dimensions: Effect on trust and commitment with prior experiences as moderating effect. *Interdisciplinary Journal of Contemporary Research in Business*, 2(11), 228-237. <https://doi.org/10.1002/nvsm.1574>

- Allen, M., & Preiss, R. (1993). Replication and meta-analysis: A necessary connection. *Journal of Social Behavior and Personality*, 8(6), 9-20. Retrieved from <https://search.proquest.com/openview/789bbad11869a66830b5a3fcefabd39b/1?pq-origsite=gsc-ho-lar&cbl=1819046>
- Alvard, M. (2003). Kinship, lineage, and an evolutionary perspective on cooperative hunting groups in Indonesia. *Human Nature and Interdisciplinary Biosocial Perspective*, 14(2), 129-163. <https://doi.org/10.1007/s12110-003-1001-5>
- Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., ... Van't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, 50, 217-224. <https://doi.org/10.1016/j.jesp.2013.10.005>
- Baker, M. (2015). Over half of psychology studies fail reproducibility test. *Nature News*, 27. Retrieved from <https://www.nature.com/news/over-half-of-psychology-studies-fail-reproducibility-test1.18248>
- Bem, S., & de Jong, H. L. (2006). *Theoretical issues in psychology: An introduction* (second edition). Sage Publication, London.
- Bem, D. J. (2011). Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100(3), 407-425. <https://doi.org/10.1037/a0021524>
- Bohannon, J. (2016). About 40% of economics experiments fail replication survey. *Science*. Retrieved from <http://www.sciencemag.org/news/2016/03/about-40-economics-experiments-fail-replication-survey>
- Bonett, D. G. (2012). Replication-extension studies. *Current Directions in Psychological Science*, 21(6), 409-412. <https://doi.org/10.1177/0963721412459512>
- Colella-Sandercock, J. A. (2017). Self-reporting in plagiarism research: How honest is this approach? *Journal of Research Practice*, 12(2), 3. Retrieved from <http://jrp.icaap.org/index.php/jrp/article/view/558/456>
- Coppock, A. (2016). *Generalizing from survey experiments conducted on Mechanical Turk: A replication approach*. In annual meeting of the Midwest Political Science Association.
- Ekman, P., & Friesen, W. V. (1986). A new pan-cultural facial expression of emotion. *Motivation and Emotion*, 10(2), 159-168. <https://doi.org/10.1007/BF00992253>
- Ekman, P., & Heider, K. G. (1988). The universality of a contempt expression: A replication. *Motivation and Emotion*, 12(3), 303-308. <https://doi.org/10.1007/BF00993116>
- Fanelli, D. (2011). Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90(3), 891-904. <https://doi.org/10.1007/s11192-011-0494-7>
- Gebauer, J. E., Sedikides, C., Schönbrodt, F. D., Bleidorn, W., Rentfrow, P. J., Potter, J., & Gosling, S. D. (2016). The religiosity as social value hypothesis: A multi-method replication and extension across 65 countries and three levels of spatial aggregation. *Journal of Personality and Social Psychology*, 113(3), e18-e39. <https://doi.org/10.1037/pspp0000104>
- Gebauer, J. E., Leary, M. R., & Neberich, W. (2012). Big two personality and big three mate preferences: Similarity attracts, but country-level mate preferences crucially matter. *Personality and Social Psychology Bulletin*, 38(12), 1579-1593. <https://doi.org/10.1177/0146167212456300>
- Gelman, A., & Geurts, H. M. (2017). The statistical crisis in science: how is it relevant to clinical neuropsychology? *The Clinical Neuropsychologist*, 31, 1000-1014. <https://doi.org/10.1080/13854046.2016.1277557>
- Gravetter, F. J., & Wallnau, L. B. (2012). *Statistics for the behavioral sciences* (9th Edition). Cengage Learning, USA.
- Hagger, M. S., Chatzisarantis, N. L., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., ... & Calvillo, D. P. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science*, 11(4), 546-573. <https://doi.org/10.1177/1745691616652873>
- Hornsey, M. J., & Imani, A. (2004). Criticizing groups from the inside and the outside: An identity perspective on the intergroup sensitivity effect. *Personality and Social Psychology Bulletin*, 30(3), 365-383. <https://doi.org/10.1177/0146167203261295>
- Kar, S. S., & Ramalingam, A. (2013). Is 30 the magic number? Issues in sample size estimation. *National Journal of Community Medicine*, 4(1), 175-179. Retrieved from http://www.njcmindia.org/uploads/4-1_175-179.pdf
- Kline, R. B. (2013). *Beyond significance testing: Statistics reform in the behavioral sciences* (2nd ed.). American Psychological Association, Washington DC.
- Kross, E., Bruehlman-Senecal, E., Park, J., Burson, A., Dougherty, A., Shablack, H., ... Ayduk, O. (2014). Self-talk as a regulatory mechanism: how you do it matters. *Journal of Personality and Social*

- Psychology*, 106(2), 304-324. <https://doi.org/10.1037/a0035173>
- La Sorte, M. A. (1972). Replication as a verification technique in survey research: A paradigm. *The Sociological Quarterly*, 13(2), 218-227. <https://doi.org/10.1111/j.1533-8525.1972.tb00805.x>
- Langit, U. S. (2017). *Ubah ancaman menjadi tantangan: Strategi berbicara kepada diri sendiri sebagai mekanisme regulasi stres di masa depan pada individu yang terbiasa menyebut diri menggunakan nama (Turning threats into challenges: Self-talk as a regulatory mechanism of future stressor in individuals using name as self-perspective)*. Unpublished Master Thesis. Faculty of Psychology, Universitas Indonesia, Depok, Indonesia.
- LeBel, E. P., & Peters, K. R. (2011). Fearing the future of empirical psychology: Bem's (2011) evidence of psi as a case study of deficiencies in modal research practice. *Review of General Psychology*, 15(4), 371-379. <https://doi.org/10.1037/a0025172>
- LeBel, E. P., Berger, D. M., Campbell, L., & Loving, T. (2017). Falsifiability is not optional. *Journal of Personality and Social Psychology*, 113(2), 254-261. <https://doi.org/10.1037/pspi0000106>
- Lilienfeld, S. O. (2017). Psychology's Replication Crisis and the Grant Culture: Righting the Ship. *Perspectives on psychological science*, 12(4), 660-664. <https://doi.org/10.1177/17456916166877>
- Machamer, P. (2002). A brief historical introduction to the philosophy of science. In P. Machamer & M. Silberstein (Eds), *The Blackwell Guide to the Philosophy of Science*. Blackwell Publisher, UK.
- Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research how often do they really occur? *Perspectives on Psychological Science*, 7(6), 537-542. <https://doi.org/10.1177/1745691612460688>
- Matosin, N., Frank, E., Engel, M., Lum, J. S., & Newell, K. A. (2014). Negativity towards negative results: a discussion of the disconnect between scientific worth and scientific culture. *Disease Models & Mechanisms*, 7(2), 171-173. <https://doi.org/10.1242/dmm.015123>
- Mantara, A. Y. (2017). *Pengaruh status keanggotaan kelompok sumber kritik dan durasi tinggal terhadap intergroup sensitivity effect (Effect of the source of critics' group status and duration of stay on intergroup sensitivity effect)*. Unpublished Master Thesis. Faculty of Psychology, Universitas Indonesia, Depok, Indonesia.
- Martin, G. N., & Clarke, R. M. (2017). Are psychology journals anti-replication? A snapshot of editorial practices. *Frontiers in Psychology*, 8, 523. <https://doi.org/10.3389/fpsyg.2017.00523>
- Nolin, D. A. (2011). Kin preference and partner choice. *Human Nature*, 22(1-2), 156-176. <https://doi.org/10.1007/s12110-011-9113-9>
- Nurfiana, A. (2017). *Pengaruh sumber kritik dan target etnis stereotip terhadap intergroup sensitivity effect (The impact of source of critic and target ethnic stereotypes to intergroup sensitivity effect)*. Unpublished Master Thesis. Faculty of Psychology, Universitas Indonesia, Depok, Indonesia.
- Proctor, R. W., & Capaldi, E. J. (2008). *Why science matters: Understanding the methods of psychological research*. Blackwell Publishing, UK.
- Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, 7(6), 657-660. <https://doi.org/10.1177/1745691612462588>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), 1-8. <https://doi.org/10.1126/science.124716>
- Ophir, E., Nass, C., & Wagner, A. D. (2009). Cognitive control in media multitaskers. *Proceedings of the National Academy of Sciences of the United States of America*, 106(37), 15583-15587. <https://doi.org/10.1073/pnas.0903620106>
- Paez, D., Liu, J. H., Techio, E., Slawuta, P., Zlobina, A., & Cabecinhas, R. (2008). "Remembering" World War II and willingness to fight: Sociocultural factors in the social representation of historical warfare across 22 societies. *Journal of Cross-Cultural Psychology*, 39(4), 373-380. <https://doi.org/10.1177/0022022108316638>
- Schuster, L., Kubacki, K., & Rundle-Thiele, S. R. (2015). A theoretical approach to segmenting children's walking behaviour. *Young Consumers*, 16(2), 159-171. <https://doi.org/10.1108/YC-07-2014-00461>
- Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, 13(2), 90-100. <https://doi.org/10.1037/a0015108>
- Smith, N. C. (1970). Replication studies: A neglected aspect of psychological research. *American Psychologist*, 25(10), 970-975. <https://doi.org/10.1037/h0029774>
- Sprecher, S. (2013). Attachment style and sexual permissiveness: The moderating role of gender. *Personality and Individual Differences*, 55(4), 428-432.

- <https://doi.org/10.1016/j.paid.2013.04.005>
- Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, 9(1), 59-71. <https://doi.org/10.1177/1745691613514450>
- Tax, S. S., Brown, S. W., & Chandrashekar, M. (1998). Customer evaluations of service complaint experiences: implications for relationship marketing. *The Journal of Marketing*, 62(2), 60-76. <https://doi.org/10.2307/1252161>
- Thompson, B. (1997). Editorial policies regarding statistical significance tests: further comments. *Educational Researcher*, 26(5), 29-32. <https://doi.org/10.3102/0013189X026005029>
- Tjiptono, F., Arli, D., & Bucic, T. (2014). Consumer confusion proneness: insights from a developing economy. *Marketing Intelligence & Planning*, 32(6), 722-734. <https://doi.org/10.1108/MIP-05-2013-0082>
- Van Hell, E. A., Kuks, J. B. M., Raat, A. N., van Lohuizen, M. T., & Cohen-Schotanus, J. (2009). Instructiveness of feedback during clerkships: Influence of supervisor, observation and student initiative. *Med Teach*, 31(1), 45-50. <https://doi.org/10.1080/01421590802144294>
- Walsh, G., & Mitchell, V. W. (2010). The effect of consumer confusion proneness on word of mouth, trust, and customer satisfaction. *European Journal of Marketing*, 44(6), 838-859. <https://doi.org/10.1108/03090561011032739>
- Winerman, L. (2016). How much of the psychology literature is wrong? *Monitor on Psychology*, 47(6), 14-15. Retrieved from <http://www.apa.org/monitor/2016/06/psychology-literature.aspx>
- Wiradhany, W., & Nieuwenstein, M. R. (2017). Cognitive control in media multitaskers: Two replication studies and a meta-analysis. *Attention, Perception, & Psychophysics*, 79(8), 2620-2641. <https://doi.org/10.3758/s13414-017-1408-4>
- Yeung, A. W. (2017). Do neuroscience journals accept replications? A survey of literature. *Frontiers in Human Neuroscience*, 11, 468. <https://doi.org/10.3389/fnhum.2017.00468>
- Yong, E. (2012). Bad copy: In the wake of high-profile controversies, psychologists are facing up to problems with replication. *Nature*, 485, 298-300. <https://doi.org/10.1038/485298a>
- Yusainy, C. (2015). Quo vadis psikologi sebagai sebuah kajian ilmiah? *Buletin Psikologi*, 23(1), 51-56. <https://doi.org/10.22146/bps.10577>

(Appendix follows)

Appendix

Table 1

Summary of Articles in Behavioral Science Reporting Replication Studies in Indonesia

No	Authors (Year)	Types of Study	Original Study: Authors (Year)	Study Descriptions and Brief Results
1	Ekman & Heider (1988)	Replication: Experiment:	Ekman & Friesen (1986)	Non-Indonesian authors. Indonesian Sample. Result: Minangnese (participants) indicated that their emotional expression was similar to Japanese and Americans.
2	Paez et al. (2008)	Conceptual Replication: Survey	Archer & Gartner (1984)	Non-Indonesian authors. Indonesian Sample. Result: the legitimacy of violence through war (World War II) was dominated by collective memory.
3	Astuti & Dharmmesta (2011)	Theoretical replication: Survey	Tax et al. (1998)	Indonesian authors. Indonesian Sample. A study on customer attitude (complaints & inconvenience) and social justice.
4	Tjiptono, Arli & Bucic (2014)	Theoretical replication: Survey	Walsh & Mitchell (2010)	Indonesian authors. Indonesian sample. Result: there were three dimensions of consumer confusion proneness (confusion of young consumers over smartphone): similarities, overload, and ambiguity.
5	Suhoyo et al. (2014)	Replication: Experiment	Van Hell et al. (2009)	Indonesian authors. Indonesian sample. Result: similar to the result of a replication study in the Netherlands, Indonesian students assumed that feedback was more constructive if it was given by an expertise and initiated by their superiors and the students themselves. Dutch students appreciated feedback if it was based on an observation.
6	Schmitt & Jonason (2015)	Replication: Survey	Sprecher (2013)	Non-Indonesian authors. A part of data collected from Indonesian sample. Result: yielding replication findings of the previous study indicating that there was a relationship between attachment and permissive sexuality (a permissive sexual position), and expanding the findings to 10 country regions across the globe.
7	Gebauer et al. (2016)	Replication: Survey	Gebauer et al. (2012)	Non-Indonesian authors. A part of data was collected from Indonesian sample. Result: three studies replicated the findings with respect to religiosity as social value (RAVS) based on cross-country data. It was confirmed that in a religious country, religiosity contained many social values.
8	Hagger et al. (2016)	Replication: Multilab Experiment	Sripada, et al. (2014)	A research collaboration with an Indonesian scholar. Indonesian sample. Result: the study indicated a very small ego-depletion effect based on several experimental replications in the lab.
9	Arli et al. (2017)	Replication theory: Survey	Ajzen and Fishbein (1980); Schuster et al. (2015)	Indonesian authors. Indonesian sample. A study on healthy life intention. Result: a segmentation analysis with different groups in a vast population and group consideration could help to provide services /interventions that met the needs of a community. This could further expand the implementation of a campaign concerning obesity.
10	Wiradhany & Nieuwenstein (2017)	Replication: Experiment	Ophir et al. (2009)	An Indonesian author. Result: two experiments and a meta-analysis showed that there was an interaction between multitasking media and distractibility; this was based on the lab result using information processing tasks and a meta-analysis review.

Table 2

A Recipe to Conduct a Replication Study Using 36 Questions as a Guideline (Brandt et al., 2014; p. 219)

The nature of effect (influence)

1. Description of the effect in the original study that will be replicated:
2. It is important to replicate this effect because:
3. *Effect size* or the magnitude of effects in the original study (for example, Cohen’s *d*, η^2 or *eta square*, etc.) is:
4. Confidence Interval (CI) of the effect in the original study is:
5. Sample size in the original study that generates the effect:
6. Where was the original study conducted? (for example, in a laboratory, field, *online*).
7. In what country/state was the original study conducted?
8. What the type of sample did the original study use? (for example, students, Mturk, representative sample).....
9. Did the original study use a paper-and-pencil survey, computerized measure, or other types of measures?

Replication Study Design

10. Is the original measure used in the study available from the author?
 - a. If not, is the original measure available from other sources (for example, the scale is published in the previous article)?
 - b. If the original measure is unavailable from researcher or other sources, then how was the measure created for replication?
11. The author understands the assumption (for instance, the meaning of stimuli) behind the original study that will be maintained in the replication study because:
12. Location of experimenter during data collection:.....
13. Experimenter’s knowledge about the condition of participants:.....
14. Experimenter’s knowledge about the hypothesis in general:
15. The author’s target concerning a sample size in the replication study:.....
16. The rationale in determining the sample size is:.....

Documenting Differences Between the Original and Replication Study

For each part of the replication study below, please indicate the type of replication study (exact, close, or conceptually different) compared to the original study.

17. The similarity/difference in the instruction is:
- (Exact/Close/Different)
18. The similarity/difference in the measure is:
- (Exact/Close/Different)
19. The similarity/difference in the stimulus is:
- (Exact/Close/Different)
20. The similarity/difference in the procedure is:
- (Exact/Close/Different)
21. The similarity/difference in the location (laboratory versus *online*, individual versus group) is:
- (Exact/Close/Different)
22. The similarity/difference in remuneration is: (Exact/Close/Different)
23. The similarity/difference in the participants is: (Exact/Close/Different)
24. Will the differences between the original and replication study influence the size and/or direction of the effect? ...
25. The steps that the author will take if the differences stated in number 24 influence the results of the replication study:.....

(appendix continues)

 Evaluation on the Analysis and Replication

26. The exclusion criteria used (for example, managing *outlier*, removing participants from the analysis):
27. The analysis that will be used is (justify if it is different from the original study):
28. Success in replicating is defined as:

 Replication Registration

29. Measurement, procedures, analysis plan, etc. of the replication study have been registered in (indicate the name of website or institution):

 Reporting a Replication Study

30. *Effect size* of replication is:
31. Confidence interval of the *effect size* is:
32. Is the *effect size* of the replication study significantly different from the effect of the original study (yes/no)?.....
33. The author considers that the replication is (succeed/having an informative failure /having a practical failure /inconclusive), because:
34. Interested experts can obtain data or syntax analysis in:
35. All reported analysis are available in:
36. The limitation of the replication study is:

Note. Source: Adapted from *Journal of Experimental Social Psychology* (2014), 50(1), 217-24. Copyright © 2013 by the Authors. Published by Elsevier Inc. Open-access article.