Science is about Corroborating Empirical Evidence, even in Academic Business Research Journals

Barry J. Babin[*]

Morris Lewis Professor of Marketing

University of Mississippi,

University, MS 38677

USA

bbabin@bus.olemiss.edu

David J. Ortinau

University of South Florida,

Tampa, FL

USA,

dortinau@usf.edu

Jean-Luc Herrmann

University of Lorraine

Metz, France

jean-luc.herrmann@univ-lorraine.fr

Carmen Lopez

University of Southampton

Southhampton, UK,

c.lopez@soton.ac.uk

[*] Corresponding Author

Science is about Corroborating Empirical Evidence, even in Academic Business Research Journals

Abstract

Corroboration of empirical insights is critical to theory development, developing generalizations from empirical findings, verifying the validity and reliability of findings, delimiting the scope of empirical findings, and increasing scientific rigor. Despite the widely acknowledged role of replication, corroboration, and synthetization in building scientific knowledge, few elite and top tier marketing and business journals seriously consider for publication papers attempting to reproduce the empirical insights of previously published studies. Consequently, a majority of the top marketing peer-review literature consists mainly of single research endeavors or multiple intra-studies within a single main study (or series of studies) with the result being that the published evidence is built largely on unverified and potentially tenuous findings. This special issue of the *Journal of Business Research* consists of thirteen articles that use different types of replication research procedures to independently attempt to corroborate previously reported marketing studies that display evidence of representing significant, relevant, theoretical and managerial contributions. We offer introspection on how corroboration fits with other movements in the statistical community to promote a better literature.

Science is about Corroborating Empirical Evidence, even in Academic Business Research Journals

1.   Introduction

In the history of academics, the Ph.D. in Business is a relatively recent phenomenon. The famous, or *infamous*, Carnegie and Ford Report of 1959 motivated many dramatic changes in higher business education allowing institutions to escape the *insulting* label of a vocational school. Part of the recommended remedy to *fix* business schools was to make them more academic, and thus, win the respect of the peer schools and colleges across campus. Consequently, academic research trumped pragmatic experience as a faculty qualification (Daniel, 1998). Therefore, publication in peer-reviewed journals (PRJs) became paramount above other considerations. Publishing in A journals became the "bottom line" for university business academicians and along with the growth in the number of business Ph.D.s came the publish-or-perish culture and the analogy of publishing as a *game* that must be not only played, but, won (Aguinis, Cummings, Ramani, and Cummings, 2020). Along the way, the rules of the game prioritized publication over science. Perhaps not coincidentally, our academic journals are increasingly criticized as practically irrelevant (November 2004).

One rule of the A journal publication game is to present a "novel" finding. Thus, authors often imply or explicitly state that their article is the first to provide evidence of some marketing phenomena. By staking the flag in the ground on an issue, the author renders others that care to follow along the same line of research powerless to defend against the reviewer charge that "if you would see so and so's article, you'd know this has already been done," therefore, disqualifying the second author from the game. But, science is not built on novelty alone. Otherwise, once an academic paper had declared the world flat, it would have stayed that way

forever. Common sense itself dictates that science is about building evidence so that conclusions can be stated based on accumulated evidence and not by being the first to show a relationship with a statistically significant p-value.

Thus, we put out a call for research in 2018 asking for submissions of research that focused on attempts to corroborate previously published marketing research of importance. Here, we elaborate on the state of business research (with an emphasis on marketing) with respect to what has been called the reproducibility crisis in academic management research (Hubbard and Carriquiry, 2019), offer insight into factors that contribute to the crisis (e.g. publication bias, backing into hypotheses, misuse of p-values and statistical inferences, lack of emphasis on managerial irrelevance), and provide an overview of the corroborative efforts of the researchers successfully responding to the call for research. In addition, we discuss possible ways to improve the academic research environment.

2.   Reproducibility of Published Findings

Scientifically-based factual knowledge builds on the accumulation of consistently verified empirical findings. Academicians, however, show confusion as to the difference between scientific facts and empirical findings (Hubbard and Carriquiry, 2019). Fundamentally, *empirical findings* (also referred to as insights) represent researchers' interpretations of statistically significant tests drawn from a single sample or multiple studies reported in an article. The main focus is on establishing the internal reliability of those reported empirical insights. In contrast, *scientific facts* are created by corroborating or reproducing empirical findings over time through independent replication research. Scientific facts depend not only on establishing internal reliability, but more importantly, they require external validity or the generalizability of reported empirical research findings. Recently, Hubbard and Carriquiry (2019) reemphasized

that the failure of academic journal authors to corroborate and reproduce statistical test results from original research studies feeds a *reproducibility crisis* that impedes theory development, theory testing, and the growth of a practically reliable body of knowledge in marketing, business, and the social sciences in general. In short, the journals are not producing the type of knowledge that would allow for confident prescriptions.

The psychology discipline, which overlaps with marketing and management, saw evidence of the disconnect between findings from articles and accumulated scientific facts in a large-scale attempt to reproduce findings from A+ psychology journals. In what could be considered shocking results, only about a third of statistically significant findings could be reproduced using the same procedures (sometimes with the cooperation of) as the original authors. Further, the corroborative attempts produced effect sizes that averaged less than half of the size of those reported in the top journals (Open Science Collaboration, 2015).

Many factors contribute to the reproducibility crisis. One of the factors lies in the obvious notion that replication is required in science but it is often against the rules of A+ publication as top journals eschew replication for novelty. A related factor lies in the connection between finding support for hypothesis and success. Thus, questioning the hypothesis by testing it again could be seen as an afront to TRUTH (Babin, Griffin, and Hair, 2016). Rather than ascribing some presumption of certainty to a null-hypothesis statistical test (NHST), often involving an implausible null, the researcher "embraces variation and … full reporting of results" (McShane et al. 2020, p. 240). Thus, the researcher expects varying results across corroborative attempts and practices, not only preaches, that hypotheses are not *proved* true by a significant NHST. True, a strict replication in behavioral research may prove difficult than in physical research. But, corroborative attempts are no less important. Other factors working aginst a climate of corroboration lie in the reporting of statistical results. Sample idiosyncrasies and deficiencies,

statistical-power levels, researchers' incorrect applications and interpretations of statistical test results and misalignment of captured data with research questions all contribute (Hubbard, 2016; Stodden, 2015). Researchers' reluctance to share original data, codes, and/or software used in producing the original statistical results also contribute (Sodden et al., 2013). Factors that negatively influence corroboration and reproducibility of originally reported empirical statistical results and insights can be categorized as due to: (1) different aspects of publication bias, (2) misuse of statistical inferences, (3) types of replications, and (4) relevance to managerial research questions (Dickerson 1990).

3.   Publication Bias

The mere existence of publication bias as a scientific stream of research signals a disconnect between publication and science. Publication bias exists because top journals are more likely to publish papers with statistically significant hypotheses tests (i.e., statistical differences) compared to papers with findings failing to reject statistical null hypotheses (Klaus and Edvardsson, 2014; Rothstein, Sutton, and Borenstein, 2005). Thus, reviewers and editors tend to prioritize statistical significance over the quality and rigorousness of the research and contribution potential of the research question itself (Houston, 2019; Hubbard and Carriquiry, 2019). Publication bias implies that the answer is more important than the question. Authors then, apparently rightly so, believe that journal editors and reviewers judge statistically insignificant results as relatively unworthy of dissemination in quality journals. Meta-analytic procedures are corroborative in nature and provide tools for assessing publication bias (Rothstein et al., 2005).

Faced with strongly motivating pressures like denial of promotion and/or tenure, or simply from other explicit and implicit rewards of winning the A-publication game, authors, often encouraged by advisors or the review process itself, play to win. An experienced academic would be naïve to deny that authors at times (or as a rule rather than the exception) undertake *p-hacking,* searching through relationships with the goal of building an article around the statistically significant results (Goodman, 1999; Hubbard and Carriquiry, 2019), and HARKing, which stands for hypothesizing after the results are known (Kerr, 1998; Murphy and Aguinis, 2019), to enable reporting of statistically significant results (Hubbard et al., 2019). How often do discussions, formal or otherwise, about potential papers hinge on "tell me, what is significant?" Thus, p-hacking proceeds the HARKing of hypotheses all then presented in a hypo-deductive package (Babin, Griffin, and Hair, 2016). In other words, hypotheses are cherry-picked out of a net that captures statistically significant results rendering other relationships unimportant (Murphy and Aguinis, 2019). Afterwards, the *theoretical* front-end of the article is born! While something like p-hacking could play a scientific role in discovery, authors, steeped in the hypo-deductive reporting of research, learn not to report p-hacking details. Authors emphasize, and academic journals show more interest in publishing, "gap-filling," "novel empirical insights" that follow from deductively-derived hypotheses (Hubbard, Haig and Parsa, 2019; MacInnis et al., 2020). Corroborative efforts or articles addressing controversial issues take a back seat by being perceived as neither gap-filling nor novel (MacInnis and Folkes, 2010; Espinosa and Ortinau, 2016).

3.   Statistical Inference versus Scientific Inference

3.1 Sampling

Another factor limiting academic researchers' ability to build marketing literature based on scientific knowledge relates to authors' misuse of *significant-difference testing* procedures and *statistical inferences*. Countless studies provide neither an accurate description of the defined target population nor the use of random sampling procedures in drawing samples. Inferences are routine but acknowledging to whom they refer is rare (Gigerenzer and Marewski, 2015; Hubbard et al., 2019). In addition, many authors confound the use of formal significant difference testing by limiting the data sources to non-probabilistic convenience samples like college students or crowdsourced internet workers, such as Amazon M-Turkers (Espinosa and Ortinau, 2016). Experimentally-oriented authors voice primarily concern over the internal reliability (or validity) of the test results and do not worry about using random samples of prospective subjects from well-defined target populations. External validity issues of reported empirical results are not a concern. Yet, the authors report data from experimental designs using statistical differences tests as though the data originated from random samples of well-defined target populations (Winer, 1999). Seldom does the article suggest the population of inference.

Researchers explicitly define the target population being studied to correctly make and discuss empirical statistical inferences. Once the target population is defined, a researcher identifies an appropriate sampling frame and then applies a probabilistic, random-sampling procedure to draw the required sample of size necessary for an inference of specific precision (Hahn and Meeker, 1993; Iacobucci and Churchill, 2010). While some error creeps in as sample, sampling frame, and target population rarely align perfectly, the ignorance of such approaches

has led the academic audience to treat probabilistic and non-probabilistic sampling approaches as if the consequences were the same.

3.2 Null Hypothesis Statistical Testing (NHST)

Hubbard et al. (2019) argue scientific-knowledge depends more on *significant sameness* than on *significant differences* (Haig and Evers, 2016; Hubbard, 2016). In other words, previous research should establish the baseline effect size about which a statistical inference should be made. With NHST, typical hypotheses (i.e., attitude is positively related to intention; ease of use is positively related to use..) are tested against a 0 baseline (prior probability). Never asked is the question of "what is really the possibility of 0 relationship in such case?" Under significant sameness, hypotheses would be examined not against 0 but on a reasoned standard. The significant sameness approach considers the heterogeneity of effect-sizes across studies and would place greater light on practically significant findings. Rosenthal and Rosnow (1984) expressed similarly a belief that replicability of previously reported empirical findings is the most important factor in building genuine scientific knowledge of phenomena. From a logical processing perspective, corroborated empirical insights lead to verifiable facts, and accumulation of confidence in those facts provides an underpinning for scientific-knowledge, theory, identification of boundary conditions, and predictive generalizations across different population settings. Hubbard and Lindsay (2013) provide an excellent treatise contrasting the *significant difference* and *significant sameness* paradigms in academic marketing and business research.

3.3 Bifurcation by p-value

Authors could benefit from a greater understanding of precision. Authors often report statistics to 3, 4, or more decimals places. Did their measures begin with that much precision?

Survey items most typically are captured with the 1-7 scale format; that is, one significant digit. With the inference on statistical difference testing, bifurcation often occurs along a p-value of unspecified precision .05; or is it .050, or .0500? In the latter case then, "good" or "successful" findings are determined with a p = 0.499. However, an author would be disappointed with the poor or unsuccessful result of p = 0.501 (Wasserstein et al., 2019). Does a difference in p-value of 0.002 really bifurcate knowledge and meaning? How often, even if it would, do academic marketing researchers have such precision? Research conducted under a false sense of precision in statistical inferences is not likely to lend itself to replication.

4.   Intra-study and Inter-study Replications

As discussed earlier, building scientific knowledge and literatures about different marketing, business, and/or consumer behavior phenomena requires the accumulation of empirical insights about phenomena that are not only internally reliable within the original study but also have external validity. In an effort to overcome the external validity limitations of empirical findings from single studies or experiments, Easley et al. (2000) advocate that academic researchers should conduct multiple *intra-study replications* within the main investigation using significant difference [p<.05] testing procedures. The replications could involve different contextual settings, populations, scale measurements, and sampling units. More recently, some researchers suggest that intra-study approaches benefit corroborative efforts by allowing researchers to incorporate multiple measures and different methods into the main study (Davis et al., 2011; Uncles and Kwok, 2013). In addition, intra-study approaches encourage the application of multiple theoretical perspectives in explaining the collected data (Lewis and Kelemen, 2002; Tadajewski, 2008). Furthermore, intra-study replications allow the random sampling of different sub-populations of subjects from the defined target population under investigation, possibly

negating convenience versus non-convenience sampling issues (Espinosa and Ortinau, 2016; Peterson, 2001). Although elite academic marketing and business journals strongly encourage researchers to employ multiple intra-study replications, the unresolved drawback of these types of replications is the fact that the corroboration attempts are not separate independent studies conducted by outside independent researchers. Park et al. (2015) point out that a *true* replication should be a duplication of an earlier published original research endeavor in either an exact, partial, or extended form by independent researchers in a separate and independent setting. Others argue that by only employing statistical difference testing procedures without inter-study (experiment) replications by independent researchers, the empirical results and insights from multiple intra-study (experiment) replications can only reinforce the internal reliability, not external validity, of the reported empirical results and insights (Hubbard and Lindsay, 2013; Kwon et al., 2017).

The use of *inter-study replication* research practices to independently corroborate previously reported findings, combined with the adoption of statistical sameness testing procedures, provides an avenue to build confidence in empirical generalizations. Furthermore, some researchers argue that using differentiated inter-study replication studies with the statistical sameness paradigm helps validate previously reported empirical insights and protects marketing and business literatures from uncritical acceptance of suspicious empirical results, speculative interpretive insights, Type I errors, other mistakes, and fraudulent practices (Hubbard and Armstrong, 1994, Hubbard and Lindsay, 2013; Kwon et al., 2017). Further, multiple study papers need to report all studies, even those that did not produce the desired results. All of these need to be included in a meta-analytic synthesis to provide a true picture of results. Otherwise, our published articles further the problems with publication bias and the authors, perhaps unintentionally, contribute to a misleading literature.

5.   **Managerial Relevance Issues**

While nearly all authors publishing in marketing and consumer research journals would agree that theory is important, the recent decades have seen a dearth of organic, or indigenous, theory development in deference to increasingly narrow applications of theory derived elsewhere (Hunt, 2020). Indigenous theory development may have been one of the casualties of the emphasis on "theoretical" research called for in the Ford and Carnegie reports. Rather, contributions became defined by increasingly narrow, and perhaps forced,applications of psychological *theories du jour* (Babin and Moulard, 2018) with an emphasis on statistical interactions over main effects (Alba, 2012). Statistically significant interactions with any size of effect are considered as support for boundary condition hypotheses. However, interactions often can be significantly different than 0 with small effect sizes so that such small effect-size interactions are practically irrelevant and prove resistant to replication (see Iyer and Griffin, 2020 in this issue).

Consequently, authors often adopt such a gap-filling approach by stating a "first-strike" contribution ("we are the first to show the interaction …"). However, the approach encourages ever narrower applications of non-indigenous theories producing so-called novel, but often practically unimportant, findings. Articles in this vein tend to focus on narrow topics analyzed with complex methods applied to convenience samples that seem to be purely academic exercises. The end result, as has been said, is that we "know more and more about less and less" and are well on track to know "everything about nothing" (Klaus and Edvardsson, 2014).

The journals have to publish something and what gets published is a sample of the research being spawned at business and management schools. Managers seldom turn to the academic literature for guidance; the research is not focused on solving real-world managerial or decision-makers' policy problems (Klaus et al., 2014). Bennis and O'Tool (2005) point out that academic marketing and business authors have created a vast wasteland of misleading insights

not useful for business decision-makers and warn practitioners to ignore academic research insights. Articles that do offer managerial implications often include recommendations that could have been reached without the findings of the study, speculation beyond the results, or self-serving opinions based on tenuous results that likely would defy corroborative attempts if there was motivation to try and take such steps (Hubbard and Lindsay, 2013; Klaus and Edvardsson, 2014). Hubbard and Carriquiry (2019) conclude that much reported empirical research does not translate into *stubborn facts* that practitioners can understand and use in management settings. They argue that there is simply a lack of repeatable information that can be generalized over a wide range of populations and conditions, suggesting most reported empirical findings and stated implications are, at best, anecdotal not applicable in practice.

6. **Corroborating previous published empirical insights**

The concern about building marketing and business literatures with accumulative scientific knowledge and facts serves, in part, as the underlying objective behind this special issue of the *Journal of Business Research* (*JBR*). The articles in this special issue demonstrate different types and levels of intra-study and inter-study replication, corroboration, and synthetization of reproducible empirical insights. The articles illustrate the benefits and difficulties of undertaking replication research. As with all special issues in journals, the success of this special issue is due to the collaborative efforts of many people. We want to thank all the authors who initially submitted their research to our extensive review process and the authors of the thirteen accepted articles for excellent work and patience through the publication review process. Next, we thank all the *JBR* reviewers for their constructive insights during the review process that helped improve the quality of the papers. This special issue would not have been possible without their strong dedication to *JBR*. While the four guest editors were instrumental in guiding the journey

of this special issue, we are indebted to Naveen Donthu, Georgia State University and Anders

Gustafsson, BI Norwegian Business School the Editors-in-Chief of *JBR* for believing in and

supporting the topic and our vision of creating this special issue.

Due to the diversity in the replication procedures used in the different articles, we

organize the special issue using categorical types similar to those classification categories

provided by Tsang and Kwan (1999). Articles represented either "exact/reproduction

replication." "conceptual/constructive replication with extension," "partial/differentiated

replication with extension," or "scale measurement/instrumental replication" attemps.

6.1   *Exact/reproduction replication article*

Exact replication studies (also referred to as a reproduction) represent studies that basically

repeat the original study using like methods and procedures and samples that ostensibly represent

the same population. The objective of this type of replication is to keep the conditions as similar

as possible to those of the original study to best assess the degree of corroboration of the original

empirical findings.

6.1.1   *Word-of-mouth usage*

Iyer and Griffin (2020) undertake two separate inter-study replications using the exact

replication approach to corroborate the Martin and Lueg (2013) article about marketing word-of-

mouth usage. Using the same methods, procedures, and drawing samples from the same

population as in the original study, neither of two replication studies fully corroborate the

original results. The authors uncover inconsistencies in the reporting of results in the original

study, including reported degrees of freedom that do not clearly match the model being tested.

In addition to summarizing the extent of corroboration, Iyer and Griffin (2020) provide a very

thoughtful and informative discussion describing the difficulties in conducting replication research.

6.2   Conceptual/constructive replication with extension article

Conceptual replication studies attempt to corroborate an original study's reported insights by using samples from the same population and theories but using different methods and procedures than the original study. In this approach, observed differences might occur due to the way of measuring constructs, structuring the relationships between constructs, and analyzing of the data.

6.2.1   Dominant logic salesperson job satisfaction

Rajabi, Boles, Brashear-Alejandro and Sarin (2020) re-analyze the Brown and Peterson (1993) meta-analytic model and its relationships between salesforce job satisfaction, organizational commitment, turnover intentions, role ambiguity and conflict. One unique aspect of the authors' conceptual-with-extension replication approach is that they use a longitudinal sample of salespeople that were tracked over a four-year time frame. The authors' replication study corroborates many of the findings reported in Brown and Peterson's (1993) meta-analysis study. Behavioral models that are decades old are in need of corroboration as the socialization process affects behavior.

6.3   Partial/differentiated replication with extensions

This form of replication occurs when researchers deliberately modify some of the methods, conditions, and/or procedures used in the original study or experiment and other elements are duplicated as in the original study. One of the aims is not only to attempt to corroborate some of the previously reported empirical insights but also to extend the range of conditions (i.e., new mediators and/or moderators) under which the original results still hold. There are four articles that use this type of replication approach.

6.3.1 People as Products

The Farshid, Ferguson, Pitt, and Plangger (2020) article corroborates Hirschman's (1987) *Journal of Marketing* article that investigates people as though they are products. The replication study uses a different context (i.e., online dating service) and extends the initial original empirical insights by using new methodologies to examine partner-seeking (also referred to as human branding phenomenon) through a different theoretical lens.

6.3.2 Self-accountability in ethical consumption

Tran and Paparoidamis (2020) attempt to replicate and extend the original empirical findings reported by Peloza, White, and Jingzhi (2013). That study concerned the relationships between self-accountability and ethical consumption behaviors. Using a series of four intra-study online samples across different national cultures (i.e., India and United States), the authors attempt to not only corroborate most of the original empirical findings but also extend results to different product types, specific advertising designs factors (i.e., message framing and colors) and contextual factors.

6.3.3 Goal-relevant versus incidental similarity

Arndt, Karande, Harrison, and Khoshadam (2020) attempt to partially corroborate then extend the empirical findings regarding outcomes from varying degrees of goal-relevant and incidental similarity between service providers and customers originally reported by Jiang, Hoegg, and Chattopadhyay (2009). The replication study includes four intra-study experiments using college student samples and the results highlight the benefits and limitations of incidental similarity and their specific conditions. One managerial implication offered is that goal-relevant (matching sex in a wait-loss context as opposed to matching birthday) similarity has a greater influence on service provider choice than does incidental similarity.

6.3.4 Range extension on attraction effect

The last partial/differentiated replication study, by Padamwar, Dawra, and Kalakbandi (2020), focuses on extending original findings about range extension significantly changing preference between core alternatives and the systematic amplification of attraction effect in both stylized and natural settings. The authors conduct three intra-studies using postgraduate college students and attempt to replicate their own empirical insights reported in studies 1 and 2 with conducting study 3. Yet, the research attempts to corroborate findings from Huber, Payne, and Puto (1982), Huber and Puto (1983).

6.4 Scale measurement/instrumental replication with extension articles

This type of replication study proved most popular among replication research methods . In a scale measurement replication study, independent researchers conduct inter-studies that duplicate (meaning complete adoption) the original scale measurement with a different population sample. Most often the main emphasis is that of validating the scale measure/instrument's reliability and validity levels by cross-validating fit or comparing the scale measurement's reliability and validity values of the replication study back to those values from the original study. When conducting a scale measurement replication study, some researchers add a specific extension that goes beyond corroborating the reliability and validity levels of an original scale measurement.

6.4.1 Hedonic and utilitarian shopping value scale

Picot-Coupey, Krey, Hure and Ackermann (2020) incorporate a series of independent inter-study scale measurement replication and extension studies corroborating the structure, reliability, and validity of the Babin, Darden, and Griffin's (1994) "hedonic and utilitarian shopping value" (PSV) scale. Following exact scale measurement replication procedures, the authors first replicate the PSV using samples representing similar populations. In addition, the corroboration mirrors the original results confirming that the shopping value scale constitutes a

two-dimensional structure matching work (utilitarian) and fun (hedonic) elements. Through extensive comparative research, the authors provide scale modifications for assessing utilitarian and hedonic shopping value in "online" and "mobile app" shopping experiences. The authors caution both researchers and practitioners about blindly incorporating or adapting scales developed in one customer experience context to others.

6.4.2   Brand love scale

Sajtos et al. (2020) replicate and extend Batra, Ahuvia, and Bagozzi's (2012) brand love scale. Using an exact scale/instrumental replication approach, the authors replicate the original scale in six cultural contexts (Australia, China, France, UK, US northeast, and US Southeast) across four continents. Results suggest corroboration of brand love as a second-order factor in five of the six samples. The results were not corroborated for Australia. Yet, there are differences in the strength of the reflective dimensions of the brand love scale across some of the contextual settings and the authors emphasize the need for careful adaption and not a mere translation of the reflective scale item indicators in each cultural context. From a nomological framework perspective, the brand love scale displayed similar networks of relationships as in the original study in four of six cases.

6.4.3   Employer brand equity and employee well-being and loyalty scale

The Benraiss-Noailles and Viot's (2020) article uses the scale measurement/instrumental replication approach to corroborate Berthon, Ewing, and Hah's (2005) employer brand equity (EBE) scale, which they refer to as the "employer attractiveness" (EmpAt) scale. The EmpAt conceptually consists of five dimensions, including economic value, interest value, social value, development value and application value. In addition to corroborating the EBE scale's validity from the original study, the authors provide an extension by testing the predictive relationships between EBE and positive/negative employee well-being and employee intentions to leave the

job. The results of the extension demonstrating relationships between EBE and employee well-being and intention to leave the company need further corroboration.

6.4.4 *Experiential value versus Perceived Value scales*

Gallarza, Maubisson, and Riviere (2020) attempt to cross compare Sweeny and Soutar's (2001) PERceived Value (PERVAL) scale with Mathwick, Malhotra, and Rigdon's (2001) Experiential Value (EVS) scale within a tourism context. Using a form of scale/instrumental replication, the authors replicate the conceptual structure of the PREVAL and EVS measurements. The authors also claim that both the PERVAL and EVS scales are suitable for measuring consumer value in extended contexts (e.g., cultural heritage experiences) and other geographical areas (e.g., non-English-speaking countries such as France).

6.4.5 *Customer experience quality scale*

Kuppelwieser and Klaus (2020) attempt to corroborate Klaus's (2014) three-dimensional customer experience scale using scale/instrumental replication procedures. The attempts include SEM results aimed at cross-validation from a B2C contextual with a new sample of 200 selected UK consumers to a second intra-study sample using a B2B. The multigroup SEM results provide some support for factor structure equivalence but do not support measurement equivalence across contexts. In addition, the authors question the dimensionality of the original scale. The authors suggest a need for further research to better refine EXQ's psychometric measurement. In particular, because evidence for metric invariance is lacking, the corresponding EXQ items should perhaps change depending on the application context.

6.4.6 *Consumer brand engagement scale*

The Obilo, Chefor, and Saleh (2020) article focuses on replicating the widely-cited Hollebeek, Glynn, and Brodie (2014) consumer brand engagement (CBE) scale. The authors closely address the psychometric issues with the CBE scale. Specifically, the authors point to

potential face and discriminant validity issues that intermingle CBE dimensions themselves and with other well-established consumer concepts. The results point to the difficulty in defining CBE uniquely from other constructs. The authors undertake an extension effort as they create an alternative conceptualization and operationalization of "customer engagement." The reported results provide empirical support for the new conceptualization and invite further corroborative attempts in the near future.

6.4.7   Trustworthiness in e-commerce scales

Hallikainen and Laukkanen (2020) attempt to investigate trustworthiness in e-commerce construct using two different scale measurements; one scale being McKnight, Choudhury, and Kacmar (2002) and the other being Genfen and Straub (2004). Hallikainen and Laukkanen study trustworthiness across two different cultural contexts, China and Finland. The authors suggest that the two cultural investigations of the two different measures of trustworthiness represent a conceptual replication with extension because they use similar methods and measurement instruments. The reported research better represents a "generalization with extension" corroborative attempt than a conceptual replication study. In essence, the authors conduct two intra-studies that allow for a comparison of the two alternative measures of trustworthiness.

7.   Some concluding comments

The *JBR* Special Issue on corroborating original empirical insights from published marketing and consumer behavior research represents a small step toward developing factual knowledge from the academic marketing research. The mindset of "novel" findings defined by statistical significance (typically $p < 0.05$) is strong and remains evident in some of the articles comprising the special issue. Perhaps the marketing discipline will follow other disciplines that are turning away from thatmindset in pursuit of a more pragmatically reliable and useful academic literature

and, therefore, develop a scientific body of knowledge containing stubborn facts upon which theory can grow, including indigenous theory. With that in mind, a few summary points can be offered that illustrate the challenges to create a a corroborative spirit in business, marketing and consumer research journals:

- The dominant formula for success in our A journals remains one strongly based on a hypo-deductive presentation of results that emphasize statistical significance of "novel" findings (mean differences and regression parameters). Corroborative efforts are often discouraged as less worthy (aside from special avenues like this one).

- Reviewers and editors likely were trained and socialized into the publishing environment that emphasized the standard presentation of journal articles. Editors, in particular, likely enjoyed success with the publish or perish authorship model and they may find it difficult to buy in to the subsequent problems. Ironically, journals emphasize novelty but even papers purporting "novel" results typically adopt methods and are written in a deductive style more appropriate for the context of justification than discovery. The top journals do not publish a lot of truly discovery oriented research.

- Doctoral programs often emphasize publication as the primary academic goal over the development of a factual, scientific knowledge base, and thus, academic researchers receive relatively little training into corroborative research approaches, including inter-study and independent replication research. Doctoral programs also greatly emphasize training in complex methodologies over the training in basics like statistical inference testing and sampling. Consequently, many learn to use a p-value but may not know exactly what it is (Wasserstein et al., 2019). How many students receiving their Ph.D. hood know that it is the probability of rejecting the null hypothesis *assuming the null is true*?

- Articles reporting results from survey research rarely draw probability samples from sampling frames representative of a specifically specified, practically relevant, target population. In contrast, many authors report results using non-probabilistic convenience samples of sometimes dubious integrity. Thus, generalizability is seldom claimed and generally ignored.

- From a relevance perspective, who really are the main audiences for the research results published in the marketing discipline's elite journals? Is the value of a journal limited when it is essentially a group of academics writing with the same group of academics as the audience?

Thus, how will the problems be overcome? Listed below are a few suggestions aimed at creating a more valuable academic marketing and business research literature. Many ideas that relate to the ways we draw statistical inferences are influenced by the special issues of the American Statistician (*Statistical Inference in the 21st Century*) that are aimed at changing the way academics use statistics and do research (see Wasserstein et al., 2019).

- Marketing and business research absolutely needs more articles attempting to corroborate important empirical insights from published studies. If researchers are to be rewarded for those efforts, journals need to be more open to publishing such research (i.e., well-designed replications reporting and interpreting with equal prominence "negative" (disconfirming) as well as "positive" (confirming) results). Guidelines for articles in need of replication need to be established. Here, we asked authors to look for articles with a high volume of citations (for example, at least 100 over 5 years). Other criteria need to be considered. We asked all authors contributing to the special issue to specify why the target article was in need of a corroborative attempt. Straight-forward meta-analytic studies that focus on synthesizing effect sizes of key relationships also should be

encouraged. They need not establish new relationships or findings, but analyze studies

reporting results on important marketing relationships.

- Doctoral students should be encouraged to conduct corroborative studies. Attempts at

  replicating others' studies provide an effective avenue for learning research methods.

  The experience would accelerate learning and the papers would help move findings

  closer to facts.

- Authors should invite and report sufficient details to facilitate corroborative attempts

  including replications and syntheses. ,Authors need to report the details necessary to

  facilitate future meta-analyses. Descriptive statistics (means, medians, standard

  deviations, correlations or covariances, sample size details (by cell size and by basic

  demographic variables), control group results for experiments. Authors also should

  report null effects including those that were part of the research project but were not

  included as part of the published article.

- Changes in the way research is reviewed for publication should be considered. One

  alternative is to mimic in form a traditional dissertation proposal defense in which the

  research is presented pro-forma and the critical decision is made at that point. In a

  process sometimes referred to as registration or pre-registration (Calin-Jageman and

  Cumming, 2019), the emphasis shifts from statistically significant results to the relevance

  and meaningfulness of the research question(s). If a question is important to ask, its value

  does not depend on the outcome (Babin et al., 2016). Conditional acceptance for

  publication then rests on the question(s) and the implications (both theoretical and

  practical) that would stem from knowing the question's answer and the

  soundness/appropriateness of the proposed methodological approach. Thus, the process

  addresses the HARKing issue and provides the benefit that reviewer suggestions can be

considered before a study is conducted. So long as the researchers follow through with their plans competently, the paper would be accepted for publication. That is, the results are reported after the submission is decisioned.

- Top journals should welcome inductively-presented discovery research that presents interesting and practically relevant data but lacks a theoretical conceptual development as a precursor. Instead, suggestions for theoretical development could be offered in the discussion. Moreover, in a climate of corroborative research, other researchers would jump in to try to reproduce the results and in the process of examining effect sizes and heterogeneity, theoretical development would be fostered.

- End traditional NHST. Researchers should not divide results into good and bad based on a threshold p-value. The significant sameness paradigm provides a more logical alternative where hypotheses should not be judged against a 0 null, but rather against a reasoned baseline (or prior). For corroborative research, a previously reported effect-size provides a reasoned basis. Thus, a hypothesis would specify the estimated effect size and not just speculate on the direction of a relationship.

- Robustness checks on effect sizes are also wise. As part of the corroborative effort, consider employing different control variables than in the original study, perhaps even use different measures of the same construct, and apply more than one statistical approach or model (Meyer, van Witteloostuijn, and Beugelsdijk, 2017).

- Researchers should place greater weight on practical effect sizes (known with confidence) than on null-hypothesis testing statistical significance. Thus, effect sizes (coefficients and marginal effects, if appropriate) should be reported. Additionally, rather than referring to a threshold p-value or reporting asterisks to signal p-value thresholds, report the actual p-values.An educated readership can interpret the actual value without

being told something obvious. Confidence intervals and prediction intervals provide

better ways of depicting the effect size (as opposed to only a point estimate) and the

expected range of effects that may be found. For instance, a small effect with a narrow

prediction interval is likely more meaningful and more amenable to corroboration than

the same size effect (or even larger) with a wide prediction interval.

Each of these suggestions is offered as a means of helping to work toward greater and more

valuable scientific knowledge development in marketing. In doing so, the academy can enhance

the credibility of our academic pursuits. Is not that the crux of the mission of our academic

community? Further, would it be that the Ford and Carnegie report would have led us to be

researchers and not just authors?

Reference Sources

Aguinis, H., Cummings, C., Ramani, R.S. and Cummings, T.G. (2020), An A is an A: The new bottom line for valuing academic research. *Academy of Management Perspectives*, 34, 135-154.

Alba, J.W. (2012). In defense of bumbling. *Journal of Consumer Research*, 981-987.

Arndt, A.D., Karande, K., Harrison, K. and Khoshgadam, L. (2020), Goal-relevant versus incidental similarity when choosing between multiple service providers. *Journal of Business Research*, forthcoming.

Babin, B.J. & Moulard, J.G. (2018), To what is the review process relevant? What's right and what's wrong with peer review for academic business journals. *European Business Review*, 30 (2), 145-156.

Babin, B. J., Darden, W. R., & Griffin, M. (1994). Work and/or fun: Measuring hedonic and utilitarian shopping value. *Journal of Consumer Research*, *20* (4), 644-656.

Babin, B.J., Griffin, M., Hair, J.F. (2016). Heresies and sacred cows in scholarly marketing publications. *Journal of Business Research*, 69, 3133-3138.

Batra, R., Ahuvia, A., & Bagozzi, R. P. (2012). Brand love. *Journal of Marketing*, *76* (2), 1-16.

Bennis, W. and O'Tool, J. (2005). How business schools lost their way. *Harvard Business Review*, 8 (3), 96-104.

Benraiss-Noailles, L., and Viot, C. (2020). Employer brand equity effects on employees' well-being and loyalty. *Journal of Business Research*, forthcoming.

Berthon, P., Ewing, M., & Hah, L. (2005). Captivating company: Dimensions of attractiveness in employer branding. *International Journal of Advertising*, 24 (2), 151-172.

Brown, S. P., & Peterson, R. A. (1993). Antecedents and consequences of salesperson job satisfaction: Meta-analysis and assessment of organizational commitment. *Journal of Marketing Research*, *30* (1), 63-77.

Calin-Jageman, R. and Cumming, G. (2019), The new statistics for better science: Ask how much, how uncertain, and what else is known. *The American Statistician*, 73, 271-280.

Daniel, C.A. (1998). *MBA:the First Century*. Bucknell Press: Lewisburg, PA.

Davis, D.F., Golicic, S.L. and Boerstler, C.N. (2011). Benefits and challenges of conducting multiple methods research in marketing. *Journal of the Academy of Marketing Science*, 39 (3), 467-479.

Dickerson, K. (1990). The existence of publication bias and risk factors for its occurrence. *Journal of the American Medical Association*, 263, 1385-1389.

Farshid, M., Ferguson, S.L., Pitt, L. and Plangger, K. (2020), People as products: exploring replication and corroboration in the dimensions of theory, method and context. *Journal of Business Research*, forthcoming.

Easley, R.W., Madden, C.S. and Dunn, M.G. (2000). Conducting marketing science: The role of replication in the research process. *Journal of Business Research*, 48 (1), 83-92.

Espinosa, J.A. and Ortinau, D.J. (2016). Debunking legendary beliefs about student samples in marketing research. *Journal of Business Research*, 69 (8), 3149-3158.

Gallarza, M.G., Maubisson, L., and Riviere, A., (2020). Replicating consumer value scales: a compariative study of EVS and PER VAL at a cultural heritage site. *Journal of Business Research*, forthcoming.

Gefen, D., & Straub, D. W. (2004). Consumer trust in B2C e-Commerce and the importance of social presence: Experiments in e-products and e-services, *Omega*, 32 (6), 407–424.

Gigerenzer, G. and Marewski, J.N. (2015). Surrogate science: The idol of a universal method for scientific inference. *Journal of Management*, 41, 421-440.

Goodman, S.N. (1999). Toward evidence-based medical statistics: The p-value fallacy, *Annals of Internal Medicine*, 130, 995-1004.

Hahn, G.J. and Meeker, W.Q. (1993). Assumptions for statistical inference. *The American Statistician*, 47, 1-11.

Haig, B.D. and Evers, C.W. (2016). *Realist Inquiry in Social Science.* Thousand Oaks, CA: SAGE Publications.

Hallikainen, H., and Laukkanen, T. (2020). Trustworthiness in e-commerce: a replication study of competing measures. *Journal of Business Research*, forthcoming.

Hirschman, E. C. (1987). People as products: Analysis of a complex marketing exchange. *Journal of Marketing*, *51* (1), 98-108.

Hollebeek, L. D., Glynn, M. S., & Brodie, R. J. (2014). Consumer brand engagement in social media: Conceptualization, scale development and validation. *Journal of Interactive Marketing*, 28 (2), 149-165.

Houston, M.B. (2019). Four facets of rigor. *Journal of the Academy of Marketing Science*, 47 (4), 570-573.

Hubbard, D.W. and Carriquiry, A.L. (2019). Quality control for scientific research: Addressing reproducibility, responsiveness and relevance. *The American Statistician*, 73 (s1), 46-55.

Hubbard, R. (2016). *Corrupt research: The case for reconceptualizing empirical management and social science.* Thousand Oaks, CA: SAGE Publications.

Hubbard, R., Haig, B.D. and Parsa, R.A. (2019). The limited role of formal statistical inference in scientific inference. *The American Statistician*, 73 (s1), 91-98.

Hubbard, R. and Armstrong, J.S. (1994). Replications and extensions in marketing: Rarely published but quite contrary. *International Journal of Research in Marketing*, 11 (3), 233-248.

Hubbard, R. and Lindsay, R.M. (2013). From significant difference to significant sameness: Proposing a paradigm shift in business research. *Journal of Business Research*, 66 (9), 1377-1388.

Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, *9* (1), 90–99.

Huber, J., Payne, J. W., & Puto, C. P. (2014). Let's be honest about the attraction effect., *Journal of Marketing Research*, *51* (4), 520–525.

Huber, J., & Puto, C. (1983). Market Boundaries and Product Choice: Illustrating Attraction and Substitution Effects. J*ournal of Consumer Research*, *10* (1), 31–44.

Hunt, S.D. (2020). Indigenous theory development in marketing: the foundational premises approach. *Academy of Marketing Science Review*, https://doi.org/10.1007/s13162-020-00165-w.

Iacobucci, D. and Churchill, Jr., G.A. (2010). *Marketing Research: Methodological Foundations*, 10th ed., Mason, Ohio: South-Western Publisher.

Iyer, R . and Griffin, M. (2020). Modeling word-of-mouth usage: a replication. J*ournal of Business Research*, in press.

Jiang, L., Hoegg, J., Dahl, D.W., & Chattopadhyay, A. (2009). The persuasive role of incidental similarity on attitudes and purchase intentions in a sales context. *Journal of Consumer Research,* 36 (5), 778-791.

Kerr, N.L. (1998). HARKing: Hypothesizing after the results are known, *Personality and Social Psychology Review*, 2, 196-217.

Klaus, P. (2014). *Measuring customer experience: How to develop and execute the most profitable customer experience strategies*: Springer.

Klaus, P. and Edvardsson, B. (2014). The road back to relevance: How to put marketing (and marketing scholars) back on the top managements' agendas. *Journal of Service Management*, 25 (2), 166-175.

Klaus, P., Edvardsson, B., Keiningham, T.L. and Gruber, T. (2014). Getting in with the 'In' crowd: How to put marketing back on the CEO's agenda. *Journal of Service Management*, 25 (2), 105-212.

Kuppelwieser, V. and Klaus, P. (2020). Measuring customer experience quality: the EXQ scale revisited. *Journal of Business Research*, forthcoming.

Kwon, E.S., Shan, Y., Lee, J.S. and Reid, L.N. (2017). Inter-study and intra-study replications in leading marketing journals: A longitudinal analysis, *European Journal of Marketing*, 51 (1), 257-278.

Lewis, M.W. and Kelemen, M.L. (2002). Multiparadigm inquiry: Exploring organizational pluralism and paradox. *Human Relations*, 55 (2), 251-275.

MacInnis, D.J. and Folkes, V.S. (2010). The disciplinary status of consumer behavior: A sociology of science perspective on key controversies. *Journal of Consumer Behavior*, 36 (6), 899-914.

MacInnis, D.J., Morwitz, V.G., Botti, S., Hoffman, D.L., Kozinets, R.V., Lehmann, D.R., Lynch, Jr., J.G., and Pechmann, C. (2020). Creating boundary-breaking, marketing-relevant consumer research. *Journal of Marketing*, 84 (2), 1-23.

Martin, W. C., & Lueg, J. E. (2013). Modeling word-of-mouth usage. *Journal of Business Research*, 66 (7), 801-808.

Mathwick, C., Malhotra, N.K., & Rigdon, E. (2001). Experiential value: conceptualization, measurement, and application in the catalog and internet shopping environment. *Journal of Retailing*, 77 (1), 39-56.

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, 13 (3), 334–359.

McShane, B.B., Gal, D., Gelman, A., Robert, C. & Tackett, J.L. (2020). Abandon statistical significance. *The American Statistician*, 73 (S1), 235-245.

Meyer, K.E., van Witteloostuijn, A., and Beugelsdijk, S. (2017). What's in a p? Reassessing best practices for conducting and reporting hypothesis-testing research? *Journal of International Business Studies*, 48, 535-551.

Murphy, K.R. and Aguinis, H. (2019). HARKing: How badly can cherry-picking and question trolling produce bias in published results? *Journal of Business and Psychology*, 34 (1), 1-17.

November, P. (2004). Seven Reasons Why Marketing Practitioners Should Ignore Marketing Academic Research, *Australasian Marketing Journal*, 12, 39-50.

Obilo, O.O., Chefor, E., and Saleh, A. (2020), Revisting the consumer brand engagement concept. *Journal of Business Research*, forthcoming.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349, aac4716.

Padamwar, P.K., Dawra, J., and Kalakbandi, V.K. (2020), The impact of range extension on the attraction effect. *Journal of Business Research*, forthcoming.

Park, J.H., Venger, O., Park, D.Y. and Reid, I.N. (2015). Replication in advertising research 1980-2012: A longitudinal analysis of leading advertising journals. *Journal of Current Issues and Research in Advertising*, 36 (2), 115-135.

Peloza, J., White, K., & Jingzhi, S. (2013). Good and Guilt-Free: The role of self-accountability in influencing preferences for products with ethical attributes. *Journal of Marketing,* 77 (1), 104-119.

Peterson, R.A. (2001). On the use of college students in social science research: Insights from a second order meta-analysis. *Journal of Consumer Research*, 28 (3), 450-461.

Picot-Coupey, K., Krey, N., Hure, E., and Ackerman, C.L. (2020). Still work and/or fun? Corroboration of the hedonic and utilitarian shopping value scale. *Journal of Business Research*, forthcoming.

Rajabi, R., Boles, J., Brashear-Alejandro, T., and Sarin, S. (2020). Revisiting and replicating the dominant logic on salesperson job satisfaction, organizational commitment, and turnover. *Journal of Business Research*, forthcoming.

Rosenthal, R. and Rosnow, R.L. (1984). *Essentials of behavioural research: Methods and meta-analysis*. New York, NY: McGraw-Hill Publisher.

Rothstein, H.R., A.J. Sutton, and M. Borenstein (2005). *Publication-Bias in Meta-Analysis – Prevention, Assessment and Adjustments*. Wiley: New York.

Sajtos, L., Cao, J.T., Espinosa, J.A., Phau, I., Rossi, P., Sung, B., and Voyer, B. (2020), Brand love, corroboration, replication, extension, cross-cultural comparison. *Journal of Business Research*, forthcoming.

Stodden, V. (2015). Reproducing statistical results. *Annual Reviews of Statistics and Its Application*, 2, 1-19.

Stodden, V., Guo, P. and Ma, Z. (2013). Toward Reproducible computational Research: An empirical analysis of data and code policy adoption by journals. *PloSOne*, 8, e6711.

Sweeney, J.C., & Soutar, G.N. (2001). Consumer perceived value: the development of a multiple item scale. *Journal of Retailing*, *77* (2), 203-220.

Tadajewski, M. (2008). Incommerisurable paradigms, cognitive bias and the politics of marketing theory. *Marketing Theory*, 8 (3), 273-297.

Tran, T.T.H. and Paparoidamis, N.G. (2020), Taking a closer look: reasserting the role of self-accountability in ethical consumption. *Journal of Business Research*, forthcoming.

Tsang, E.W. and Kwan, K.M. (1999). Replication and theory development in organizational science: A critical realist perspective. *Academy of Management Review*, 24 (4), 759-780.

Uncles, M.D. and Kwok, S. (2013). Designing research with in-built differentiated replication. *Journal of Business Research*, 66 (9), 1398-1405.

Wasserstein, R.L., Shim, A.L., and Lazar, N.A. (2019). Moving to a world beyond "p < 0.05." *The American Statistician*, 73 (S1), 1-19.

Winer, R. (1999). Experimentation in the 21st century: The importance of external validity. *Journal of the Academy of Marketing Science*, 27 (2), 349-358.